

ANDREW R. MCCLUSKEY  
for the degree of DOCTOR OF PHILOSOPHY

# COARSE-GRAINED MODELLING FOR SOFT MATTER SCATTERING

UNIVERSITY OF BATH  
DEPARTMENT OF CHEMISTRY

DIAMOND LIGHT SOURCE

Copyright © 2019 Andrew R. McCluskey

Attention is drawn to the fact that copyright of this thesis rests with the author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that they must not copy it or use material from it except as permitted by law or with the consent of the author.

This work is licensed under a Creative Commons “Attribution 4.0 International” license.



*First printing, June 2019*

"ATTICUS TOLD ME TO DELETE THE ADJECTIVES AND I'D HAVE THE FACTS."

Scout Finch – To Kill a Mockingbird







## **DECLARATION OF AUTHORSHIP**

I, Andrew R. McCluskey, declare that this thesis titled "Coarse-grained modelling for soft matter scattering" and the work presented in it are my own. I confirm that:

- where the thesis or any part of the thesis such as a published paper, has been produced jointly with others, that a substantial part is the original work of myself, and
- where the thesis incorporates material already submitted for another degree, the extent of that material and the degree, if any, obtained.

Signed:

---

Date:

---







## PUBLICATIONS

Some of the information in Chapter 2 has been previously published in:

- A. R. McCluskey *et al.* *Curr. Org. Chem.* 22.8 (2018), pp. 750–757.

Some of the work covered in Chapter 3 has been previously published in:

- A. R. McCluskey *et al.* *Phys. Chem. Chem. Phys.* 21.11 (2019), pp. 6133–6141.

Some of the work covered in Chapter 4 has been previously published in:

- A. R. McCluskey *et al.* *J. Phys. Comm.* Accepted (2019).

Some of the work covered in Chapter 6 has been previously published in:

- A. R. McCluskey *et al.* *J. Open Source Educ.* 1.2 (2018), pp. 19–21.
- A. R. McCluskey *et al.* *J. Appl. Crystallogr.* 52.3 (2019), pp. 665–668.



## REPRODUCIBILITY STATEMENT

This thesis exists as a piece of completely reproducible research. I have endeavoured to include as much algorithmic and methodological detail within the text, and where relevant working Code Blocks written in Python.

However, in order to provide complete, and easy, reproducibility an electronic supplementary information (ESI) is available online in the form of a Git repository. This ESI provides full details of the analyses performed in this work and access to an automated analysis workflow.

In order to avoid incorrect self-plagiarism detection this Git repository will not be publically available until after the *viva voce*, however, if either examiners would like access to the repository before the *viva voce*, this may be arranged.







## ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors at Bath; Karen Edler and Steve Parker, without whom I would never have gotten here. The freedom they have given me and the gentle guidance they have bestowed on me, has allowed me to develop during these past three and a half years more than I ever thought was possible. I hope that Karen will forgive me for being a butterfly chaser and Steve will forgive me for using Python.

I should thank my Diamond supervisors; Andrew Smith and Jonathan Rawle, for hosting me at DLS for 6 months in my first year and always being available for a chat or a drink or to gift me some beamtime.

Thanks to James Grant for being my mentor in the first year and for introducing me to the magical world of computational science.

Tom Arnold, Richard Campbell, and Andrew Jackson made the work in Chapter 3 possible with their help and advice. Chapter 4 built on some great data collected by Dave Barlow and Jayne Lawrence. Ben Morgan & Fiona Dickinson, thank you for all the help and encouragement that allowed my participation in the pythoninchemistry project.

Thanks to all the members of the Edler Group that I have gotten to know over the years; Cecilia, Kerrie, Vincenzo, Julien, Marcelo, Andi, and Ria. Thanks for the laughs on beamtime and for the willingness to suffer my chat during beamwine. Adrian for arguments at 5 am on beamtime about data analysis and drunken chats about everything at NMSUM. Oli for being there to discuss the futility of academia and for getting drunk at any opportunity, regardless of the day. Nom for being my personal chauffeur far too often and for all night beamtimes.

Cheers to CompChem for all the drunken nights, long may it continue; Phil, Georgie, Alex, Dan, Jacob. Particular shout out to Victoria; despite your late arrival, your attendance made up for it. And Adam for being my friend and housemate, listening to me complain, and for always being keen for Parade.

To all of the members of 0.15 throughout my time here; Richard, Philippe, Megan, Leyorla, Elliot, and Jack for ensuring a happy and friendly place to work every day, and for not complaining about my gym towel lying about. And to all the Chemistry football crowd, for an escape from the monotony and the allowance to run around like a headless chicken.

Ioli, you've changed my life, and I wouldn't have managed to get here without you.

Thanks to Sadie and Penny (and Sophie before that) to being a particularly beautiful way to end a presentation, and for helping me stay sane during the writing of this.

Finally, to my parents (all of them) for suffering with me as I chase this dream, but particularly my mother for letting me occupy the dining room for the last three months as I wrote this up and buying me Ben & Jerry's ice cream whenever I wanted it.



# CONTENTS

|   |  |     |
|---|--|-----|
| 1 | <i>Introduction</i>  | 1   |
| 2 | <i>Theory</i>  | 7   |
| 3 | <i>Chemically consistent modelling of X-ray and neutron reflectometry</i>                | 43  |
| 4 | <i>Applying atomistic and coarse-grained simulation to reflectometry analysis</i>        | 65  |
| 5 | <i>Using particle swarm methods for small angle scattering analysis</i>                  | 85  |
| 6 | <i>Developing open-source teaching materials for classical simulation and scattering</i> | 99  |
| 7 | <i>Summary &amp; Future Work</i>   | 111 |
| A | <i>Additional Code Blocks for Differential Evolution</i>                                 | 117 |
| B | <i>Additional tables and plots from Chapter 3</i>  | 119 |
| C | <i>Additional tables and plot from Chapter 4</i>   | 125 |



## LIST OF ABBREVIATIONS

|                                     |   |
|-------------------------------------|---|
| <b>ACMW</b>                         | air-contrast matched water                                      |
| <b>APM</b>                          | area per molecule   |
| <b>cmc</b>                          | critical micelle concentration                                  |
| <b>C<sub>10</sub>TA<sup>+</sup></b> | <i>n</i> -decyltrimethylammonium                                |
| <b>CCP-SAS</b>                      | collaborative computational project for small angle scattering  |
| <b>CPU</b>                          | central processing unit   |
| <b>DE</b>                           | differential evolution  |
| <b>DES</b>                          | deep eutectic solvent   |
| <b>DLPC</b>                         | 1,2-dilauroyl- <i>sn</i> -glycero-3-phosphocholine              |
| <b>DMPC</b>                         | 1,2-dimyristoyl- <i>sn</i> -glycero-3-phosphocholine            |
| <b>DMPG</b>                         | 1,2-dimyristoyl- <i>sn</i> -glycero-3-phospho-(1'-rac-glycerol) |
| <b>DSPC</b>                         | 1,2-distearoyl- <i>sn</i> -glycero-3-phosphocholine             |
| <b>DPPC</b>                         | 1,2-dipalmitoyl- <i>sn</i> -glycero-3-phosphocholine            |
| <b>DVTD</b>                         | differential vibrating tube densimetry                          |
| <b>GIXD</b>                         | grazing-incidence X-ray diffraction                             |
| <b>GPU</b>                          | graphical processing units                                      |
| <b>hdDES</b>                        | partially deuterated deep eutectic solvent                      |
| <b>hDES</b>                         | hydrogenated deep eutectic solvent                              |
| <b>ILL</b>                          | Institut Laue-Langevin  |
| <b>LC</b>                           | Liquid-Condensed  |
| <b>LE</b>                           | Liquid-Expanded   |
| <b>MCMC</b>                         | Markov chain Monte Carlo  |
| <b>MD</b>                           | molecular dynamics  |
| <b>MPI</b>                          | message passing interface                                       |
| <b>NB</b>                           | neutral buoyancy  |
| <b>NPT</b>                          | constant number, pressure, and temperature                      |
| <b>NR</b>                           | neutron reflectometry   |
| <b>NVE</b>                          | constant number, volume, and energy                             |
| <b>NVT</b>                          | constant number, volume, and temperature                        |
| <b>OER</b>                          | open educational resource                                       |
| <b>PBC</b>                          | periodic boundary condition                                     |
| <b>PC</b>                           | phosphocholine  |
| <b>PDF</b>                          | probability distribution function                               |
| <b>PFTE</b>                         | polytetrafluoroethylene   |
| <b>PG</b>                           | phosphatidylglycerol  |
| <b>PSO</b>                          | particle swarm optimisation                                     |
| <b>RDF</b>                          | radial distribution function                                    |

|             |   |
|-------------|---|
| <b>SANS</b> | small angle neutron scattering          |
| <b>SAS</b>  | small angle scattering                  |
| <b>SDS</b>  | sodium dodecyl sulfate                  |
| <b>SLD</b>  | scattering length density               |
| <b>SAXS</b> | small angle X-ray scattering            |
| <b>SP</b>   | surface pressure                        |
| <b>SPC</b>  | single point charge                     |
| <b>STFC</b> | Science & Technology Facilities Council |
| <b>ToF</b>  | time-of-flight                          |
| <b>VMD</b>  | visual molecular dynamics               |
| <b>wph</b>  | water molecules per head group          |
| <b>WAXS</b> | wide angle X-ray scattering             |
| <b>XRR</b>  | X-ray reflectometry                     |

# PHYSICAL CONSTANTS

|                                       |   |
|---------------------------------------|---|
|                                       | $\pi = 3.1415\dots$   |
| Planck constant                       | $h = 6.626 \dots \times 10^{-34} \text{ Js}$                |
| Golden ratio                          | $\Phi = 1.618 \dots$  |
| dielectric permittivity of the vacuum | $\epsilon_0 = 8.854 \dots \times 10^{-12} \text{ F m}^{-1}$ |
| charge of electron                    | $e = 1.602 \dots \times 10^{-19} \text{ C}$                 |
| Boltzmann constant                    | $k_B = 1.381 \dots \times 10^{-23} \text{ J K}^{-1}$        |



## LIST OF SYMBOLS

|                         |                                   |                     |
|-------------------------|-----------------------------------|---------------------|
| $a$                     | step size                         |                     |
| $a_0$                   | optimum head-group area           | $\text{m}^2$        |
| $b$                     | scattering length                 | $\text{m}$          |
| $b$                     | bond length                       | $\text{m}$          |
| $b_i$                   | best candidate solution           |                     |
| $b_0$                   | equilibrium bond length           | $\text{m}$          |
| $d_h$                   | phospholipid head layer thickness | $\text{m}$          |
| $d_t$                   | phospholipid tail layer thickness | $\text{m}$          |
| $d_x$                   | length of object                  | $\text{m}$          |
| $k_m$                   | mutation constant                 |                     |
| $l_0$                   | length of hydrophobic tail        | $\text{m}$          |
| $m$                     | mass                              | $\text{kg}$         |
| $n$                     | number of scattering vectors      |                     |
| $n_i$                   | refractive index                  |                     |
| $p$                     | packing parameter                 |                     |
| $p$                     | parallel component                |                     |
| $q$                     | scattering magnitude              | $\text{m}^{-1}$     |
| $q_{i,j}$               | electronic charge                 |                     |
| $r$                     | displacement vector magnitude     | $\text{m}^{-1}$     |
| $r_{n,n+1}$             | Fresnel equation coefficient      |                     |
| $s$                     | serial component                  |                     |
| $t$                     | time                              | $\text{s}$          |
| $t_F$                   | time-of-flight                    | $\text{s}$          |
| $A_{1,2,3}$             | dihedral constants                | $\text{J}$          |
| $B$                     | resultant matrix                  |                     |
| $E_{\text{bonded}}$     | bonded energy                     | $\text{J}$          |
| $E_k$                   | kinetic energy                    | $\text{J}$          |
| $E_{\text{non-bonded}}$ | non-bonded energy                 | $\text{J}$          |
| $E_{\text{total}}$      | total energy                      | $\text{J}$          |
| $F$                     | some function                     |                     |
| $I$                     | intensity                         |                     |
| $K_b$                   | bond force constant               | $\text{J m}^{-2}$   |
| $K_\theta$              | angle force constant              | $\text{J rad}^{-2}$ |
| $L_F$                   | length-of-flight                  | $\text{m}$          |
| $M_n$                   | layer matrix                      |                     |
| $N$                     | number of some item               |                     |
| $N_P$                   | number of magnets                 |                     |

|                      |                                     |                    |
|----------------------|-------------------------------------|--------------------|
| $N_\rho$             | number density                      |                    |
| $P(q)$               | form factor                         |                    |
| $R$                  | reflected intensity                 |                    |
| $R_{1,2}$            | some random number                  |                    |
| $R_g$                | radius of gyration                  | m                  |
| $R_i$                | incidence rate                      | s <sup>-1</sup>    |
| $R_s$                | radius of sphere                    | m                  |
| $S$                  | surface area                        | m <sup>2</sup>     |
| $S(q)$               | structure factor                    |                    |
| $T$                  | temperature                         | K                  |
| $U$                  | uniform distribution                |                    |
| $V$                  | volume                              | m <sup>3</sup>     |
| $V_c$                | volume of hydrophobic tail          | m <sup>3</sup>     |
| $V_h$                | phospholipid head volume            | m <sup>3</sup>     |
| $V_n$                | volume of layer                     | m <sup>3</sup>     |
| $V_p$                | volume of particle                  | m <sup>3</sup>     |
| $V_t$                | phospholipid tail volume            | m <sup>3</sup>     |
| <b>a</b>             | acceleration                        | m s <sup>-2</sup>  |
| <b>f</b>             | force                               | N                  |
| <b>g</b>             | global best                         |                    |
| <b>k<sub>i</sub></b> | incident wavevector                 | m <sup>-1</sup>    |
| <b>k<sub>f</sub></b> | final wavevector                    | m <sup>-1</sup>    |
| <b>m</b>             | mutant vector                       |                    |
| <b>o</b>             | offspring population                |                    |
| <b>p</b>             | parent population                   |                    |
| <b>q</b>             | scattering wavevector               | m <sup>-1</sup>    |
| <b>r</b>             | displacement                        | m                  |
| <b>s</b>             | personal best                       |                    |
| <b>v</b>             | velocity                            | m s <sup>-1</sup>  |
| <b>x</b>             | position                            | m                  |
| <b>R</b>             | particle position                   | m                  |
| $\beta_c$            | fraction of the speed of light      |                    |
| $\beta_n$            | phase factor                        |                    |
| $\delta$             | infinitesimally small change        |                    |
| $\varepsilon$        | depth of potential well             | J                  |
| $\zeta$              | figure of merit                     |                    |
| $2\theta$            | scattering angle                    | rad                |
| $\theta$             | angle                               | rad                |
| $\theta$             | some solution                       |                    |
| $\theta_0$           | equilibrium angle                   | rad                |
| $\theta_c$           | critical angle                      | rad                |
| $\theta_e$           | angle between electron and photon   | rad                |
| $\lambda$            | wavelength                          | m                  |
| $\pi$                | surface pressure                    | mN m <sup>-1</sup> |
| $\sigma$             | distance at zero                    | m                  |
| $\sigma_{coh}$       | coherent scattering cross-section   | m                  |
| $\sigma_{incoh}$     | incoherent scattering cross-section | m                  |
| $\sigma_{n,n+1}$     | interfacial roughness               | m                  |

|                      |                                   |          |
|----------------------|-----------------------------------|----------|
| $\phi$               | scattering angle                  |          |
| $\phi$               | dihedral angle                    | rad      |
| $\phi_g$             | global acceleration coefficient   |          |
| $\phi_i$             | volume fraction of solvation      |          |
| $\phi_p$             | personal acceleration coefficient |          |
| $\chi^2$             | a figure of merit                 |          |
| $\psi_0$             | wave at point 0                   |          |
| $\psi_i$             | wave at point $i$                 |          |
| $\psi_f$             | final scattered wave              |          |
| $\omega$             | frequency                         | $s^{-1}$ |
| $\omega$             | initial weight                    |          |
| $\omega_i$           | incident frequency                | $s^{-1}$ |
| $\omega_f$           | final frequency                   | $s^{-1}$ |
| $\Delta t$           | timestep                          | s        |
| $\Theta$             | new solution                      |          |
| $\Phi$               | incident flux                     |          |
| $d\sigma(q)/d\Omega$ | differential cross-section        |          |







## ABSTRACT

This work uses different coarse-graining methodologies to assist in the analysis of scattering measurements from soft matter aggregates, such as surfactant monolayers and micelles. The term “coarse-graining” is used broadly in this work; to describe the coarse-graining of a classical simulation potential model, a surfactant monolayer model that consists two layers, and a severe coarse-graining methodology that describes a surfactant molecule as just a position and direction in space. In all three cases, the aim of the applied coarse-graining is to improve the methods of analysis that may be performed for reflectometry and small angle scattering measurements.

A surfactant monolayer model, that was considerate of the chemical bonds between the surfactant heads and tails, was developed for the analysis of reflectometry measurements from a phospholipid monolayer at the interface between air and a deep eutectic solvent. This model allowed for a unique insight into the structure of the monolayer at the given interface. To assess the utility of coarse-grained potential models in the analysis of neutron reflectometry, a phospholipid monolayer was simulated using three different potential models, of different particle grain-size. This allowed for a better understanding of the simulation resolution necessary to accurately and successfully apply simulation-driven analysis to reflectometry. Finally, a severely coarse-grained description of a surfactant was used in a particle swarm optimisation to try and develop a starting structure for a multiple micelle simulation, where the experimental scattering profile was the optimisation target.

Alongside the development and application of coarse-graining methodologies, the final chapter of this work describes the development of open-source software and teaching materials for the introduction of classical molecular simulation. These educational resources introduce scattering users to simulation and its utility in scattering analysis, to enable a future where simulation driven analysis may be performed accurately by all.



# 1

## INTRODUCTION

The aim of this work is to investigate a series of different coarse-graining methodologies that can be used in the analysis of scattering data from soft matter. The coarse-graining used in this work includes traditional potential model coarse-graining,<sup>1</sup> applying a layer-based coarse-grained description of a surfactant monolayer to the analysis of reflectometry, and developing a severely coarse-grained description of a surfactant molecule to allow the easy generation of realistic starting structures<sup>2</sup> from experimental data. From this work, I hope that those interested in using coarse-grained approaches for the analysis of scattering data will be able to better understand the avenues that are open to them and the possible pitfalls that are present.

<sup>1</sup> Such as the use of the MARTINI potential model.

<sup>2</sup> For use in classical simulation.

### 1.1 SOFT MATTER

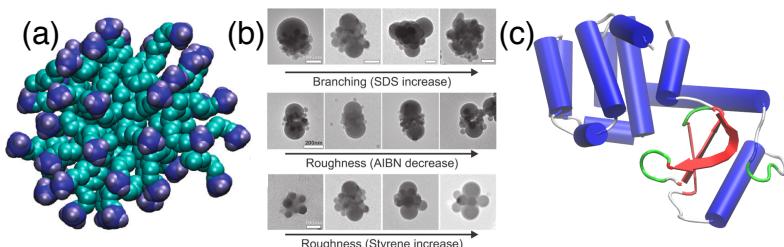
Soft matter is an umbrella term for many different types of material. These include micelles; sub-micron sized, dynamic agglomerates of amphiphilic molecules such as surfactants or block copolymers, colloidal solutions; where the interaction between the colloids may be controlled through chemical modification, or proteins; where the polar nature of different amino acids leads to the protein folding into a highly organised, and biologically relevant, shape. Some examples of these soft matter systems are shown in Figure 1.1. These species, initially, appear rather disparate, however, there are a few important commonalities among soft matter systems<sup>3</sup>:

- the lengths scales are intermediate between atomistic and macroscopic,<sup>4</sup>
- for soft matter systems the energy of a structural distortion is similar to thermal energy, so the material is in constant flux,
- this thermal motion can lead to the formation of complex, hierarchical structures due to the balance between enthalpy and entropy, this process is referred to as self-assembly.

<sup>3</sup> R. A. L. Jones. *Soft Condensed Matter*. 2002.

<sup>4</sup> typically  $1 \times 10^{-8}$ - $1 \times 10^{-5}$  m.

Figure 1.1: Three examples of soft matter species; (a) a 43 C<sub>10</sub>TAB surfactant micelle, reprinted with permission from R. Hargreaves *et al.* *J. Am. Chem. Soc.* 133.41 (2011), pp. 16524–16536, copyright 2011 American Chemical Society, (b) the tunable interactions of colloids, reprinted with permission from D. J. Kraft *et al.* *J. Phys. Chem. B* 115.22 (2011), pp. 7175–7181, copyright 2011 American Chemical Society, and (c) generated using VMD (W. Humphrey *et al.* *J. Mol. Graph.* 14.1 (1996), pp. 33–38) from the crystal structure of T4-lysozyme D. R. Rose *et al.* *Protein Eng. Des. Sel.* 2.4 (1988), pp. 277–282.



<sup>5</sup> L. L. Schramm *et al.* *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* 99 (2003), pp. 3–48.

<sup>6</sup> K. Simons *et al.* *Nat. Rev. Mol. Cell Biol.* 1 (2000), pp. 31–39.

<sup>7</sup> J. Israelachvili. *Intermolecular and Surface Forces*. 2011.

<sup>8</sup> D. Schmaljohann. *Adv. Drug Deliv. Rev.* 58.15 (2006), pp. 1655–1670. M. Sammalkorpi *et al.* *J. Phys. Chem. B* 113.17 (2009), pp. 5863–5870.

<sup>9</sup> M. J. Rosen *et al.* *Surfactants and Interfacial Phenomena*. 2012.

<sup>10</sup> J. Goodwin. *Colloids and Interfaces with Surfactants and Polymers*. 2009.

<sup>11</sup> Goodwin, see n. 10.

Soft matter self-assembly is the ability for soft matter systems to form organised structures in solution. These are of particular interest industrially, where surfactant and polymer self-assembly play an import role in food, commodity, and speciality chemicals.<sup>5</sup> Self-assembly processes are important from a biological perspective as it is phospholipids, a family of surface-active biomolecules, which make up the bilayers that protect cells.<sup>6</sup> The structures that result from the self-assembly of soft matter species have fluid-like properties. This is due to the fact that the subunits are held together by weak forces such as the van der Waals, hydrophobic, hydrogen-bonding, and screened electrostatic interactions.<sup>7</sup> This means that the structure of a self-assembled species is susceptible to changes in the local chemical environment, such as pH or salt concentration.<sup>8</sup>

The focus of this work is on the self-assembly of surfactant molecules. Surfactant is a general term for any molecule which is “surface-active”, to say that they will interact at an interface.<sup>9</sup> Surfactants are generally made up of two components; one part is highly soluble in one of the interfacial phases, while the other is not.<sup>10</sup> Usually, surfactants consist of a hydrocarbon tail, which is hydrophobic, and some hydrophilic head group, which can be ionic or non-ionic. When surfactants are present in water, the two components will interact differently with the solvent. A hydration sphere of water molecules will form around the hydrophilic head group, effectively allowing the head group to take part in the water’s hydrogen-bonding network. Whereas, the lyophilic tail has a structure-breaking effect on the hydrogen bonding network, termed the “hydrophobic effect”. The free energy deficit of this structure-breaking can be reduced through the aggregation of these hydrophobic groups, as the van der Waals attraction between tail groups is larger than that present between tail groups and water molecules. There is a decrease in entropy from the tail organisation, however, this is offset by the entropic increase from the water structure breakup. Finally, by considering the effect of the, often charged, head groups being close together, it is thought that the majority of the charge can be screened by the presence of a counter-ion, or water molecules, bound to the head group.<sup>11</sup> This means that at low concentrations, where it is statistically unlikely

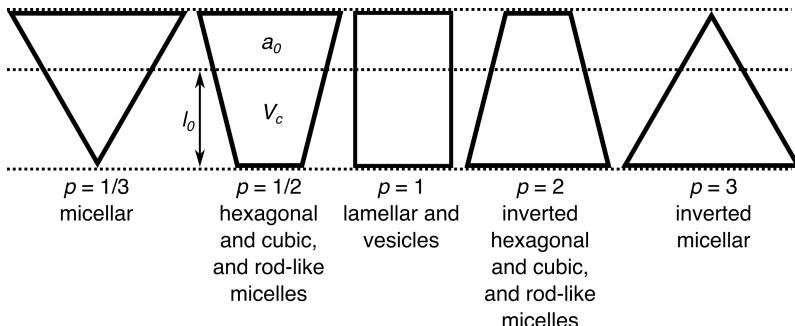


Figure 1.2: A graphical representation of the packing parameters and information of the resulting self-assembled structure.

for an agglomerate to form, there will be a higher concentration of surfactants at the air-water interface, and as the concentration is increased, assuming the system is above the Krafft temperature,<sup>12</sup> organised structures will begin to appear in solution.

The structures that can be formed from surfactant solutions are diverse; featuring micellar, hexagonal, cubic, and lamellar mesophases. These mesophases have a significant impact on the macroscopic properties of the system, for example, the liquid crystalline hexagonal phase can present interesting viscoelastic behaviour.<sup>13</sup> The mesophase that is formed is dependent on the shape of the underlying surfactants. Israelachvili described this dependency in terms of the dimensionless surfactant packing parameter,  $p$ .<sup>14</sup>

$$p = \frac{V_c}{a_0 l_0}, \quad (1.1)$$

where,  $V_c$  is the volume of the hydrophobic tail,  $l_0$  is the length of the tail, and  $a_0$  is the optimum head group area. This parameter can be used to estimate the geometry of the resulting self-assembled structure, detailed in Figure 1.2. It is important to note that the optimum head group area accounts for the hydration sphere of the head group. A short tail surfactant<sup>15</sup> will have a very small packing parameter resulting in small spherical micelles. Whereas, the twin-tailed phospholipids, such as 1,2-dipalmitoyl-sn-glycero-3-phosphocholine,<sup>16</sup> will have a much larger packing parameter due to the larger tail volume and length, therefore this surfactant will form a lamellar bilayer in solution.

This work will focus on the investigation of surfactant monolayers and micellar systems. These represent interesting model systems of significant relevance both technologically<sup>17</sup> and biologically.<sup>18</sup> Both of these systems are regularly investigated using X-ray and neutron elastic scattering techniques, with analysis performed in a model-dependent fashion.<sup>19</sup>

## 1.2 ANALYSIS OF SOFT MATTER SCATTERING

The use of neutron and X-ray scattering experiments for the study of soft matter is well developed, with early research into the struc-

<sup>12</sup> The lowest temperature at which agglomerates will form.

<sup>13</sup> D. Jurašin *et al.* *Soft Matter* 9.12 (2013), p. 3349; F. Cordobés *et al.* *J. Colloid Interface Sci.* 187.2 (1997), pp. 401–417.

<sup>14</sup> Israelachvili, see n. 7.

<sup>15</sup> Such as *n*-decyltrimethylammonium bromide.

<sup>16</sup> Known as DPPC.

<sup>17</sup> N. Anton *et al.* *Int. J. Pharm.* 398.1-2 (2010), pp. 204–209; M. Zagnoni. *Lab on a Chip* 12.6 (2012), p. 1026.

<sup>18</sup> K. Kataoka *et al.* *Adv. Drug Deliv. Rev.* 64 (2012), pp. 37–48; H. Mohwald. *Annu. Rev. Phys. Chem.* 41 (1990), pp. 441–476; S. Kewalramani *et al.* *J. Phys. Chem. Lett.* 1.2 (2010), pp. 489–495.

<sup>19</sup> E. Pambou *et al.* *Langmuir* 31.36 (2015), pp. 9781–9789; D. W. Hayward *et al.* *Macromolecules* 48.5 (2015), pp. 1579–1591; I. Rodriguez-Loureiro *et al.* *Soft Matter* 13.34 (2017), pp. 5767–5777; G. Hazell *et al.* *J. Colloid Interface Sci.* 474 (2016), pp. 190–198.

<sup>20</sup> O. Albrecht *et al.* *J. Phys. France* 39.3 (1978), pp. 301–313.

<sup>21</sup> O Kratky *et al.* *J. Colloid Sci.* 4.1 (1949), pp. 35–70.

<sup>22</sup> Generally abbreviated to SAS; with SAXS indicating the use of X-rays and SANS neutrons

<sup>23</sup> G. V. Jensen *et al.* *Angew. Chemie Int. Ed.* 53.43 (2014), pp. 11524–11528.

<sup>24</sup> V. Rondelli *et al.* *J. Phys. Conf. Ser.* 340 (2012), p. 012083.

<sup>25</sup> P. A. Hassan *et al.* *J. Colloid Interface Sci.* 257.1 (2003), pp. 154–162, see Section 2.2.7.

<sup>26</sup> R. A. Campbell *et al.* *J. Colloid Interface Sci.* 531 (2018), pp. 98–108; J. R. Lu *et al.* *Acta Crystallogr. A* 52.1 (1996), pp. 11–41, see Section 2.2.7.

<sup>27</sup> J. S. Pedersen. “Monte Carlo Simulation Techniques Applied in the Analysis of Small-Angle Scattering Data from Colloids and Polymer System”. In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter.* 2002, pp. 381–390.

<sup>28</sup> M. Wormington *et al.* *Philos. Trans. R. Soc. London Ser. A* 357.1761 (1999), pp. 2827–2848, abbreviated to DE.

<sup>29</sup> A. R. J. Nelson *et al.* *J. Appl. Crystallogr.* 52.1 (2019), pp. 193–200; A. H. Larsen *et al.* *J. Appl. Crystallogr.* 51.4 (2018), pp. 1151–1161.

<sup>30</sup> Such as molecular dynamics (MD).

<sup>31</sup> E. Scoppola *et al.* *Curr. Opin. Colloid Interface Sci.* 37 (2018), pp. 88–100.

<sup>32</sup> The historical context of this is discussed briefly in Chapter 5.

<sup>33</sup> S. Bowerman *et al.* *J. Chem. Theory Comput.* 13.6 (2017), pp. 2418–2429.

<sup>34</sup> In the form of Empirical Potential Structure Refinement.

<sup>35</sup> R. Hargreaves *et al.* *J. Am. Chem. Soc.* 133.41 (2011), pp. 16524–16536.

<sup>36</sup> M. T. Ivanović *et al.* *Angew. Chemie Int. Ed.* 57.20 (2018), pp. 5635–5639.

<sup>37</sup> A. F. Miller *et al.* *Mol. Phys.* 101.8 (2003), pp. 1131–1138; P. M. Anderson *et al.* *J. Chem. Phys.* 121.17 (2004), p. 8503.

<sup>38</sup> Monte Carlo and MD respectively.

<sup>39</sup> Abbreviated to NR.

<sup>40</sup> A. P. Dabkowska *et al.* *Langmuir* 30.29 (2014), pp. 8803–8811.

ture of phospholipid monolayers by reflectometry methods being conducted in the late 1970s by Albrecht *et al.*<sup>20</sup> While, the work of Kratky and Porod,<sup>21</sup> who used small angle X-ray scattering<sup>22</sup> for the study of colloidal systems was published in 1949. Since these early works, instrumentation developments have enabled more challenging experiments to be conducted, such as time-resolved studies<sup>23</sup> and the study of floating phospholipid bilayers.<sup>24</sup>

However, the analysis of soft matter scattering has changed little since these early works, still typically involving the use of very coarse models. These include the shape-based modelling common in SAS<sup>25</sup> and reflectometry analysis.<sup>26</sup> More sophisticated model refinements have been developed, such as the use of Monte-Carlo sampling,<sup>27</sup> differential evolution optimisation,<sup>28</sup> and Bayesian inference.<sup>29</sup> However, there has been little change in the definition of the models that unpin the analysis processes. Recently, there have been movements towards the use of atomistic modelling techniques<sup>30</sup> to augment, and assist, the analysis of soft matter scattering measurements, in a multi-modal approach.<sup>31</sup>

Much of the work relating to the use of atomistic simulation for the analysis of SAS measurements has been focused on the study of protein molecules in solution.<sup>32</sup> This has allowed for more profound understanding aspects of biology such as the conformational states available to protein molecules in solution.<sup>33</sup> The uptake of atomistic simulation for the analysis of SAS from systems such as micelles has been slower, in part due to the more complex conformation landscape available to these systems under standard conditions. However, the work of Hargreaves *et al.* paired atomistic simulation with total scattering measurements<sup>34</sup> to resolve the structure of a simple short-tail surfactant micelle.<sup>35</sup> Furthermore, the work of Ivanović *et al.* used scattering experiments to refine the output of MD simulations of micelles of a pre-defined size.<sup>36</sup> Both of these examples required significant computational resource; in the former case, the computational time taken was quoted as 200 days, while the later required the running of multiple simulations at different micelle sizes in order to determine the appropriate simulation.

The use of atomistic simulation for the analysis of reflectometry measurements of soft matter systems began with the work of Miller *et al.* and Anderson and Wilson,<sup>37</sup> where atomistic simulations<sup>38</sup> were used to study polymer self-assembly at the oil-water interface. These simulation trajectories were then compared with experimental neutron reflectometry<sup>39</sup> measurements. Dabkowska *et al.* also used atomistic simulation and NR measurements to study the structure of a surfactant monolayer at the air-water interface, providing the first example of a direct comparison between experimental reflectometry data and that determined from simulation.<sup>40</sup> To date, there is only one work that has used coarse-grained MD simulation to aid in the analysis of NR, this is the work of Kout-

sioubas.<sup>41</sup> This work made use of the MARTINI coarse-grained potential model to simulate a phospholipid bilayer and was compared with experimental NR measurements.

<sup>41</sup> A. Koutsoubas. *J. Phys. Chem. B* 120.44 (2016), pp. 11474–11483.

### 1.3 COARSE-GRAINING OF SOFT MATTER SYSTEMS

The characteristic non-atomistic length scales associated with soft matter systems make them ideal for the application of coarse-graining protocols. Coarse-graining is where the dimensionality of a problem is reduced by the removal of certain degrees of freedom from a set. The most common method of coarse-graining is the re-parameterisation of an atomistic potential model in terms of this reduced parameter space. An example of this is the MARTINI potential model,<sup>42</sup> where the aim is to reparameterise the system without significant loss of chemical information.<sup>43</sup> A result of coarse-graining is the creation of a flatter potential energy landscape, as shown in Figure 2.20. The availability of coarse-grained potential models and tools for coarse-graining has allowed for very large simulations to be performed, such as the simulation of large polymeric species,<sup>44</sup> phospholipid nanodiscs,<sup>45</sup> and realistic biomembranes.<sup>46</sup> Furthermore, these coarse-grained potential model have shown the ability to reproduce experimental measurements.<sup>47</sup>

However, in this work I have used the term “coarse-graining” broadly to include the applications of a chemically-consistent coarse-grained monolayer model for the analysis of reflectometry data,<sup>48</sup> In addition to the assessment of different atomistic and coarse-grained potential models for the analysis of NR, building on the work of Dabkowska *et al.* and Koutsoubas.<sup>49</sup>

### 1.4 OPTIMISATION METHODOLOGIES

The availability of high-performance computing has increased significantly in recent years, in particular, due to cloud-based infrastructures. Furthermore, highly parallelisable optimisation algorithms are now available such as the particle swarm<sup>50</sup> and DE<sup>51</sup> optimisations. As mentioned above, previous work has shown that the simulation of a surfactant micelle and comparison with experimental data requires significant computational expense<sup>52</sup> In the interest of reducing this, and improving the applicability of high-performance computing to the simulation-driven analysis of SAS I have investigated the use of a particle swarm optimisation<sup>53</sup> to produce a realistic, near-atomistic micelle structure based on experimental data alone. This has made use of a coarse-grained description of a surfactant molecule on two levels; one for the PSO and another for the scattering profile calculation.<sup>54</sup>

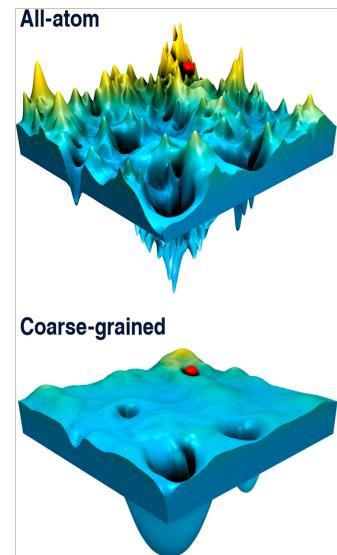


Figure 1.3: Potential energy surfaces for an all-atom vs a coarse-grained potential model, reprinted with permission of the American Chemical Society from S. Kmiecik *et al.* *Chem. Rev.* 116.14 (2016), pp. 7898–7936.

<sup>42</sup> This specific model is discussed in greater detail in Section 2.3.3.

<sup>43</sup> S. J. Marrink *et al.* *J. Phys. Chem. B* 111.27 (2007), pp. 7812–7824.

<sup>44</sup> P. Carbone *et al.* *J. Chem. Phys.* 128.6 (2008), p. 064904.

<sup>45</sup> M. Xue *et al.* *Biophys. J.* 115.3 (2018), pp. 494–502.

<sup>46</sup> S. J. Marrink *et al.* *Chem. Rev.* 119.9 (2019), pp. 6184–6226.

<sup>47</sup> E. Negro *et al.* *Soft Matter* 10.43 (2014), pp. 8685–8697; S. Nawaz *et al.* *Soft Matter* 8.25 (2012), p. 6744.

<sup>48</sup> This work is the focus of Chapter 3.

<sup>49</sup> Dabkowska *et al.*, see n. 40; Koutsoubas, see n. 41, see Chapter 4.

<sup>50</sup> J. Kennedy *et al.* In: *Proceedings of ICNN'95*. International Conference on Neural Networks. Perth, AU, 1995, pp. 1942–1948; Y. Shi *et al.* In: *1998 IEEE International Conference on Evolutionary Computation Proceedings*. IEEE World Congress on Computational Intelligence. Anchorage, US, 1998, pp. 69–73.

<sup>51</sup> R. Storn *et al.* *J. Global Optim.* 11 (1997), pp. 341–359.

<sup>52</sup> Hargreaves *et al.*, see n. 35; Ivanović *et al.*, see n. 36.

<sup>53</sup> Abbreviated to PSO.

<sup>54</sup> See Chapter 5.

## 1.5 EDUCATIONAL MATERIALS

While the development of analytical methods and infrastructure are important for the development and uptake of the simulation-driven analysis methods applied in this work, the development of informative resources is also necessary. Current experimental users of SAS are typically not familiar with detailed aspects of classical simulation, however, they are often interested in applying it to assist with their analyses. Therefore, alongside traditional research applications, I have been developing open educational resources<sup>55</sup> designed to introduce classical simulation techniques, and to allow users of scattering techniques to become familiar with these. The ambition being that as the availability of simulation-driven analysis for SAS grows, so will the user base that is familiar with the underlying methods.<sup>56</sup>

<sup>55</sup> Abbreviated to OERs.

<sup>56</sup> See Chapter 6.

## 2

# THEORY

### 2.1 PROBING RADIATION

This work is focussed on the use of X-ray and neutron scattering to probe soft matter systems; in particular surfactant monolayers and micelles. Therefore, it is pertinent to discuss how each of these probing radiation is produced and detail the advantages of each with respect to the other.

#### 2.1.1 Generation of X-rays

X-rays are a form of electromagnetic radiation similar to visible light, albeit with a much shorter wavelength.<sup>1</sup> There are four common ways to produce X-rays; three are available within the laboratory, while the other is exclusive to large scale facilities.

The three laboratory source X-ray generation techniques are the X-ray tube, the rotating anode, and the liquid jet. An X-ray tube consists of a filament and an anode within a vacuum chamber, by passing a high voltage electrical current across the filament electrons are emitted which accelerate towards the anode. On collision with the anode, the rapid deceleration results in the emission of X-rays at a characteristic wavelength based on the anode material.<sup>2</sup> The most common material for an X-ray tube anode is copper which gives off radiation with an energy of  $\sim 8 \text{ keV}$ .<sup>3</sup>

Another common laboratory method for the generation of X-rays is the rotating anode.<sup>4</sup> In the X-ray tube, each time that an electron collides with anode there is some energy transfer, this means that over many millions of collisions the temperature of the anode can rise significantly, which can cause the anode material to melt. Resulting in a temperature-based limitation to the available X-ray flux. This lead to the development of the rotating anode, which is simply where the anode is made from a rotating wheel, such that the bombardment is spread across the whole wheel reducing the energy localisation. The use of a rotating anode can allow for an increase in the photon flux by about an order of magnitude.<sup>5</sup>

The final laboratory method for X-ray generation is the liquid jet source.<sup>6</sup> For the liquid jet X-ray source, an electron beam is

<sup>1</sup> Typically 0.01-10 nm.

<sup>2</sup> H. Schnablegger *et al.* *The SAXS Guide: Getting Acquainted with the Principles*. 2017.

<sup>3</sup> This has a wavelength of  $\sim 1.5 \text{ nm}$ .

<sup>4</sup> Essentially an improvement on the X-ray tube.

<sup>5</sup> Schnablegger *et al.*, see n. 2.

<sup>6</sup> MetalJet X-Ray Source Technology.  
URL: <https://www.excillum.com/technology/> (Accessed 2018-12-6), Branded MetalJet by excillum.

Figure 2.1: A schematic representation of a synchrotron radiation source, identifying the Linac, the booster ring, the radio-frequency cavities (rf), the bending magnet (BM) and the insertion device (ID). Reprinted, with permission of Springer Nature Customer Service Centre GmbH: Springer Nature, from M. C. Garcia-Gutierrez *et al.* "Bases of Synchrotron Radiation, Light Sources, and Features of X-Ray Scattering Beamlines". In: *Applications of Synchrotron Light to Scattering and Diffraction in Materials and Life Sciences*. 2009, pp. 1–22.



<sup>7</sup> Usually an gallium or indium alloy.

incident on a liquid metal sample,<sup>7</sup> rather than traditional solid metal, which can dissipate heat more efficiently. This means that the electron intensity and therefore X-ray brightness, available to the liquid jet source is much greater than a rotating anode source.

The method of X-ray generation that is not available in a typical laboratory is at a synchrotron, the use of this method has the drawback that it requires access to a national or international facility.<sup>8</sup> The way in which X-rays are generated at a synchrotron involves the acceleration of an electron, rather than the deceleration as with the laboratory sources. This is achieved by having relativistic electrons travel on a curve, from Newtonian mechanics it is known that travelling on a curve at constant speed is equivalent to acceleration. First the electrons are accelerated, after being produced in a linear accelerator, to near the speed of light in a booster synchrotron before injecting them into the storage ring. In the storage ring, the electrons are kept at relativistic speeds with bending magnets and straight sections making up a ring as shown in Figure 2.1. How circular the ring is depends on the number of bending magnets that make it up; for example, DLS had 48 bending magnets with 48 straight sections at the time of construction.

When an electron accelerates (or travels on a curve), Cherenkov radiation is emitted in accordance with the Cherenkov relation,

$$n_i \beta_c \cos \theta_e = 1, \quad (2.1)$$

where,  $n_i$  is the refractive index for the dielectric medium,  $\beta_c$  is the fraction of the speed of light at which that electron is travelling, and  $\theta_e$  is the angle between the electron trajectory and the trajectory of the resulting photon.<sup>9</sup> The curve is the result of a bending magnet, meaning that at each bending magnet there can be a beamline which uses the synchrotron light. The light that is given off from a bending magnet is continuous and broad, covering a wide range of the electromagnetic spectrum. The alternative to a bending magnet beamline is that which is served by an insertion device.

<sup>8</sup> M. C. Garcia-Gutierrez *et al.* "Bases of Synchrotron Radiation, Light Sources, and Features of X-Ray Scattering Beamlines". In: *Applications of Synchrotron Light to Scattering and Diffraction in Materials and Life Sciences*. 2009, pp. 1–22.

An insertion device is able to offer more specific radiation characteristics (photon energy, narrower band) than a bending magnet, and are placed on the straight sections of the synchrotron. Common insertion devices include wavelength shifters, wigglers, and undulators.

The type of insertion device that is present at both I07 and I22 at DLS is an undulator. An undulator consists of a series of magnets of opposing polarity that causes the electrons to ‘wiggle’ back and forth as shown in Figure 2.2. This results in a superposition of radiation from  $N_P$  sources, where  $N_P$  is the number of magnets, yielding quasi-monochromatic radiation. The brilliance of different X-ray sources are compared in Table 2.1, this shows the significant benefit that an undulator can offer in terms of photon brilliance.

### 2.1.2 Generation of neutrons

Neutrons hold an advantage over X-rays, particularly for application to the study of soft matter, due to the ability to use contrast variation to increase the quantity of information from the sample.<sup>10</sup> However, neutrons cannot be produced safely on a laboratory scale, therefore it is always necessary to visit large scale facilities to harness neutrons for scattering experiments. These facilities come in two flavours; the reactor source and the spallation source, each offering unique benefits.

Neutron reactor sources<sup>11</sup> are currently the most common format of neutron source and are capable of producing the highest average neutron flux.<sup>12</sup> The High-Flux Reactor at the ILL is capable of producing a neutron flux of  $1.5 \times 10^{15}$  neutrons  $s^{-1}cm^{-2}$ .<sup>13</sup> A reactor source operates on the principle of nuclear fission, where an atomic nucleus is capable of breaking down into smaller nuclei, overcoming the strong nuclear force. This often involves using uranium enriched with its fissile isotope,  $^{235}U$ , which after the initial absorption of a stray neutron<sup>14</sup> will undergo fission to release, on average, 2.5 daughter neutrons, an example of a possible uranium fission mechanism is:



This type of mechanism is the basis for both research, and nuclear power reactors.<sup>15</sup> One of the major drawbacks for reactor neutron sources is the perceived public opinion towards such facilities. Major safety concerns, such as “nuclear meltdown” and the resulting nuclear waste, mean that reactor sources are often unpopular and therefore struggle to obtain funding required for operation.

The other form of neutron source is a spallation source, this is much less controversial as it does not require fissile materials and hence there is no risk of a nuclear disaster. The ISIS Neutron and Muon Source is an example of a spallation source, where high energy protons, 800 MeV,<sup>16</sup> are accelerated towards a tungsten target.

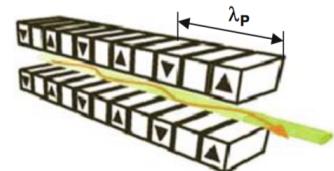


Figure 2.2: A diagram of an undulator insertion device, such as that on I07 and I22, where  $\lambda_P$  is the period length between opposing magnets. Reprinted, with permission of Springer Nature Customer Service Centre GmbH: Springer Nature, from Garcia-Gutierrez et al., “Bases of Synchrotron Radiation, Light Sources, and Features of X-Ray Scattering Beamlines”.

<sup>10</sup> This is discussed in detail in Section 2.1.3.

<sup>11</sup> Such as the Institut Laue-Langevin (ILL) in Grenoble, France.

<sup>12</sup> The number of neutrons per second per unit area.

<sup>13</sup> ILL: *Neutron for Science: Technical Characteristics*. URL: <https://www.ill.eu/reactor-environment-safety/high-flux-reactor/technical-characteristics/> (Accessed 2016-8-8).

<sup>14</sup> Arising from a cosmic ray, or spontaneous fission.

<sup>15</sup> D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.

<sup>16</sup> ISIS – How ISIS Works. URL: <https://www.isis.stfc.ac.uk/Pages/What-does-ISIS-Neutron-Muon-Source-do.aspx> (Accessed 2018-9-25).

Table 2.1: A comparision of the photon brilliance from different light sources.  
Adapted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users.* 2011.

| Light source   | Approximate brilliance/<br>photons s <sup>-1</sup> mrad <sup>-2</sup> 0.1%bandwidth <sup>-1</sup> |
|----------------|---|
| Candle         | $1 \times 10^5$   |
| X-ray tube     | $1 \times 10^8$   |
| Sun            | $1 \times 10^{10}$  |
| Bending magnet | $1 \times 10^{15}$  |
| Undulator      | $1 \times 10^{20}$  |

When the protons strike the target, they can cause the release of a series of neutrons, the first batch of neutrons are given off with too high an energy to be useful, however, less excited neutrons are given off by secondary emissions. In addition to the public perception benefit, spallation sources also offer a technological advantage in the time-of-flight<sup>17</sup> technique. The ToF technique relies the fact that at a spallation source, it is possible to know the time at which the neutron was ejected from the target to a high level of precision and therefore it is possible to measure the time taken for the neutron to reach the instrument. Since the neutron is a particle of a finite mass,  $m$ , it is possible to correlate the velocity,  $\mathbf{v}$ , of the particle with the kinetic energy,  $E_k$ ,

$$E_k = \frac{m\mathbf{v}^2}{2}, \quad (2.2)$$

and with knowledge of the energy of the particle, its wavelength,  $\lambda$ , can be determined by the de Broglie relation,<sup>18</sup>

$$E = h\omega = \frac{h\mathbf{v}}{\lambda}, \quad (2.3)$$

where,  $h$  is Planck constant and  $\omega$  is the neutron frequency. Therefore, the wavelength of the neutron is proportional to the inverse of the particle's velocity, and hence the time-of-flight,  $t_F$ ,

$$\lambda = \frac{h}{m\mathbf{v}} = \frac{ht_F}{mL_F}, \quad (2.4)$$

where,  $L_F$  is the distance between the target and the instrument. The fact that the neutrons can spread out in the flight from the target means that wavelength-dispersive techniques, where the neutron wavelength is measured rather than the scattering angle, are possible at spallation sources which cannot be carried out natively at reactor sources. The weakness of current spallation sources is that they have a lower average flux than reactor sources, however, the construction of the European Spallation Source will change this as it offers an average flux similar to that of a reactor source with the benefits of the spallation technique.

A problem that is inherent for both reactor and spallation sources is that the energy of the neutrons given off is usually too high to be used to study condensed materials, such as soft matter. This means that moderation must be used to reduce the energy of the neutrons passing through the sample. The neutrons that are

<sup>17</sup> Abbreviated to ToF.

<sup>18</sup> L. de Broglie. *Ann. Phys. (Paris)* 10.3 (1925), pp. 22–125.

considered to be optimal for the study of condensed materials are thermal in nature, named because their energy is approximately that of ambient temperature. Thermal neutrons are achieved by allowing the neutrons to pass through a large volume of moderator material; usually, graphite, D<sub>2</sub>O, methane or H<sub>2</sub>, stored at 300 K before they reach the instrument.<sup>19</sup>

<sup>19</sup> Sivia, see n. 15.

### 2.1.3 Contrast variation

The scattering profile generated by the interaction of some system with radiation depends on three factors:

- the spatial arrangement of the atoms in the system,
- the instrument being used to measure the pattern; instrumental resolution function, and
- the interaction between the radiation and the matter under investigation.

This final factor is perhaps better known as the “scattering contrast”, this is an extremely important factor in the study of soft matter, particularly when the probing radiation is the neutron. The scattering contrast makes it possible to select individual components of the system and investigate their structural properties.<sup>20</sup> The differential cross-section,  $d\sigma/d\Omega$  of a point scatterer, as shown in Equation 2.21, varies only with respect to the scattering length of the species,  $b$ ,

$$\frac{d\sigma}{d\Omega} \propto b^2. \quad (2.5)$$

However, as discussed in Section 2.2.4, it is often easier to use the scattering length density.<sup>21</sup>

When an X-ray interacts with an atom, it is scattered by the interaction with the electrons, this is due to the X-ray being a form of electromagnetic radiation. Furthermore, it means that the scattering length of an atom by an X-ray is directly proportional to the number of electrons in the atom, so it is therefore it is difficult to discern between the scattering from a carbon atom (6 electrons) and a nitrogen atom.<sup>22</sup> Additionally, the scattering from hydrogen atoms is practically non-existent.

The scattering length of a neutron by an atom varies unsystematically with respect to the atomic number of a species, this is shown in Figure 2.3. In addition to the apparently random variation with changes in atomic number, there is also significant variation with mass number.<sup>23</sup> This is also dependent on to the magnetic state of the atom, however, this is normally unimportant for soft matter. The scattering lengths differ with the nuclear spin energy level, this leads to an average scattering length,  $\langle b \rangle$ , for isotopes where the nuclear spin is non-zero ( $S \neq 0$ ). There are two forms of scattering, coherent and incoherent, for which the

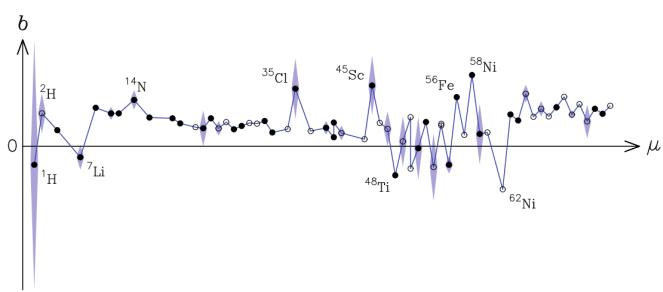
<sup>20</sup> P. Schurtenberger. “Contrast and Contrast Variation in Neutron, x-Ray and Light Scattering”. In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*. 2002, pp. 145–170.

<sup>21</sup> Abbreviated to SLD.

<sup>22</sup> As there is a difference of just a single electron between these atoms.

<sup>23</sup> Leading to variation between isotopes of the same atom.

Figure 2.3: The variation of the average neutron scattering length,  $\langle b \rangle$  (circles), with atomic mass,  $\mu$ . The standard deviation,  $\Delta b$ , is indicated with the shaded regions. Reprinted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.



scattering cross-sections,  $\sigma$ , are determined by,

$$\begin{aligned}\sigma_{\text{coh}} &= 4\pi\langle b \rangle^2 \\ \sigma_{\text{incoh}} &= 4\pi(\langle b^2 \rangle - \langle b \rangle^2)\end{aligned}\quad (2.6)$$

The coherent scattering is the scattering from nuclei that all have the same value of  $\langle b \rangle$ , and leads to the important scattering pattern. Whereas, the incoherent scattering is caused by the ‘disorder’ between the isotopes, and is the cause of the background present in the measurement. Examples of these scattering cross-sections for nuclei relevant to soft matter are shown in Table 2.2. It can be seen that the incoherent scattering from the  ${}^1\text{H}$  nuclei is more than forty times the coherent scattering. This leads to a large, intrusive background present in the scattering pattern of hydrogenous samples. The difference between the scattering of  ${}^1\text{H}$  and  ${}^2\text{H}$ , evident in Table 2.2, can lead to a very useful technique if soft matter scattering, known as contrast variation. The idea of contrast variation is based on the substitution of one isotope of an atom for another, while not introducing significant change to the properties of the material. Traditionally, the benefit of this came in terms of contrast matching out a part of the system to reduce the dimensionality of the problem for analysis. For example, by matching the solvent SLD to that of the tails of the surfactants at the centre of a micelle there would only be scattering from the heads, and conversely, there would only be scattering from the tails if the solvent had the same SLD as the head groups. This means that the problem becomes more straightforward as there are fewer variable parameters when fitting the data. This idea is represented graphically in Figure 2.4. The technique of contrast variation may also be used in terms of data analysis. By increasing the number of data sets corresponding to a single model at different contrasts, the solution for the true structure of the model from the scattering data becomes

Table 2.2: Examples of coherent and incoherent scattering cross-sections.  
Reprinted, with permission of Elsevier, from P. Schurtenberger. “Contrast and Contrast Variation in Neutron, x-Ray and Light Scattering”. In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*. 2002, pp. 145–170.

| Isotope           | $S$   | $\sigma_{\text{coh}}/10^{-28} \text{ m}$ | $\sigma_{\text{incoh}}/10^{-28} \text{ m}$ |
|-------------------|-------|--|--|
| ${}^1\text{H}$    | $1/2$ | 1.8                                      | 79.7                                       |
| ${}^2\text{H}$    | 1     | 5.6                                      | 2.0  |
| ${}^{12}\text{C}$ | 0     | 5.6                                      | —  |
| ${}^{14}\text{N}$ | 1     | 11.6                                     | 0.3  |
| ${}^{16}\text{O}$ | 0     | 4.2                                      | —  |

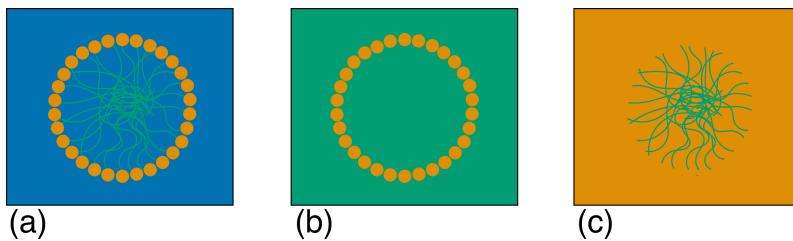


Figure 2.4: The effect of varying the SLD of the solvent in a micelle system, (a) the system in a pure solvent, (b) the solvent is contrast matched to the surfactant tails, and (c) the solvent is contrast matched to the surfactant heads.

more robust. This is due to the fact that each different contrasts can be considered as an independent measurement of the same system, and hence each set of scattering data can be used within the data analysis procedure to obtain the best global agreement to the experiment. This co-refinement of multiple experiments can, under the right conditions, be used to simultaneously consider both neutron and X-ray datasets.<sup>24</sup>

There is also the possibility of using contrast variation when the probing radiation is the X-ray, through the use of anomalous scattering. This is where different wavelengths of radiation give different scattering when the wavelengths are on opposite sides of an X-ray absorption edge. This is not frequently used for soft matter species, as the X-ray absorption edges for elements common in soft matter<sup>25</sup> are at very low X-ray energies so generally outside of the accessible range of a standard SAXS instrument.<sup>26</sup>

## 2.2 SCATTERING

The use of scattering techniques to probe soft condensed matter systems is commonplace. This work will focus on the use of small angle scattering<sup>27</sup> and reflectometry techniques. These are particularly appropriate for application to soft condensed matter systems due to the length scales capable of being probed being similar to the persistence length of the soft condensed matter systems. The length scale covered for such techniques is from around 1-300 nm, as is shown in Figure 2.5. The focus is on the equilibrium structure(s) of a material, and therefore there is no interest in the system dynamics, meaning that exclusively elastic scattering techniques may be used, where there is no energy transfer between the probing radiation and the material. This is in contrast to inelastic scattering where energy transfer occurs; facilitating the measurement of system dynamics, such as the dynamical modes of polymers and phospholipid bilayers<sup>28</sup>

Both X-ray and neutron scattering techniques are discussed and used in this work. From an experimental viewpoint, there are significant differences between an X-ray scattering and a neutron scattering experiment. However, there is little variation in terms of the data analysis, where the differences are limited to; the nature of the scattering lengths<sup>29</sup> and the higher background that is present

<sup>24</sup> A. R. J. Nelson. *J. Appl. Crystallogr.* 39.2 (2006), pp. 273–276.

<sup>25</sup> H, C, N, O, etc.

<sup>26</sup> Schurtenberger, see n. 20.

<sup>27</sup> SAS.

<sup>28</sup> V. Garcia Sakai *et al.* *Curr. Opin. Colloid Interface Sci.* 14.6 (2009), pp. 381–390; B. Farago. *Curr. Opin. Colloid Interface Sci.* 14.6 (2009), pp. 391–395.

<sup>29</sup> See Section 2.1.3.

Figure 2.5: A representation of how different techniques can be used to probe various length scales. Reprinted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.



in the neutron scattering experiments.

### 2.2.1 The scattering vector

<sup>30</sup> Not just X-rays and neutrons, but any wave.

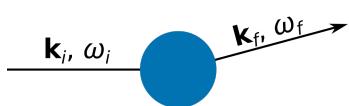


Figure 2.6: A schematic of the scattering of some probing radiation by a sample (blue circle). Adapted, with permission of Oxford University Press, from Sivia, *Elementary Scattering Theory: For X-Ray and Neutron Users*.

<sup>31</sup> Throughout this work, units of Å<sup>-1</sup> will be wherever possible.

The scattering of some probing radiation,<sup>30</sup> by some sample, can be represented as shown in Figure 2.6. Since only elastic scattering is being considered, there will be no change in the frequency of the radiation,  $\omega_i = \omega_f$ . This means that only the wavevector,  $\mathbf{k}$ , can change,  $\mathbf{k}_i \neq \mathbf{k}_f$ . The difference between the incident and final wavevectors is the scattering vector,  $\mathbf{q}$ , where,

$$\mathbf{q} = \mathbf{k}_i - \mathbf{k}_f. \quad (2.7)$$

The scattering vector strictly has units of  $\text{m}^{-1}$ , however it is often more practical to use  $\text{nm}^{-1}$  or  $\text{\AA}^{-1}$ .<sup>31</sup> Since the frequency of the probing radiation does not change during an elastic scattering event, the wavelength,  $\lambda$ , will also not change, meaning that the moduli of the incident and final wavevectors are,

$$|\mathbf{k}_i| = |\mathbf{k}_f| = \frac{2\pi}{\lambda}. \quad (2.8)$$

This also means that only the angle will change during the elastic scattering event. The vector diagram in Figure 2.7 can be used to describe the geometry of an elastic scattering event. From this, and



Figure 2.7: A vector diagram describing an elastic scattering event, where  $\mathbf{k}_i$  is the incident wavevector,  $\mathbf{k}_f$  is the final wavevector,  $2\theta$  is the scattering angle, and  $\mathbf{q}$  is the scattering vector. Adapted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.

Equation 2.8, the value of  $q$ , where  $q = |\mathbf{q}|$  can be shown as,

$$q = \frac{4\pi \sin \theta}{\lambda}. \quad (2.9)$$

However, this fails to fully capture the three dimensional nature of the scattering event. Hence, it is necessary to describe the scattering with spherical coordinates,  $2\theta$ , and  $\phi$ , such that the incoming and outgoing radiation can be described as,

$$\begin{aligned} \mathbf{k}_i &= \left(0, 0, \frac{2\pi}{\lambda}\right), \\ \mathbf{k}_f &= \frac{2\pi}{\lambda}(\sin 2\theta \cos \phi, \sin 2\theta \sin \phi, \cos 2\theta), \end{aligned} \quad (2.10)$$

where,  $|\mathbf{k}_f| = 2\pi/\lambda$ . This allows the scattering vector to be written,

$$\mathbf{q} = \frac{4\pi \sin \theta}{\lambda}(-\cos \theta \cos \phi, -\cos \theta \sin \phi, \sin \theta). \quad (2.11)$$

For an isotropic scattering pattern, it is the magnitude of the scattering vector,  $q$ , that is measured. In practical terms, the scattering vector allows for easy comparison of measurements made at different radiation wavelengths.

The basic quantity measured in a scattering experiment is the differential cross section,  $d\sigma(q)/d\Omega$ . This is the fraction of particles of probing radiation that is scattered with a particular set of polar coordinates,  $2\theta$  and  $\phi$ ,

$$\frac{d\sigma(q)}{d\Omega} = \frac{R_i(2\theta, \phi)}{NV\Phi\Delta\Omega}, \quad (2.12)$$

where,  $R_i(2\theta, \phi)$  is the rate of arrival of the scattered particles at the position  $2\theta, \phi$ ,  $V$  is the illuminated volume of the sample,  $\Phi$  is incident flux,  $\Delta\Omega$  is some small solid angle, and  $N$  is the number of scattering particles of interest, in the case of elastically scattered radiation,  $N = N\%_{el}$ , where  $\%_{el}$  is the fraction of elastically scattered radiation.

### 2.2.2 Scattering from a single fixed particle

It is possible to describe a steady stream X-ray photons or neutrons of wavelength,  $\lambda$ , travelling through space as follows,

$$\psi_i = \psi_o \exp(i k z), \quad (2.13)$$

where,  $z$  is the direction of travel,  $k = 2\pi/\lambda$ , and the incident flux is the magnitude of the wave squared,  $\Phi = |\psi_o|^2$ . This wave then interacts with a single fixed particle elastically, propagating the wave radially outwards, as shown in Figure 2.8. This propagation is centred on the atom, therefore the wavevector,  $\mathbf{k}_f$  is parallel to the displacement vector,  $\mathbf{r}$ , and the following holds,

$$\exp(i \mathbf{k}_f \cdot \mathbf{r}) = \exp(i k r). \quad (2.14)$$

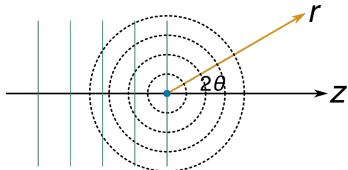


Figure 2.8: A schematic showing the propagation of the wave of probing radiation (green lines) radially outwards following the scattering event, where  $r$  is the magnitude of the displacement vector. Adapted, with permission of Oxford University Press, from Sivia, *Elementary Scattering Theory: For X-Ray and Neutron Users*.

This final wave is no longer collimated and therefore diminishes with distance,  $r$ . Hence the final scattered wave has the form,

$$\psi_f = \psi_o b \frac{\exp(i k r)}{r}, \quad (2.15)$$

where,  $b$  is the scattering length discussed in Section 2.1.3.

### 2.2.3 Scattering from multiple particles

It is important to consider how the probing radiation would interact with a real system, consisting of many particles. If the incident beam has the form of Equation 2.13, with the wavevector  $\mathbf{k}_i = (0, 0, k)$ , each particle,  $j$ , will contribute the following to the total scattered wave,  $\psi_f$ , made up of the scattering from all,  $N$ , atoms,

$$[\delta\psi_f]_j = \psi_o \exp(i \mathbf{k}_i \cdot \mathbf{R}_j) b_j \frac{\exp\{i \mathbf{k}_f \cdot (\mathbf{r} - \mathbf{R}_j)\}}{|\mathbf{r} - \mathbf{R}_j|}, \quad (2.16)$$

where,  $\mathbf{R}_j$  is the position of particle  $j$ ,  $\mathbf{r}$  is some arbitrary position, and  $\mathbf{k}_f$  is the wavevector of the scattered wave, described graphically in Figure 2.9. This allows the total scattered wave to be defined as a summation of the contributions from the individual waves,

$$\psi_f = \psi_o \exp(i \mathbf{k}_f \cdot \mathbf{r}) \sum_{j=1}^N \left\{ b_j \frac{\exp(i \mathbf{q} \cdot \mathbf{R}_j)}{|\mathbf{r} - \mathbf{R}_j|} \right\}. \quad (2.17)$$

Equation 2.17 holds true, within the Born approximation, where the scattered wave has no impact on the incident wave and each wave is scattered only once.

The sample-detector distance is usually much larger than the typical particle size, allowing for the following approximation,

$$|\mathbf{r} - \mathbf{R}_j| = |\mathbf{r}| = r. \quad (2.18)$$

This is termed the Fraunhofer, or far-field limit, and allows Equation 2.17 to be simplified,

$$|\psi_f|^2 = \frac{\Phi}{r^2} \left| \sum_{j=1}^N b_j \exp(i \mathbf{q} \cdot \mathbf{R}_j) \right|^2. \quad (2.19)$$



Figure 2.9: A schematic showing the interaction of raditon scattered by two particles that are separated by the vector  $\mathbf{R}_j$ . Adapted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.

In the scattering experiment, radiation is deflected elastically into a detector with a small area,  $\delta A$ , at the polar coordinates,  $2\theta$  and  $\phi$ , at a rate of  $R_{\text{el}}$ ,

$$R_{\text{el}}(2\theta, \phi) = |\psi_f|^2 \delta A = \Phi \delta \Omega \left| \sum_{j=1}^N b_j \exp(\mathbf{i} \mathbf{q} \cdot \mathbf{R}_j) \right|^2, \quad (2.20)$$

where,  $\delta \Omega = \delta A / r^2$ . Therefore, the differential cross section, defined in Equation 2.12 can be related to the scattering from the sample as,

$$\left( \frac{d\sigma(q)}{d\Omega} \right)_{\text{el}} = \frac{1}{V} \left| \sum_{j=1}^N b_j \exp(\mathbf{i} \mathbf{q} \cdot \mathbf{R}_j) \right|^2. \quad (2.21)$$

## 2.2.4 Scattering length density

While it may be helpful to consider the scattering from multiple particles individually, where each particle has a scattering length,  $b$ . In practice, due to low experimental resolution at small angles, it is more common to consider the SLD of the system,

$$\text{SLD} = \frac{1}{V} \sum_{i=0}^N b_i, \quad (2.22)$$

where  $N$  is the total number of particles in the volume  $V$ . A result of this equation is the ability to rewrite Equation 2.21 as,

$$\left( \frac{d\sigma(q)}{d\Omega} \right)_{\text{el}} = \frac{1}{V} \left| \iiint_V \text{SLD} \exp(\mathbf{i} \mathbf{q} \cdot \mathbf{R}) d^3 \mathbf{R} \right|^2. \quad (2.23)$$

This equation shows that the scattering differential cross-section from some object is related to the SLD profile of that object by a Fourier transform.

### 2.2.5 Model-dependent analysis

All types of scattering patterns can be analysed by two analysis methods; model independent and model-dependent. Model-independent analysis is where there is no *a priori* information used in the analysis, when there are no assumptions made about the underlying structure of the sample. However, model-dependent analysis is when reasonable assumptions are made about the structure before the analysis is considered. The nature of this work means that it will focus on model-dependent analysis methods.<sup>32</sup> Model-dependent analysis has significant benefits over model-independent methods, such as improved resolution and more detailed information about the structure. However, the necessity of the inclusion of *a priori* information within model-dependent analysis may act to bias the result. While this is undesirable, these assumptions can, and should, be educated based on the chemical information present.<sup>33</sup>

<sup>32</sup> With the model usually being derived from some atomistic, or coarse-grained simulation.

<sup>33</sup> McCluskey *et al.*, see n. ??, such as the propensity for twin-tailed phospholipid molecules to form monolayers at an air-water interface or small surfactants to form micelles in solution.

<sup>34</sup> This typically uses some optimisation algorithm to determine the best solution, the particular algorithms used in this work are discussed in Section 2.4.

<sup>35</sup> A. R. J. Nelson *et al.* *J. Appl. Crystallogr.* 46.5 (2013), pp. 1338–1346; A. R. J. Nelson. *J. Appl. Crystallogr.* 47.3 (2014), p. 1162.

The scattering from the model system is determined, using technique specific methods that are discussed in detail in later sections. This is then compared with the experimental data using some figure of merit, the model is then varied to find the best possible model for the data provided.<sup>34</sup> In order to accurately reproduce the experimental measurement, it is necessary to include some instrumental resolution function,  $res(q)$ , in the modelling procedure. This is instrument-specific, although it may be approximated by convolving the experimental dataset with some Gaussian smearing function, the modelled intensity can then be determined from,<sup>35</sup>

$$I(q) = res(q) * \frac{d\sigma(q)}{d\Omega}, \quad (2.24)$$

where,  $d\sigma(q)/d\Omega$  is the differential cross-section, a measure of the number of scattering particles hitting a given solid angle of the detector.

The aim of model-dependent analysis is to obtain a model for the system which agrees well with the experimentally measured scattering data while producing something that is chemically and physically relevant.

### 2.2.6 Reflectometry

Reflectometry involves the interaction of the probing radiation with some interface, from which the radiation is reflected. The geometry of a reflectometry experiment is shown in Figure 2.10, where the reflectometry instrument is in the horizontal configuration, ideal for the study of liquid interfaces.<sup>36</sup> Reflectometry measurements give information about the structure perpendicular to the interface, the  $z$ -dimension in Figure 2.10, and therefore the analysis of reflectometry data is founded on the assumption that the layers will be completely homogenous in the plane of the interface, the  $xy$ -plane in Figure 2.10. In reality, since the layers are usually not completely

<sup>36</sup> Such at those investigated in Chapters 3 and 4.



Figure 2.10: A schematic showing the geometry of a typical specular reflectometry experiment from a layered sample. Reprinted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.

homogeneous, an average is obtained for the area in the radiation beam.

A reflectometry instrument operates by measuring the intensity of specular reflected radiation at a series of different angles,  $\theta$ , or wavelengths,  $\lambda$ . The reflected intensity is described in terms of the  $q$ -vector<sup>37</sup>, and is defined as follows,

$$R(q) = \frac{\text{specular reflected radiation intensity at } q}{\text{incident radiation intensity}}. \quad (2.25)$$

It is clear from Equation 2.25 that the value of the measured reflectometry cannot be greater than one, as this would mean that more particles of probing radiation were being reflected than were incident.

There are two model-dependent analysis techniques that can be applied to the understanding of a reflectometry dataset. The first is the kinematic approach, which can be described with Equation 2.21, from the assumption that  $q_x = 0$  and  $q_y = 0$ , as only the specular scattering is being measured. This approach models the reflectometry as a function of the SLD profile in the  $z$ -dimension,  $\text{SLD}(z)$ ,

$$R(q) \approx \frac{16\pi^2}{q^4} \left| \int_{-\infty}^{+\infty} \frac{d\text{SLD}(z)}{dz} \exp(-izq_z) dz \right|^2, \quad (2.26)$$

where,  $d\text{SLD}(z)/dz$  is the first derivative of the SLD profile. However, this method has a significant problem, which can be demonstrated by applying Equation 2.26 to the SLD profile of a bare silicon substrate, which can be modelled as a Heaviside function, as shown in Figure 2.11(a),

$$\text{SLD}(z) = \begin{cases} 0, & \text{where } z < 0 \\ \text{SLD}_{\text{Si}}, & \text{otherwise} \end{cases} \quad (2.27)$$

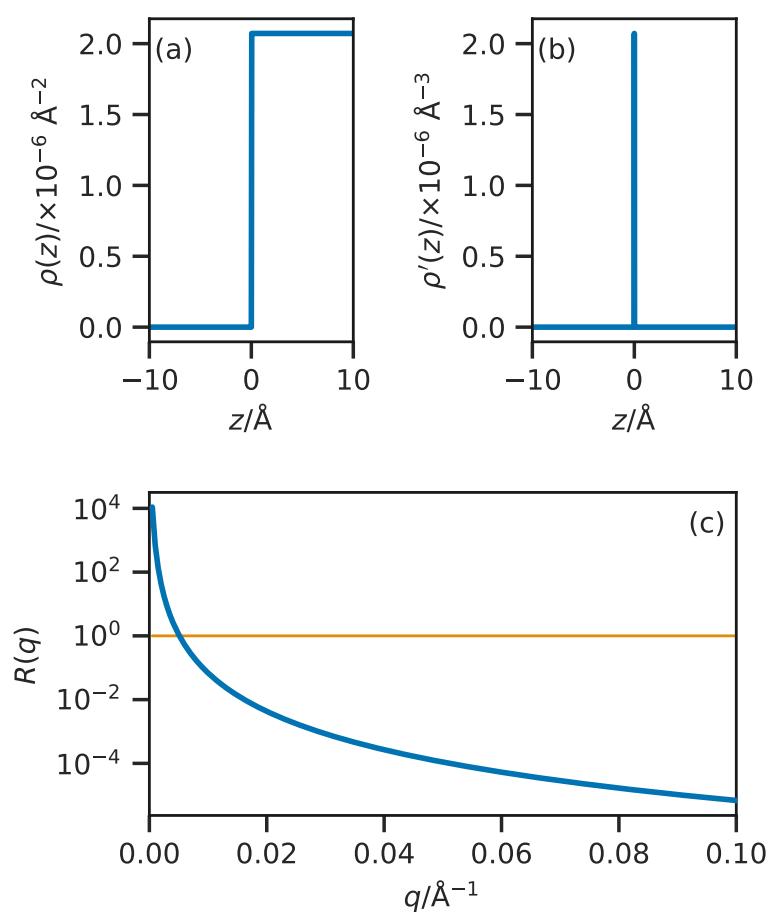
where,  $\text{SLD}_{\text{Si}}$  is the SLD of pure silicon.<sup>38</sup> The derivative of a step-wise Heaviside function is a scaled  $\delta$ -function, as shown in Figure 2.11(b),

$$\text{SLD}'(z) = \text{SLD}_{\text{Si}}\delta(z). \quad (2.28)$$

<sup>37</sup> Using Equation 2.9

<sup>38</sup> This is  $2.1 \times 10^{-6} \text{ \AA}^{-2}$  for neutrons.

Figure 2.11: A graphical representation of the kinematic approach; (a) the Heaviside function describing the SLD profile of a bare silicon substrate, (b) the  $\delta$ -function arising from the first derivative of the function in (a), and (c) the reflectometry profile resulting from the kinematic approach, where the orange line at  $R = 1$  identifies the breakdown between experimental and theory in the kinematic approach. Adapted, with permission of Oxford University Press, from D. S. Sivia. *Elementary Scattering Theory: For X-Ray and Neutron Users*. 2011.



Then, as in Equation 2.26, the Fourier transform of this  $\delta$ -function is taken,

$$\text{SLD}_{\text{Si}} \int_{-\infty}^{+\infty} \delta(z) \exp(-\mathbf{i}zq_z) dz = \text{SLD}_{\text{Si}} \exp(0) = \text{SLD}_{\text{Si}}. \quad (2.29)$$

This means that, using Equations 2.29 and 2.26, the reflectometry profile could be calculated from the following relationship,

$$R(q) \approx \frac{16\pi^2 \text{SLD}_{\text{Si}}^2}{q_z^4}. \quad (2.30)$$

The curve from this relationship is shown in Figure 2.11, where it is clear that the agreement with an experimental profile would be poor as  $q \rightarrow 0$ . It can be seen that for low values of  $q$  the calculated reflectometry is greater than 1, which violates the physical constraint imposed with Equation 2.25. This breakdown of the kinematic approach is due to the assumption present in this approach that the Born approximation<sup>39</sup> will hold. However, in the reflectometry scattering geometry, this is no longer true rendering the kinematic approach invalid.

This breakdown of the kinematic approach has led to the application of the Abelès, or Parratt, model for the reflection of light

<sup>39</sup> Mentioned previously in Section 2.2.3.

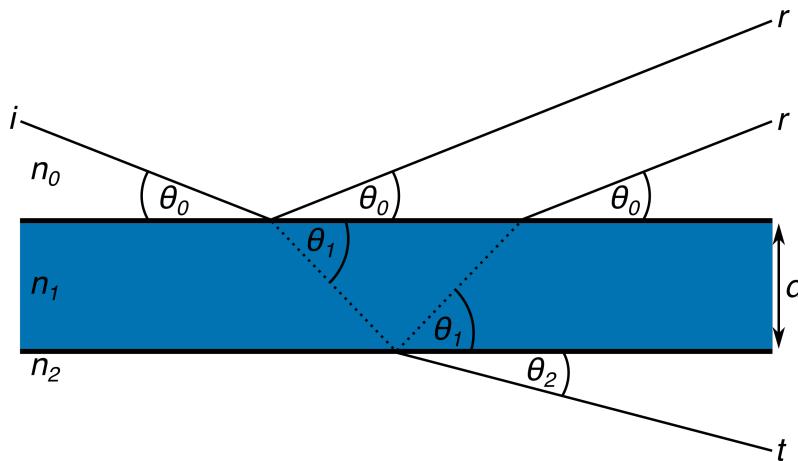


Figure 2.12: A schematic diagram showing the reflected ( $r$ ) and transmitted ( $t$ ) waves when an incident ( $i$ ) wave enters an interface of thickness  $d$ , where the refractive indices of each layer are  $n_0$ ,  $n_1$ , and  $n_2$ . Adapted, with permission of Elsevier, from F. Foglia et al. *Curr. Opin. Colloid Interface Sci.* 20.4 (2015), pp. 235–243.

at a given number of stratified interfaces.<sup>40</sup> This method involves considering the system as a layered structure at the interfaces of which, the probing radiation can either be reflected or refracted, by some refractive index,  $n_i$ . Figure 2.12 shows this process for a system of two layers, where the layer 0 is the air or vacuum above the sample, it is clear to see how the two waves labelled  $r$  could interfere constructively or destructively depending on the thickness of layer 1,  $d$ . This means that for a single interface,<sup>41</sup> the reflectometry can be described by the Fresnel equation,

$$R(q) = \left| \frac{n_0 \sin \theta_0 - n_1 \sin \theta_1}{n_0 \sin \theta_0 + n_1 \sin \theta_1} \right|^2. \quad (2.31)$$

Additionally at the point of total reflection, where  $\theta_0 = \theta_c$ , the critical angle, there will be no transmitted wave so,

$$n_1 \sin \theta_1 = 0, \quad (2.32)$$

and therefore the reflected radiation will never be greater than 1, while the critical angle can be defined as,

$$\cos^2 \theta_c = \frac{n_1^2}{n_0^2}. \quad (2.33)$$

This is the angle below which a reflectometry profile will be measured.

The above method can then be generalised to a structure of an arbitrary number of layers, as shown in Code Block 2.1.<sup>42</sup> For each value of  $q$  for which the reflectometry is to be calculated, the system is considered in terms of  $n_{\max}$  layers. The incident radiation beam will be refracted by each of the layers, giving wavevectors values for each layer,  $k_n$ ,

$$k_n = \sqrt{k_0^2 + 4\pi(\text{SLD}_n - \text{SLD}_0)}, \quad (2.34)$$

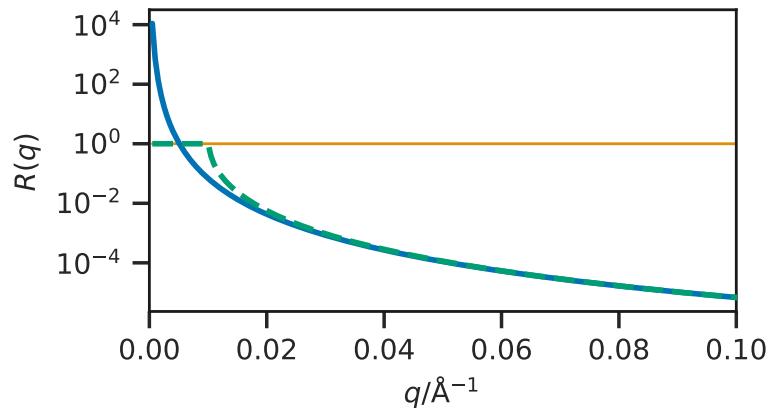
where,  $k_0 = q/2$ . The Fresnel equation coefficient between layers  $n$  and  $n+1$ ,  $r_{n,n+1}$  can then be found along with the phase factor,  $\beta_n$ ,

<sup>40</sup> F. Abelès. *Ann. Phys.* 12.3 (1948), pp. 504–520; L. G. Parratt. *Phys. Rev.* 95.2 (1954), pp. 359–369, also known as dynamical theory.

<sup>41</sup> Such as that between layers 0 and 1 in Figure 2.12.

<sup>42</sup> The purpose of the Code Blocks throughout this work is to ensure transparency and reproducibility, these are written as Python functions.

Figure 2.13: A comparison of the kinematic approach (blue solid line), and the dynamical approach (green dashed line), to determine the reflected intensity from the material with the SLD profile given in Figure 2.11(a). It is clear that at low  $q$ , there is a noticeable deviation between the two.



as the Born approximation holds for SAS, which is dependent on the thickness of the layer,  $d_n$ ,

$$r_{n,n+1} = \frac{k_n - k_{n+1}}{k_n + k_{n+1}}, \quad (2.35)$$

$$\beta_n = k_n d_n. \quad (2.36)$$

The means that a matrix can be evaluated for each layer,  $M_n$ ,

$$M_n = \begin{bmatrix} \exp \beta_n & r_{n,n+1} \exp -\beta_n \\ r_{n,n+1} \exp \beta_n & \exp -\beta_n \end{bmatrix} \quad (2.37)$$

The resultant matrix,  $B$ , is then found as a product of the matrix from each layer,

$$B = \prod_{n=0}^{n_{\max}} M_n, \quad (2.38)$$

and from this the reflected intensity at the given value of  $q$  can be found,

$$R(q_z) = \frac{B_{1,2}}{B_{1,1}}. \quad (2.39)$$

This algorithm models the layers as perfectly flat layers, which will not be strictly true.<sup>43</sup> This resulted in the correction term being added to Equation 2.35 to account for the roughness of the layers. This adapts Equation 2.35 to the form,

$$r_{n,n+1} = \frac{k_n - k_{n+1}}{k_n + k_{n+1}} \exp (-2k_n k_{n+1} \sigma_{n,n+1}^2), \quad (2.40)$$

where,  $\sigma_{n,n+1}$  is the interfacial roughness between layers  $n$  and  $n + 1$ .<sup>44</sup> This has the effect of Gaussian broadening the layers into each other, as a result. This method<sup>45</sup> is currently implemented in a variety of reflectometry modelling software packages, such as `refnx`, `MOTOFIT`, `RasCAL`, and `Aurore`<sup>46</sup>. Applying this method to the SLD profile shown in Figure 2.11 gives the reflectometry profile shown with the dashed green line in Figure 2.13.

<sup>43</sup> Particularly for soft matter systems.

<sup>44</sup> L. Nérot *et al.* *Rev. Phys. Appl. (Paris)* 15, 3 (1980), pp. 761–779. That is given programmatically in Code Block 2.1.

<sup>45</sup> Nelson *et al.*, “Refnx: Neutron and X-Ray Reflectometry Analysis in Python”; Nelson, “Co-Refinement of Multiple-Contrast Neutron/X-Ray Reflectivity Data Using MOTOFIT”; A. V. Hughes. *RasCAL*. URL: <https://sourceforge.net/projects/rscl/> (Accessed 2016-8-8); Y. Gerelli. *J. Appl. Crystallogr.* 49, 1 (2016), pp. 330–339; Y. Gerelli. *J. Appl. Crystallogr.* 49, 2 (2016), pp. 712–712.

```

# Copyright 2015-2019 A. R. J. Nelson
# Australian Nuclear Science and Technology Organisation
# Licensed under the BSD 3-Clause "New" or "Revised" License

import numpy as np

def abeles(q_values, sld, d):
    R = np.zeros_like(q_values)
    kn = np.sqrt(
        q_values[:, np.newaxis] ** 2.0 / 4.0 - 4.0 * np.pi * sld
    )
    B = np.zeros((2, 2, q_values.size))
    B[0, 0, :] = 1
    B[1, 1, :] = 1
    k = kn[:, 0]
    nmax = sld.size
    for n in range(1, nmax):
        kn1 = kn[:, n]
        r = (k - kn1) / (k + kn1)
        betan = k * d[n]
        if n > 0:
            Mn = np.array([
                [np.exp(betan * 1j), r * np.exp(betan * 1j)],
                [r * np.exp(-betan * 1j), np.exp(-betan * 1j)],
            ])
        else:
            Mn = np.array([[1, r], [r, 1]])
        p0 = B[0, 0, :] * Mn[0, 0, :] + B[1, 0, :] * Mn[0, 1, :]
        p1 = B[0, 0, :] * Mn[1, 0, :] + B[1, 0, :] * Mn[1, 1, :]
        B[0, 0, :] = p0
        B[1, 0, :] = p1
        p0 = B[0, 1, :] * Mn[0, 0, :] + B[1, 1, :] * Mn[0, 1, :]
        p1 = B[0, 1, :] * Mn[1, 0, :] + B[1, 1, :] * Mn[1, 1, :]
        B[0, 1, :] = p0
        B[1, 1, :] = p1
        k = kn1
    R = (B[0, 1, :] * np.conj(B[0, 1, :])) / (
        B[0, 0, :] * np.conj(B[0, 0, :])
    )
    R[np.where(np.isnan(R))] = 1.0
    return np.real(R)

```

**Code Block 2.1:** An example Python code block for the Abelès method for the calculation of reflectometry, adapted from A. R. J. Nelson *et al.* *J. Appl. Crystallogr.* 52.1 (2019), pp. 193–200. The input variables are *q\_values* which are the *q*-vectors at which the reflected intensity should be calculated, *sld* which is the array of scattering length densities for the layers, and *d* which is the array of thicknesses for the layers. This will return an array of floats that is the same size as the *q\_values* and contains the reflected intensities.

## 2.2.7 Small angle scattering

Equation 2.23 identified that the scattering differential cross-section for some object was related to the SLD by a Fourier transform, which is shown graphically in Figure 2.14. This figure shows that there is a reciprocal relationship between the size of the object and the scattered intensity, decaying significantly up to values of  $2\pi/d_x$ , where  $d_x$  is the size of the object. This means that in order to probe large-scale structural features that are of interest in the study of soft materials, it is necessary to consider small values of *q*. When considering the nature of *q* in Equation 2.9, it is clear that such experiments would benefit from small values of  $\theta$  and large values of  $\lambda$ . Hence, the use of scattering at small angles.

A SAS experiment generally involves some sample being placed in the path of the probing radiation; the scattering pattern that results from this transmission is measured at some distance, as is shown in Figure 2.15 for the D22 SANS instrument of the ILL. SAS instruments are usually very large, due to the large post sample flight path that is necessary to reach the small angles being measured.<sup>47</sup> Transmission SAS can provide information about the size, shape and orientation of the sample's components.<sup>48</sup> The range of *q* that is typically covered by a SAS instrument is usually around  $2 \times 10^{-3}$ – $0.5 \text{ \AA}^{-1}$ , which corresponds to 10–3000  $\text{\AA}$  in real-space. The neutron or X-ray detector of a SAS instrument is often two-dimensional, meaning that for an isotropic scattering profile, the detector image is radially averaged to give an  $I(q)$  scattering profile. It is possible to increase the *q*-range of a SAS instrument

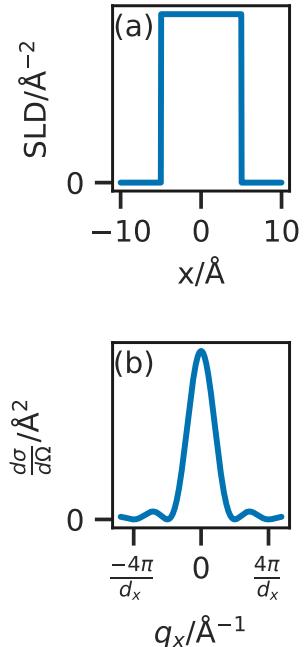
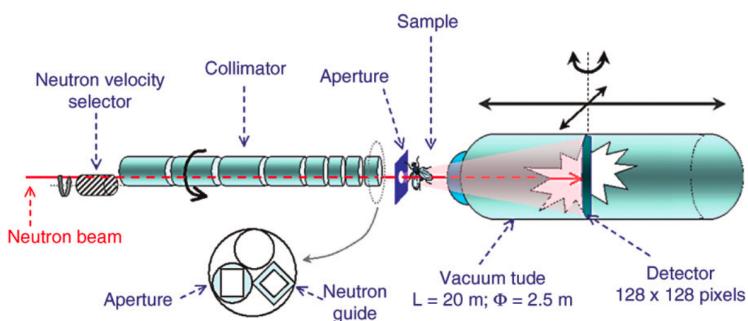


Figure 2.14: The effect of a Fourier transform (a) the SLD profile for some object with a width of 10  $\text{\AA}$ , (b) the Fourier transform of this object showing the minima in the differential cross section at values of  $2n\pi/10$ , where *n* is some integer.

<sup>47</sup> A longer flight path allows more space for angular divergence.

<sup>48</sup> B. T. M. Willis *et al.* *Experimental Neutron Scattering*. 2009.

Figure 2.15: A schematic of the D22 instrument of the ILL. Reprinted with permission of Springer Nature Customer Service Centre GmbH: Springer Nature from I. Grillo. "Small-Angle Neutron Scattering and Applications in Soft Condensed Matter". In: *Soft-Matter Characterization*. 2008, pp. 723–782.



<sup>49</sup> ISIS – SANS2D. URL: <https://www.isis.stfc.ac.uk/Pages/Sans2d.aspx> (Accessed 2018-9-25).

<sup>50</sup> K. J. Edler et al. *Curr. Opin. Colloid Interface Sci.* 20.4 (2015), pp. 227–234.

through the introduction of wide- $q$  detector banks close to the sample or small- $q$  detector banks further away. This allows the SANS2D instrument, at the ISIS Neutron and Muon Source, to have a total range from  $2 \times 10^{-3}$ – $2 \text{ \AA}^{-1}$ .<sup>49</sup> Furthermore, the SANS2D instrument may leverage the ToF method discussed in Section 2.1.2 to allow for a much shorter post sample flight path than is present at D22.

A radially averaged SAS pattern can be considered as consisting of two sections that arise from the form and structure factors for the scattering species. The form factor gives information about the average shape of the scattering particle, while the structure factor is a measure of the interaction present between the objects. It is often possible to control the presence of the structure factor by changing the concentration of the sample, eventually, the concentration will be so low that all interparticle interaction is screened by the solvent.<sup>50</sup> This method is frequently applied in biological SAXS applications, where the interactions between the protein molecules are of less interest than the overall structure of the complex. However, with micelles, it is not always possible to remove the structure factor, as the critical micelle concentration may be higher than the minimum concentration at which the structure factor is present. It is possible to deconvolute the structure and form factors for a micellar solution by studying different concentrations, assuming that the form of the micelle is concentration independent, over the measured concentration range.

The rigorous, model-independent method for the analysis of SAS involves taking the inverse Fourier transform of the scattering profile, to give an auto-correlation function of the average particle in the system, which following a deconvolution procedure will resolve the radially averaged SLD profile. However, this is often cumbersome and has a low information density, when compared to model-dependent techniques. Additionally, if the experimental data lacks information at wide enough  $q$  to cover all features of the sample, artefacts may be present in the inverse Fourier transform of the scattering.

There are two common and straight-forward analysis proce-

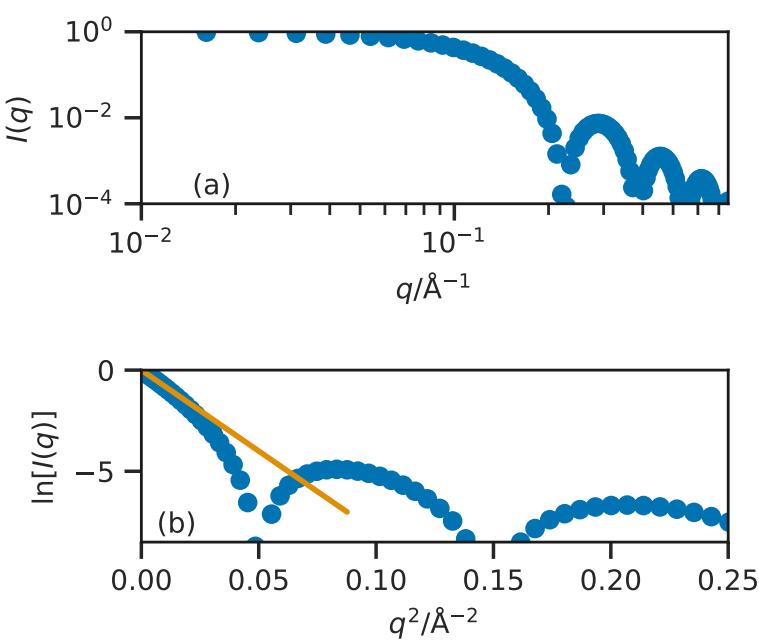


Figure 2.16: The Guinier plot, (a) the ideal scattering profile from a sphere of radius 20 Å, (b) the associated Guinier plot, with a straight line (orange) at low- $q$  showing the radius of gyration to be  $\sim 15.5$  Å.

dures that can be used to give an understanding of the scattering species structure. The first is the Guinier approximation, which is used in the determination of the radius of gyration,  $R_g$ , of the scattering species at “infinite dilution”. This scattering law is only valid at very small values of  $q$ , where  $q < R_g^{-1}$ ,<sup>51</sup>

$$\ln[I(q)] = \ln[I(0)] - \left(\frac{R_g^2}{3}\right)q^2. \quad (2.41)$$

This relationship allows the radius of gyration to be found by plotting the scattering profile transformed into  $\ln[I(q)]$  vs.  $q^2$ , and evaluating the gradient at low  $q$ . The Guinier plot for the scattering from a sphere with a radius of 20 Å is shown in Figure 2.16, where the radius of gyration correlates with the radius of the sphere,  $R_s$ , as follows,

$$R_g = \sqrt{\frac{3}{5}}R_s. \quad (2.42)$$

The Guinier analysis is very common in the study of proteins by SAS, as it allows for the determination of the protein size in the native, solution phase.<sup>52</sup> Another common analysis of SAS data comes in the form of Porod’s law, which states that for large values of  $q$ , the scattering intensity becomes proportional to  $Sq^{-4}$ , where  $S$  is the surface area of the sample. This means that by plotting  $I(q)q^4$  vs.  $q$  and extrapolating to  $q \rightarrow \infty$ , it is possible to determine the external surface area of the system.<sup>53</sup> Using the surface area, it is then possible to qualitatively determine the “roughness” of the system based on the relation of the surface area to the particle size.

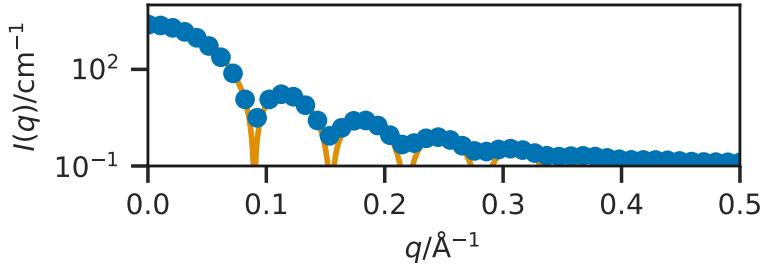
In the calculation of a SAS pattern, both the structure and form factors will contribute. Therefore, when the pattern is modelled,

<sup>51</sup> Sivia, see n. 15.

<sup>52</sup> S. Skou *et al.* *Nat. Protoc.* 9.7 (2014), pp. 1727–1739.

<sup>53</sup> Willis *et al.*, see n. 48.

Figure 2.17: The SANS profile of a micelle of  $C_{16}$ TAB with radius  $(50 \pm 3) \text{ \AA}$  (circles, generated using SASView *SASview for Small Angle Scattering*. URL: <http://www.sasview.org> (Accessed 2016-10-26), with instrumental smearing) compared with a curve of Equation 2.44, where  $R = 50 \text{ \AA}$  (solid line).



the differential cross-section, which is measured experimentally, has the following form.<sup>54</sup>

$$\frac{d\sigma(q)}{d\Omega} = N_p \Delta SLD^2 V_p^2 P(q) S(q), \quad (2.43)$$

where  $N_p$  is the number density of the particles,  $\Delta SLD$  is the difference in scattering length density between the particles and the solvent,  $V_p$  is the particle volume,  $P(q)$  is the particle form factor, and  $S(q)$  is the structure factor.<sup>55</sup> Therefore, it is necessary to understand the form and structure factors individually.

The most common method for the modelling of the form factor is by using very coarse shapes; such as spheres, cylinders, or ellipses. This involves the evaluation of analytical or quasi-analytical solutions for the scattering, which have been derived for many common shapes. The solution for a sphere was solved in the early 19th century by Lord Rayleigh.<sup>56</sup>

$$P(q) = \left\{ \frac{3\{\sin(qR_s) - qR_s \cos(qR_s)\}}{(qR_s)^3} \right\}^2, \quad (2.44)$$

where  $R_s$  is the radius of the sphere. A comparison between a possible experimental scattering pattern and the scattering generated from Equation 2.44 is shown in Figure 2.17. Analytical form factors exist for a wide variety of shapes; these can be found in software such as SASView and SASFit.<sup>57</sup>

The structure factor accounts for the scattering interference that arises from the interaction of different particles. This is modelled using expressions which depend on the nature of the scattering particles; hard-sphere, sticky hard-spheres, screened Coulomb, etc. Structure factor expressions are generated as solutions to the Ornstein-Zernike Equation.<sup>58</sup> The most relevant structure factor in terms of micelle modelling is probably the Hayter-Penfold Mean Spherical Approximation,<sup>59</sup> this is where the micelles are modelled as like-charged, soft spheres and is valid for dilute solutions. Again, a whole range of these structure factor functions are built into the SASView package.<sup>60</sup>

In order to evaluate the scattering profile from an atomic structure,<sup>61</sup> it is possible to use the Debye equation.<sup>62</sup> This is an analyti-

<sup>54</sup> When the system is centrosymmetric.

<sup>55</sup> Pedersen, see n. 27.

<sup>56</sup> Pedersen, see n. 27.

<sup>57</sup> SASview for Small Angle Scattering. URL: <http://www.sasview.org> (Accessed 2016-10-26); SASfit. URL: <https://kur.web.psi.ch/sans1/SANSSoft/sasfit.html> (Accessed 2018-11-11).

<sup>58</sup> R. Klein. "Interacting Colloidal Suspensions". In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*. 2002, pp. 351–380.

<sup>59</sup> J. B. Hayter *et al.* *Mol. Phys.* 42.1 (1981), pp. 109–118.

<sup>60</sup> SASview for Small Angle Scattering, see n. ??.

<sup>61</sup> Or any coordinate system, such as a coarse-grained simulation.

<sup>62</sup> P. Debye. *Ann. Phys.* 351.6 (1915), pp. 809–823.

cal relationship for the determination of the scattering profile based on the atomic positions,  $\mathbf{r}_i$ ,

$$I(q) = \sum_i^N \sum_j^N b_i b_j \frac{\sin(q|\mathbf{r}_i - \mathbf{r}_j|)}{q|\mathbf{r}_i - \mathbf{r}_j|}, \quad (2.45)$$

where,  $N$  is the number of particles,  $q$  is the scattering vector, and  $b_i$  and  $b_j$  are the scattering lengths of particles  $i$  and  $j$  respectively.<sup>63</sup> The Debye equation is powerful, however, it is not intrinsically parallelisable and scales as  $\mathcal{O}(N^2)$ . In order to improve the efficiency of the calculation of the scattering profile, a variety of methods have been developed that offer a sufficiently accurate approximation.<sup>64</sup> The Golden Vector method, developed by Watson and Curtis,<sup>65</sup> scales as  $\mathcal{O}(Nn)$ , where  $n$  is the number of scattering vectors used in the calculation. In this method, the scattering amplitude is calculated numerically for a single  $\mathbf{q}$ -vector,

$$I(\mathbf{q}) = \left[ \sum_i^N b_i \cos(\mathbf{q} \cdot \mathbf{r}_i) \right]^2 + \left[ \sum_i^N b_i \sin(\mathbf{q} \cdot \mathbf{r}_i) \right]^2. \quad (2.46)$$

This is carried out for  $n$  scattering vectors that are selected in an orientationally averaged fashion from a quasi-uniform lattice on a sphere. This was first developed such that  $n$  is a number from the Fibonacci sequence,<sup>66</sup> however for the Golden Vector method,  $n$  may be any positive integer. This leads to the scattering vectors being calculated as,

$$\begin{aligned} q_x^{(k)} &= q \cos \left[ \sin^{-1} \left( \frac{2k}{n} \right) \right] \cos \left( \frac{2\pi k}{\Phi} \right), \\ q_y^{(k)} &= q \cos \left[ \sin^{-1} \left( \frac{2k}{n} \right) \right] \sin \left( \frac{2\pi k}{\Phi} \right), \\ q_z^{(k)} &= \frac{2kq}{n}, \end{aligned} \quad (2.47)$$

where,  $k$  runs from  $-(n-1)/2, \dots, 0, \dots, (n-1)/2$  and  $\Phi$  is the golden ratio,

$$\Phi = \frac{1 + \sqrt{5}}{2}. \quad (2.48)$$

The approximate orientationally averaged scattering is then found as an average of the scattering from each of the individual  $q$ -vectors,

$$I(q) = \frac{1}{n} \left\{ \sum_{k=(1-n)/2}^{(n-1)/2} I[\mathbf{q}^{(k)}] \right\}. \quad (2.49)$$

The accuracy of this calculation increases with  $n$ , however agreement, comparable to the Debye equation, between experiment and simulation has been shown for  $n < 100$  even for highly anisotropic systems.<sup>67</sup>

<sup>63</sup> In the case of a coarse-grained particle, this can be obtained from summing the scattering lengths of the constituent atoms.

<sup>64</sup> D. I. Svergun. *Acta Crystallogr. A* 50.3 (1994), pp. 391–402; M. C. Watson *et al.* *J. Appl. Crystallogr.* 46.4 (2013), pp. 1171–1177.

<sup>65</sup> Watson *et al.*, see n. 64.

<sup>66</sup> Svergun, see n. 64.

<sup>67</sup> Watson *et al.*, see n. 64.

## 2.3 CLASSICAL SIMULATION

<sup>68</sup> Ivanović *et al.*, see n. 36; Scoppola *et al.*, see n. 31; Dabkowska *et al.*, see n. 40; J. S. Hub. *Curr. Opin. Struct. Biol.* 49 (2018), pp. 18–26.

While the currently applied traditional methods for the analysis of experimental scattering data discussed previously are popular. There is growing interest in the use of multi-modal analysis methods that leverage classical simulation to assist in the analysis of scattering data.<sup>68</sup> This would involve the simulation of the chemical system in order to educate the analysis of the experimental data. These systems, especially when the materials being simulated are soft in nature, are often highly complex and typically cover large length scales. Classical simulation, particularly in combination with coarse-grained potential models, can feasibly enable the simulation of these systems.

In order to simulate the complexity of a real chemical system, it is necessary to model the electrons of the molecules and their interactions. This is usually achieved using quantum mechanical calculations, where the energy of the system is calculated by finding some approximate solution to the Schrödinger equation. However, quantum mechanical calculations are very computationally expensive and are realistically limited to hundreds of atoms. In order to simulate a soft matter system such as a phospholipid monolayer or polymer nanoparticles, it is necessary to simplify the calculation being performed. This leads to the use of classical simulations, where parameterised analytical functions are used to represent the potential energy of the system. Classical simulations are used substantially in this work, in terms of molecular dynamics simulations.<sup>69</sup> Therefore, it is necessary to introduce the underlying theory on which this method is defined.

<sup>69</sup> Molecular dynamics (MD) is discussed in detail in Section 2.4.

### 2.3.1 Potential models

Potential modelling is a more computationally efficient method for the calculation of the potential energy of a chemical system. A potential model consists of a series of mathematical functions that depend on the atomic positions,  $\mathbf{r}$ . Each of the functions represents the potential energy of a different interaction for a given atom. Broadly, these interactions can be split into bonded and non-bonded, such that the total energy may be described as follows,

$$E_{\text{total}}(\mathbf{r}) = \sum_{\text{bonded pairs}} E_{\text{bonded}}(\mathbf{r}) + \sum_{\text{atom pairs}} E_{\text{non-bonded}}(\mathbf{r}). \quad (2.50)$$

The total potential energy is then the sum of the potential energy for each of the individual atoms.

The bonded terms are used to describe different aspects of chemical bonds. These typically consist of bond stretches, angle bends and dihedral torsions, these interactions have the following mathematical form,<sup>70</sup>

<sup>70</sup> These forms are specific to the OPLS2005 potential model (J. L. Banks *et al.* *J. Comput. Chem.* 26.16 (2005), pp. 1752–1780), other potential models may have different functions.

$$\begin{aligned}
E_{\text{bonded}}(b, \theta, \phi) = & \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{\text{dihedrals}} \frac{1}{2} \{ A_1[1 + \cos(\phi)] \\
& + A_2[1 - \cos(2\phi)] + A_3[1 + \cos(3\phi)] \},
\end{aligned} \tag{2.51}$$

where,  $K_b$  and  $b_0$ ,  $K_\theta$ ,  $\theta_0$ , and  $A_1$ ,  $A_2$ , and  $A_3$  are interaction dependent parameters for the bonds, angles, and dihedrals respectively, while  $b$ ,  $\theta$ , and  $\phi$  are the bond lengths, the size of the angles, and the size of the dihedrals that depend on the atom positions.<sup>71</sup> It can be seen that both the bond stretch and angle bend have harmonic functions, whereas the dihedral consists of a more complex multiple cosine functions.

The non-bonded terms are a series of functions that describe the potential energy of intermolecular interactions, such as electrostatics and London dispersion forces. The potential energies of the short-range interactions are usually modelled as a combination of the attractive London dispersion interaction and the repulsive exchange forces that arise from the Pauli exclusions principle.<sup>72</sup> These are often forms such as shown below for the Lennard-Jones potential model,<sup>73</sup>

$$\begin{aligned}
E_{\text{non-bonded}}(r) = & E_{\text{repulsive}} + E_{\text{attractive}} \\
= & \frac{A}{r^{12}} - \frac{B}{r^6} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]
\end{aligned} \tag{2.52}$$

where,  $r$  is the distance between two particles,  $A$  and  $B$  are interaction dependent parameters, and  $\sigma$  and  $\epsilon$  are simple reformations of these parameters,

$$A = 4\epsilon\sigma^{12} \quad B = 4\epsilon\sigma^6. \tag{2.53}$$

Figure 2.18 shows each component of the Lennard-Jones potential model for atoms of argon.<sup>74</sup> The Lennard-Jones potential model is not the only form that may be used for the modelling of the short-range non-bonded interactions, others such as the Buckingham and Morse potentials exist.<sup>75</sup> In each case, there is a short ranged repulsive interaction to describe the electrostatic repulsion between the electron clouds, and a longer range attractive component that represents dispersion interactions. However, the Lennard-Jones model has been used heavily in this work.

While the short-range interactions can be accounted for by a function such as the Lennard-Jones potential model, the potential energy of the long-range electrostatic interactions are usually modelled, more consistently, using Coulomb's law for classical electrostatic interaction between point particles,<sup>76</sup>

$$E_{\text{Coulomb}}(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j e^2}{r^2}, \tag{2.54}$$

where,  $r$  is the distance between the two particles,  $\epsilon_0$  is the dielectric permittivity of the vacuum,  $e$  is the charge of the electron, and

<sup>71</sup> The values for the interaction dependent parameters are determined as outlined in Section 2.3.2.

<sup>72</sup> A. R. Leach. *Molecular Modelling: Principles and Applications*. 1996.

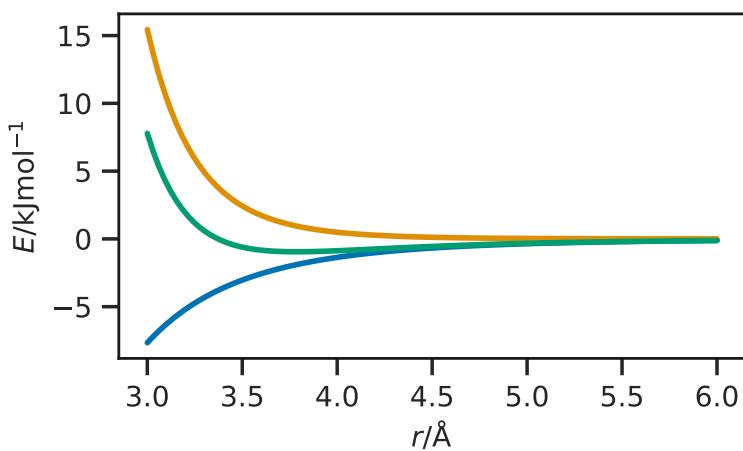
<sup>73</sup> J. E. Lennard-Jones. *Proc. Royal Soc. Lond. A.* 106.738 (1924), pp. 463–477.

<sup>74</sup> Using parameters for  $A$  and  $B$  determined in A. Rahman. *Phys. Rev.* 136 (2A 1964), A405–A411.

<sup>75</sup> R. A. Buckingham. *Proc. Royal Soc. Lond. A.* 168.933 (1938), pp. 264–283; P. M. Morse. *Phys. Rev.* 34.1 (1929), pp. 57–64.

<sup>76</sup> C. A. Coulomb. *Histoire de l'Académie Royale des Sciences*. Imprimerie Royale (1788), pp. 569–577; C. A. Coulomb. *Histoire de l'Académie Royale des Sciences*. Imprimerie Royale (1788), pp. 578–611.

Figure 2.18: The form of each component; attractive (blue), repulsive (orange), of the Lennard-Jones potential model (green) for argon, using parameters in A. Rahman. *Phys. Rev.* 136 (2A 1964), A405–A411.



$q_i$  and  $q_j$  are the electronic charges on each of the particles. It is clear that when  $q_i$  and  $q_j$  have the opposite signs Coulomb's law is always attractive. The fact that Equation 2.54 contains a factor of  $r^2$  indicates that this is a much longer range interaction than those modelled with the Lennard-Jones model, make the Coulomb potential more complex to compute.<sup>77</sup>

An example of a very large classical simulation would be  $\sim 3$  million atoms.<sup>78</sup> However, this is still only  $1.8 \times 10^{-16}$  mol which is not remotely realistic as a simulation of a “real” system. A common method to allow for the apparent simulation of a much larger system is the use of periodic boundary conditions.<sup>79</sup> This is where a boundary condition is applied to the edges of the simulation cell, such as to mimic an infinite system, such that the simulation cell is surrounded by identical images of itself, this is shown pictorially in Figure 2.19. Using the PBC means that atomic diffusion is conserved as when an atom reaches the edge of the simulation cell, it will appear on the other side as though it came from the adjacent periodic cell. The use of a PBC is particularly powerful in the simulation of homogenous systems, such as liquids.

The cut-off is another important factor for classical simulation, this is the distance after which the energy between two particles

<sup>77</sup> D. Frenkel *et al.* *Understanding Molecular Simulation: From Algorithms to Applications*. 1996.

<sup>78</sup> J. Gumbart *et al.* *Structure* 17.11 (2009), pp. 1453–1464.

<sup>79</sup> Abbreviated to PBC.

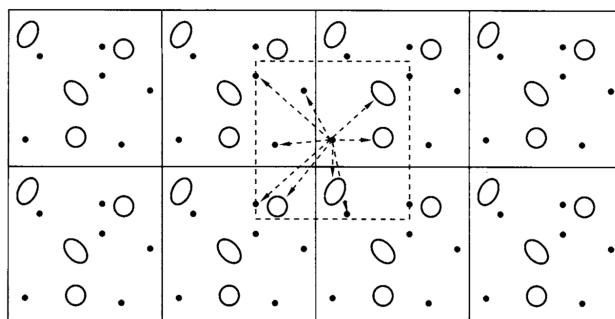


Figure 2.19: A graphical representation of the PBCs. Reprinted, with permission of Elsevier, from D. Frenkel *et al.* *Understanding Molecular Simulation: From Algorithms to Applications*. 1996.

is considered to be zero. Therefore, for distances greater than the cut-off, it is not necessary to calculate the energy between the two particles as it is taken to be zero.<sup>80</sup> Code Block 2.2 gives an example of some code that could be used to calculate the Lennard-Jones energy of an atomistic system, where both the PBC and the energy cut-off distance are considered.

The use of the PBC may be problematic for systems containing long-range interactions, such as classical electrostatics, due to the fact that the range of the electrostatic interaction may be much greater than the size of half of the simulation cell, which can be taken to be the energy cut-off distance. In order to avoid truncation artefacts, the Ewald summation is often used for the calculation of the electrostatic contribution to the potential energy.<sup>81</sup> The Ewald summation involves performing the summation of the contributing interaction energies in reciprocal space rather than in real space as is the case for the short-range interactions. Most modern MD simulation software packages implement the Ewald summation using a particle mesh Ewald method.<sup>82</sup>

<sup>80</sup> This leads to an increase in computational efficiency.

<sup>81</sup> P. P. Ewald. *Ann. Phys.* 369.3 (1921), pp. 253–287.

<sup>82</sup> U. Essmann *et al.* *J. Chem. Phys.* 103.19 (1995), pp. 8577–8593.

### 2.3.2 Parameterisation

Section 2.3.1 introduced the idea of potential models that may be used to evaluate the potential energy of a given system, requiring much less time than methods that rely on the use of quantum mechanics. However, for these methods to be effective, it is important that the potential models used are able to model the system under study accurately. This is achieved initially by selecting the correct potential model for a given interaction, and then by ensuring that the interaction-dependent parameters are accurate for a given interaction. The method of obtaining such parameters is referred to as “parameterising” the model. Model parameterisation is important for all types of potential models, for example it is necessary to determine the equilibrium bond length  $b_0$  and the force constant  $K_b$  for a given covalent bonds, or the partial electrostatic charge that is present on a carbonyl oxygen atom when it interacts with the hydrogen atom from a neighbouring hydroxyl group.

Parameterisation of a potential model is usually achieved by fitting the potential model functions to energetic data obtained using

```
import numpy as np

def lj_energy(coordinates, cell, cut_off, A, B):
    energy = np.zeros((coordinates.shape[0]))
    for i in range(coordinates.shape[0] - 1):
        for j in range(i + 1, coordinates.shape[0]):
            d = coordinates[j] - coordinates[i]
            d = d % cell
            r = np.sqrt(np.sum(np.square(d)))
            if r > cut_off:
                continue
            else:
                energy[i] += A / np.power(r, 12) - B / np.power(r, 6)
                energy[j] += A / np.power(r, 12) - B / np.power(r, 6)
    return energy
```

Code Block 2.2: Code that may be used to generate the Lennard-Jones energy for a given atomistic system, which accounts for the PBC and the energy cut-off distance. The input variables are `coordinates` which is an array of floats describing the position of the  $N$  particles, `cell` which are the unit cell vectors, `cut_off` which is potential energy cut-off, and `A` and `B` which are the Lennard-Jones potential parameters. This returns an array with the energy for each particle.

<sup>83</sup> These may be quantum mechanical calculations or experimental methods.

<sup>84</sup> This is discussed in detail in many textbooks such as J. Harvey. *Computational Chemistry*. 2018; Leach, see n. 72

<sup>85</sup> W. L. Jorgensen *et al.* *J. Am. Chem. Soc.* 110.6 (1988), pp. 1657–1666.

<sup>86</sup> In particular for emergent properties that depend on large system sizes and long simulation times.

<sup>87</sup> Marrink *et al.*, see n. 43.

<sup>88</sup> Therefore requiring fewer integrations of the equations of motion

<sup>89</sup> K. Pluhackova *et al.* *J. Phys. Condens. Matter* 27.32 (2015), p. 323103.

a higher accuracy technique.<sup>83</sup> We will not dwell on the details of potential model parameterisation,<sup>84</sup> however, it is important to note that the parameters used in MD simulation are not absolute and depend heavily on the merits of the parameterisation method.

In this work, we have focused heavily on the use of off-the-shelf potential models, to ensure the easy replicability of the work. Off-the-shelf potential models are those that are determined to be applied to a wide range of chemical systems. An example includes the OPLS potential model which was parameterised by comparison to quantum mechanical measurements and crystallographic data.<sup>85</sup> While these off-the-shelf potential models are useful for their ease-of-use, it is noted that often these forcefields may require optimisation for the particular system.

### 2.3.3 Coarse-graining

The atomistic simulation of very large systems, such as multiple surfactant micelles or large phospholipid monolayers, require a huge number of atoms. While computational efficiency improvements such as the PBC or the energy cut-off distance are able to reduce the time taken to simulate these systems, it is often still not possible to produce physically meaningful simulations,<sup>86</sup> without including some other efficiency improvements.

This has led to the use of coarse-graining of molecules in simulations. This is the definition of super-atoms, in the place of groups of atoms, known as “beading”, some examples are shown for the MARTINI force field<sup>87</sup> in Figure 2.20. Each of the super-atoms must correspond to the chemistry of the underlying atoms. For example, the MARTINI potential model introduces five different apolar, beads to represent the polarity of the carbon atoms that make up the super-atom. Additionally, there are thirteen other super-atom types that can be used to model polar, nonpolar, and charged atomic groups.

In addition to the computational benefit of having fewer particles in the simulation,<sup>88</sup> there is also the opportunity to increase the timestep length for the simulation.<sup>89</sup> This can be achieved as the highest frequency vibrations that must be modelled in the system are integrated out. For example, another coarse-grained model called the united atom potential model, where the hydrogen atoms have been integrated out, the timestep may be larger than for the same all-atom system as it is no longer necessary to model the high-frequency C–H bond.

The technique of coarse-graining a molecule can range from the integration of the hydrogen atoms into the heavier atoms to which they are bound, all the way to the treatment of entire molecules as a single “bead”, with the inclusion of an implicit solvent. The parameterisation of a coarse-grained potential model is carried out in much the same way as discussed in Section 2.3.2 for all-atom

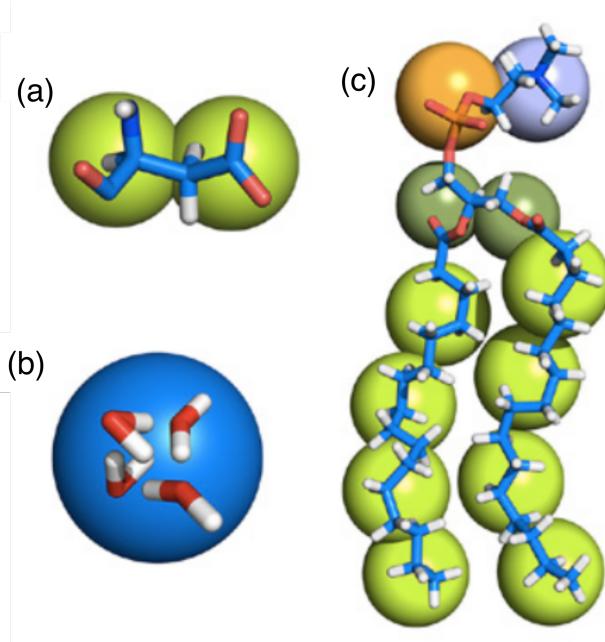


Figure 2.20: Three examples of the MARTINI coarse-graining mechanism for (a) aspartic acid, (b) a water cluster, and (c) a molecule of DPPC. Reprinted with permission of the Institute of Physics, from K. Pluhackova *et al.* *J. Phys. Condens. Matter* 27.32 (2015), p. 323103.

potential models. The coarse-grained parameters are determined by comparison with a higher-resolution technique.<sup>90</sup>

<sup>90</sup> Often this is all-atom MD simulations.

## 2.4 OPTIMISATION & SAMPLING METHODS

In this work, computational modelling methods have been applied to important scattering problems. The aim of many modelling problems is to optimise a series of parameters such that a minimum in some parameter-dependent metric is found. While, in other circumstances, the aim is to sample the parametric search-space of a particular problem. The problem of parameter optimisation and sampling is a massive area of mathematics and computer science and is it not possible to introduce the whole field. Therefore, I will introduce two optimisation methods and two sampling methods that are applied within this work.

Both optimisation algorithms in this work are population-based, making use of a population of candidate solutions. These populations of candidate solutions often have knowledge of the state of each other through some interaction method. The interaction method is often used to characterise the algorithms, into evolutionary algorithms and swarm intelligence algorithms.<sup>91</sup> These population methods are usually more efficient at finding the global minimum for a given search space, than a single candidate method.

### 2.4.1 Differential evolution

Differential evolution<sup>92</sup> is a common, iterative optimisation algorithm, that was first applied to the analysis of reflectometry and diffraction data by Wormington *et al.*<sup>93</sup> Since then, it has proven

<sup>91</sup> G. Wu *et al.* *Swarm Evol. Comput.* 44 (2019), pp. 695–711.

<sup>92</sup> DE.

<sup>93</sup> Wormington *et al.*, see n. 28.

<sup>94</sup> M. Björck. *J. Appl. Crystallogr.* 44.6 (2011), pp. 1198–1204; M. Björck *et al.* *J. Appl. Crystallogr.* 40.6 (2007), pp. 1174–1178; Nelson, “Co-Refinement of Multiple-Contrast Neutron/X-Ray Reflectivity Data Using MOTOFIT”, see n. 24; Nelson *et al.*, see n. 29; F. Ott. *SimulReflec*. URL: <http://www-l1b.cea.fr/prism/programs/simulreflec/simulreflec.html> (Accessed 2019-3-4); P. A. Kienzle *et al.* *NCNR Reflectometry Software*. URL: <http://www.ncnr.nist.gov/reflpak> (Accessed 2018-3-4).

<sup>95</sup> Storn *et al.*, see n. 51.

<sup>96</sup> J. H. Holland. *Adaptation in Natural and Artificial Systems*. 1992.

<sup>97</sup> Many of these exist however discussion will be limited to a simple classical trial method, details of other methods may be found in Björck, see n. 94.

very popular for the optimisation of reflectometry data and is included in many common analysis programs.<sup>94</sup> The DE algorithm is designed to more ably determine the global minimum of a particular function.<sup>95</sup>

DE is an example of a genetic algorithm, one that is designed to mimic the evolution processes observed in biology.<sup>96</sup> The method consists of two vectors, the parent population,  $\mathbf{p}$ , the offspring population,  $\mathbf{o}$ . These vectors are of a dimension  $(i \times j)$ , where  $i$  is the number of variables being optimised and  $j$  is the number of candidate solutions being used. The offspring population vector is created through some trial methods.<sup>97</sup>

A classical trial method consists of two stages, mutation and recombination. The mutation stage involves performing some mutation on the parent population to create a mutant vector,  $\mathbf{m}$ , analogous to the mutation in biological evolutionary theory. The magnitude of the mutation is dependent on the mutation constant,  $k_m$ ,

$$\mathbf{m}_{i,j} = b_i + k_m(\mathbf{p}_{i,R1} - \mathbf{p}_{i,R2}), \quad (2.55)$$

where  $b_i$  is the best candidate solution in the parent population, and  $\mathbf{p}_{i,R1}$  and  $\mathbf{p}_{i,R2}$  are randomly chosen members of the parent population. The mutation constant can be considered as a control variable for the size of the search radius, with a large  $k_m$  corresponding to a larger search radius.

The recombination step creates the offspring population vector by taking a sample from either the parent population or mutant vectors with some frequency, which depends on the recombination constant,  $k_r$ ,

$$\mathbf{o}_{i,j} = \begin{cases} \mathbf{m}_{i,j}, & \text{where } X < k_r \\ \mathbf{p}_{i,j}, & \text{otherwise} \end{cases} \quad (2.56)$$

where,  $X \sim U[0, 1]$ . The recombination constant controls the progress of the algorithm as it impacts the frequency with which mutation is introduced into the offspring population vector.

The final stage is to compare the offspring and parent population vectors, in the selection stage to create the new parent population for the next iteration. The selection stage comprises of using some figure of merit,  $\zeta$ , to choose between the subunit from the offspring or parent population vector.

$$\mathbf{p}_{*,j} \leftarrow \begin{cases} \mathbf{o}_{*,j}, & \text{where } \zeta_{\mathbf{o}_{*,j}} < \zeta_{\mathbf{p}_{*,j}} \\ \mathbf{p}_{*,j}, & \text{otherwise} \end{cases} \quad (2.57)$$

where, the \* notation indicates all objects in the given population, and  $\zeta_{\mathbf{o}_{*,j}}$  and  $\zeta_{\mathbf{p}_{*,j}}$  are the figures of merit for the offspring and population candidate solutions respectively. In our example, that figure of merit may be the agreement between some experimental data and our model, or for the example in Figure 2.21 it is the value of the Ackley function,<sup>98</sup> which is being minimised. The Ackley

<sup>98</sup> D. H. Ackley. “A Connectionist Machine for Genetic Hillclimbing”. PhD. Michigan, US: University of Michigan, 1987. 260 pp.

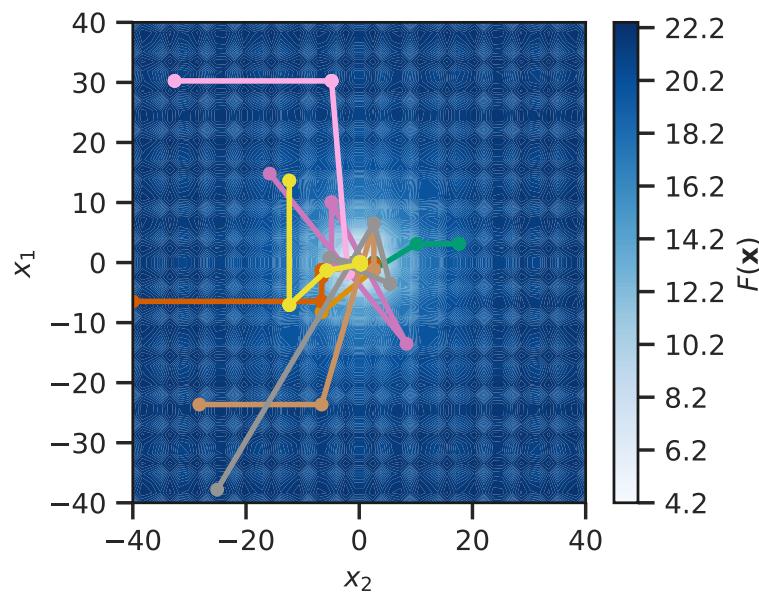


Figure 2.21: An example of a DE algorithm as applied to an Ackley function, where  $a = 20$ ,  $b = 0.2$ , and  $c = 2\pi$ . The mutation and recombination constant in this implementation are both 0.5. Each different coloured line represents a different candidate solution. The optimisation was stopped after 100 iterations had run.

function is a common function used for assessing the utility of global optimisation functions, which has the following form in the two-dimensional case,

$$f(x, y) = -a \exp \left[ -b \sqrt{0.5(x^2 + y^2)} \right] - \exp [0.5(\cos cx + \cos cy)] + e + a, \quad (2.58)$$

where,  $a$ ,  $b$ , and  $c$  are constants defined by the user, and  $e$  is the base of the natural logarithm.

It is noted that it is often the case,<sup>99</sup> that there should be some bounds applied to the variables within the populations. However, the DE algorithm may disregard these bounds due to the nature of the mutation step. Therefore, it is common in DE algorithms, where bounds must be set, that if the search space moves outside that expected it is necessary to reinitialise the parameter. An implementation of the DE algorithm is given programmatically in Code Block 2.3,<sup>100</sup> where this reinitialisation is achieved by obtaining a new random number within the given bounds.

<sup>99</sup> In particular for the optimisation of experimental data.

#### 2.4.2 Particle swarm

Particle swarm optimisation<sup>101</sup> is a type of swarm intelligence population-based optimisation method. This optimisation method was originally developed by Kennedy, Eberhart, and Shi.<sup>102</sup> Particle swarm methods are particularly suitable for the optimisation, and sampling, of parametric search-spaces with a large number of similar minima. Therefore, I believe that it will be useful for the study of the self-assembly of soft matter materials.<sup>103</sup>

These methods consist of a population vector, similar to that described for the DE, that moves around the parametric search-space.

<sup>100</sup> Additional Code Blocks showing the mutation, recombination, selection steps may be found in Appendix A.

<sup>101</sup> PSO.

<sup>102</sup> Kennedy *et al.*, see n. 50; Shi *et al.*, see n. 50, the initial purpose of the algorithm was to simulate social organisms such as bird flocks.

<sup>103</sup> This is the focus of Chapter 5.

<sup>104</sup> R. Poli. *J. Artif. Evol. Appl.* 2008 (2008), pp. 1–10.

The motions of these “particles” are influenced by the positions of the other particles in the vector.<sup>104</sup> It is anticipated that this will lead the swarm to optimise the function under investigation.

Particles in the swarm are under the influence of two elastic forces. The first attracts the particle to the best location in the search space that the particular particle has found, while the other attracts the particle to the best search-space location found by any particle of the swarm. The magnitudes of these forces are randomised but modulated by a pair of acceleration coefficients;  $\psi_p$  that influences the attraction towards the personal best location and  $\psi_g$  that influences the attraction to the global best location. The position of a particle changes between iterations of the algorithm based on the following relation,

$$\mathbf{p}_{*,j} \leftarrow \mathbf{p}_{*,j} + \mathbf{v}_{*,j}, \quad (2.59)$$

where,  $\mathbf{p}_{*,j}$  is the position of the particle, and  $\mathbf{v}_{*,j}$  is the velocity of the particle. This velocity is determined as shown below,

$$\mathbf{v}_{*,j} \leftarrow \omega \mathbf{v}_{*,j} + \psi_g R1(\mathbf{g}_* - \mathbf{p}_{*,j}) + \psi_p R2(\mathbf{s}_{*,j} - \mathbf{p}_{*,j}), \quad (2.60)$$

where,  $\omega$  a constant known as the interia weight,  $R1 \sim U[0, 1]$  and  $R2 \sim U[0, 1]$  are random numbers,  $\mathbf{g}_*$  is the best position occupied by any particle in the swarm and  $\mathbf{s}_{*,j}$  is the personal best for the particle  $j$ .

Figure 2.22 shows an example of the PSO in action, applied to the Ackley function.<sup>105</sup> Code Block 2.4 shows a functional programmatic implementation of a PSO algorithm.

<sup>105</sup> Ackley, see n. 98.

<sup>106</sup> Abbreviated to MCMC.

<sup>107</sup> W. Krauth. *Statistical Mechanics: Algorithms and Computations*. Oxford Master Series in Statistical, Computational, and Theoretical Physics 13. 2006.

<sup>108</sup> D. S. Sivia et al. *Data Analysis: A Bayesian Tutorial*. 2006.  
Code Block 2.3: An example of a simple implementation for a DE algorithm as described in M. Björck. *J. Appl. Crystallogr.* 44.6 (2011), pp. 1198–1204. The input variables are `population` which is an array of floats containing the initial parent population, `f` which is the figure of merit function to be minimised, `km` which is the mutation constant, `kr` which is the recombination constant, `bounds` which is an array of floats giving the minimum and maximum values for the variables, and `max_iter` which is the maximum number of iterations that should be performed. This will return `history` which is a history of the variables that are being fit during the DE algorithm.

#### 2.4.3 Markov chain Monte-Carlo

Markov chain Monte Carlo<sup>106</sup> is a sampling methodology, derived from direct sampling Monte-Carlo.<sup>107</sup> The aim of an MCMC algorithm is to sample a probability distribution, when parameters are described in terms of their degree of probability.<sup>108</sup> Similar to molecular dynamics,<sup>109</sup> in practical terms, MCMC should not be

```
import numpy as np
import mutation as mut
import recombination as recomb
import selection as sel

def differential_evolution(population, f, km, kr, bounds, max_iter):
    history = np.array([population])
    best = population[:, np.argmax(f(population))]
    i = 0
    while i < max_iter:
        mutant = mut.mutation(population, best, km)
        offspring = recomb.recombination(population, mutant, kr)
        offspring[
            np.where(offspring >= bounds[1])
            | np.where(offspring < bounds[0])
        ] = np.random.uniform(bounds[0], bounds[1], 1)
        selected = sel.selection(population, offspring, f)
        history = np.append(history, selected)
        history = np.reshape(
            history, (i + 2, population.shape[0], population.shape[1])
        )
        population = np.array(selected)
        best = population[:, np.argmax(f(population))]
        i += 1
    return history
```

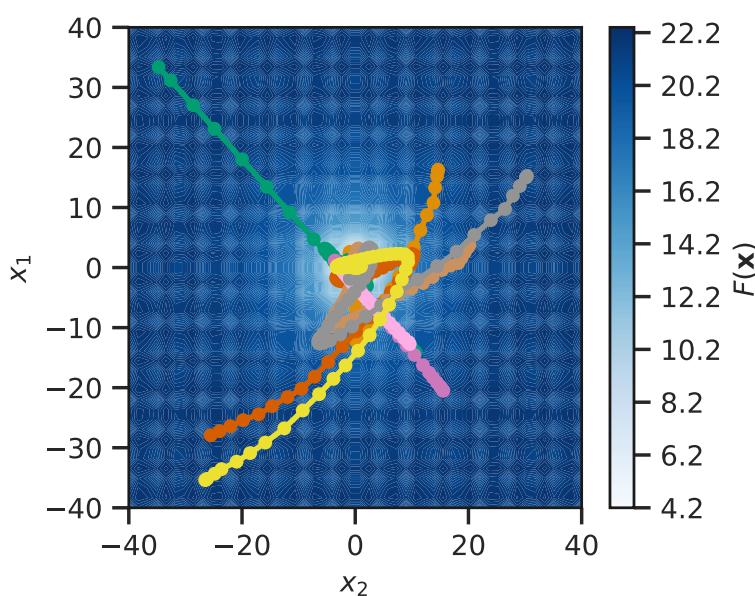


Figure 2.22: An example of a PSO as applied to an Ackley function, where  $a = 20$ ,  $b = 0.2$ , and  $c = 2\pi$ . For the particle swarm, the following parameters were used  $\omega = 0.9$ ,  $\psi_g = 0.05$ , and  $\psi_p = 0.05$ . Each different coloured line represents a different candidate solution. The optimisation was stopped after 100 iterations had run.

used on a system that is not already optimised, as its purpose is probability distribution sampling rather than minimisation. Generally, the approach would be to optimise using, for example, one of the approaches described above, then to use MCMC or MD to sample the appropriate search-space. For example, in this work MCMC is used following the optimisation of a reflectometry model using a DE algorithm, to quantify the inverse uncertainties of the model.<sup>110</sup> In addition to being able to give information about the inverse uncertainties, MCMC also offers a more complete understanding of the correlations present between the different parameters,<sup>111</sup> as the interactions between the parameter variation has been quantified.

The aim of MCMC is to only sample configurations of a given function that are within the experimental uncertainty. Figure 2.23 shows an example of the possible output that may be obtained from the application of an MCMC sampling method. This was generated using a Metropolis-Hastings MCMC algorithm,<sup>112</sup> shown in Code Block 2.5. Initially, a Levenberg–Marquardt algorithm<sup>113</sup> was used to optimise the positions and integral of the two Gaussian functions that make up the data. The MCMC was used to sample the values that were within the experimental uncertainty.

Once an optimised solution,  $\theta$ , is obtained, the figure of merit is calculated, in Code Block 2.5 this is the agreement between the model and the experimental data,  $\chi^2$ , where,

$$\chi^2 = \sum \frac{(y_{\text{exp}} - y_{\text{calc}})^2}{dy_{\text{exp}}}, \quad (2.61)$$

and  $y_{\text{exp}}$  is the experimental data, and  $dy_{\text{exp}}$  the uncertainty in the experimental data, while  $y_{\text{calc}}$  is the model solution. Some random

<sup>110</sup> This is the name given to the uncertainties in the parameters fitted in the modelling process.

<sup>111</sup> W. Gilks *et al.* *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. 1995.

<sup>112</sup> N. Metropolis *et al.* *J. Chem. Phys.* 21.6 (1953), pp. 1087–1092; W. K. Hastings. *Biometrika* 57.1 (1970), pp. 97–109.

<sup>113</sup> K. Levenberg. *Quart. Appl. Math.* 2.2 (1944), pp. 164–168; D. W. Marquardt. *J. Soc. Indust. Appl. Math.* 11.2 (1963), pp. 431–441.

Code Block 2.4: An example of the PSO algorithm from R. Poli. *J. Artif. Evol. Appl.* 2008 (2008), pp. 1–10. The input variables are position which is the initial position vector, f which is the figure of merit function to be minimised, omega which is the interia weight, psig which is the global acceleration constant, psip which is the personal acceleration constant, and the max\_iter which is the maximum number of iterations that should be performed. This will return the history which is a history of the variables that are being fit during the PSO.

```
import numpy as np

def particle_swarm(position, f, omega, psig, psip, max_iter):
    history = np.array([position])
    velocity = np.zeros_like(position)
    g_best = position[:, np.argmin(f(position))]
    p_best = np.array(position)
    i = 0
    while i < max_iter:
        for j in range(velocity.shape[1]):
            velocity[:, j] = (
                omega * velocity[:, j]
                + psig * np.random.rand() * (g_best - position[:, j])
                + psip
                * np.random.rand()
                * (p_best[:, j] - position[:, j])
            )
        position[:, j] = position[:, j] + velocity[:, j]
        history = np.append(history, position)
        history = np.reshape(
            history, (i + 2, position.shape[0], position.shape[1]))
        test_g_best = position[:, np.argmin(f(position))]
        if f(test_g_best) < f(g_best):
            g_best = test_g_best
        test_p_best = np.array(position)
        for j in range(position.shape[1]):
            if f(test_p_best[:, j]) < f(p_best[:, j]):
                p_best[:, j] = test_p_best[:, j]
        i += 1
    return history
```

perturbation is then applied to the optimised solution,

$$\Theta = \theta + aR, \quad (2.62)$$

where  $R \sim N(0, 1)$  and  $a$  is the step size. A new  $\chi^2$  is found for  $\Theta$ , and the probability that this transition will occur is found,

$$p = \exp\left(\frac{-\chi^2(\Theta) + \chi^2(\theta)}{2}\right). \quad (2.63)$$

This probability is then compared with a random number  $n \sim U[0, 1]$ , and if  $n$  is less than the probability, the new solution is stored,

$$\theta \leftarrow \Theta. \quad (2.64)$$

This process is repeated until some desired number of samples has been obtained. It should be noted that in the event on a poorly optimised initial value of  $\theta$ , it may be necessary to “burn”<sup>114</sup> the first series of solutions while the MCMC algorithm settles into the search-space.

<sup>114</sup> This means to ignore.

Code Block 2.5: An example of the Metropolis-Hastings MCMC algorithm from N. Metropolis *et al. J. Chem. Phys.* 21.6 (1953), pp. 1087–1092; W. K. Hastings. *Biometrika* 57.1 (1970), pp. 97–109. The input variables are theta which is an array of floats giving the initial values for the variables, f which is the figure of merit function to be minimised, a which is the step size for the changes, data which is the experimental data, iterations which is the number of accepted iterations to obtain, and nburn which is the number of accepted iterations to ignore in the burn-in phase. This will return the history which is a history of the variables that are being fit during the PSO.

```
import numpy as np

def mcmc(theta, f, a, data, iterations, nburn):
    accepted = np.array([])
    calc_y = f(data[0], theta)
    chi2 = np.sum(np.square((data[1] - calc_y) / data[2]))
    i = 0
    while i < iterations:
        new_theta = theta + a * np.random.randn(theta.size)
        new_calc_y = f(data[0], new_theta)
        new_chi2 = np.sum(np.square((data[1] - new_calc_y) / data[2]))
        prob = np.exp((-new_chi2 + chi2) / 2)
        n = np.random.rand()
        if n < prob:
            i += 1
            theta = new_theta
            chi2 = new_chi2
            if i > nburn:
                accepted = np.append(accepted, theta)
                accepted = np.reshape(
                    accepted, (i - nburn, theta.size))
    return accepted
```

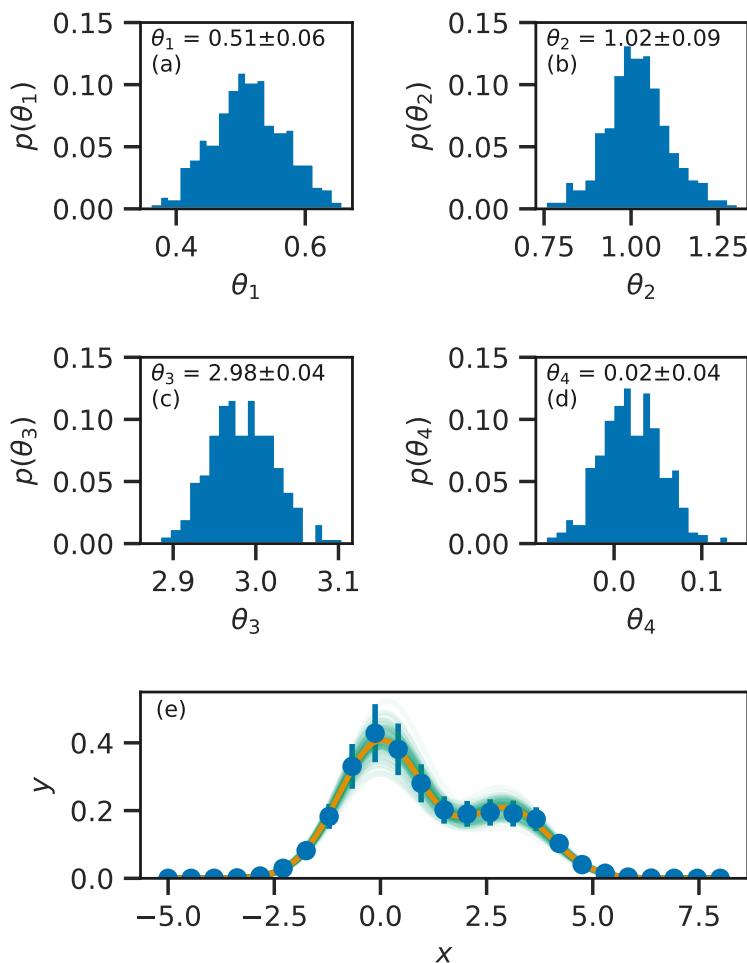


Figure 2.23: An example of a four variable (two nearby Gaussian functions of different sizes with added random noise and some fractional uncertainty) problem probed using a MCMC method, using values of  $a = 0.1$ ,  $\theta_1$  and  $\theta_2$  correspond to the integral of the Gaussian function, while  $\theta_3$  and  $\theta_4$  indicate their positions; (a)-(d) histograms of the probability distribution function for each of the variables, and (e) the data (blue circles), the optimised solution (orange line), and a series of probable solutions (green lines) showing the variability present in the data uncertainty.

#### 2.4.4 Molecular dynamics

Section 2.3 introduced classical potential models as a method for the evaluation of the interaction energy of a given chemical system. Any of the optimisation methods discussed above could be used alongside these classical potential models to find an energy minimum structure for the system or to sample the potential energy landscape. However, it is often the case that a thermodynamically relevant structure is of interest at a given temperature. This is where MD simulations are a useful and important tool.

The aim of an MD simulation is to probe the positions, velocities, and accelerations on each of the atoms, or coarse-grained particles, as a simulation progresses. The acceleration on a given particle,  $\mathbf{a}$  is defined by the force on that particle,  $\mathbf{f}$ , in agreement with Newton's second law of motion,

$$\mathbf{f} = m\mathbf{a}, \quad (2.65)$$

where,  $m$  is the mass of the particle. In order to determine the acceleration on the particle, it is necessary to know the force on that

particle. The force,  $f$ , is a function of the potential energy,  $E$ , as found from a classical potential, of that atom,

$$f(r) = \frac{-\delta E_{\text{total}}(r)}{\delta r}, \quad (2.66)$$

<sup>115</sup> The force is the negative of the first derivative of the energy with respect to the atomic configuration.

where,  $r$  is the configuration of the atoms.<sup>115</sup> The force found from Equation 2.66 is a scalar, however, the force vector is present in Equation 2.65. To determine the force in a given direction, it is necessary to find the product of the force,  $f$ , and the unit vector in that direction,

$$\mathbf{f}_x = f \hat{\mathbf{r}}_x, \quad \text{where } \hat{\mathbf{r}}_x = \frac{\mathbf{r}_x}{|\mathbf{r}|}, \quad (2.67)$$

where  $\mathbf{r}_x$  is the atomic configuration in the  $x$ -dimension, and  $|\mathbf{r}|$  is the magnitude of the atomic configuration vector.

The potential model, which is defined for a given system, allows for the calculation of the acceleration on each particle in that system. The next step is to use this acceleration to iterate through the trajectory of our system. This is achieved by applying Newtonian equations of motion, for example in the Velocity-Verlet algorithm.<sup>116</sup>

<sup>116</sup> W. C. Swope *et al.* *J. Chem. Phys.* 76.1 (1982), pp. 637–649.

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2, \quad (2.68)$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{1}{2}[\mathbf{a}(t) + \mathbf{a}(t + \Delta t)]\Delta t, \quad (2.69)$$

where,  $\mathbf{x}$  is the position the particle  $\mathbf{v}$  is the particle's velocity, and  $\mathbf{a}$  is the particle's acceleration, while  $t$  is current simulation time and  $\Delta t$  is the timestep. These equations constitute the Velocity-Verlet algorithm,

1. calculate the force, and therefore the acceleration, on each particle,<sup>117</sup>
2. find the position of the particle after some timestep,<sup>118</sup>
3. determine the new velocity for each particle, based on the average acceleration at the current and new positions,<sup>119</sup>
4. overwrite the old acceleration values with the new ones,
5. go to 1.

Following an equilibration period, this algorithm may be iterated as many times as is required to obtain sufficient statistics for the measurement quantity of interest, e.g. particle positions for structural techniques such as elastic scattering.

The above analytical process is known as the integration step, and the Velocity-Verlet is the integrator. If the size of the timestep  $\Delta t$  is too large, the step size for a given iteration will not be accurate, as the forces on the atoms will change too significantly during it. Therefore, the values of the timestep are usually on the order

<sup>117</sup> Using Equations 2.65 & 2.66.

<sup>118</sup> Using Equation 2.68.

<sup>119</sup> Using Equation 2.69.

of  $10 \times 10^{-15}$  s.<sup>120</sup> This means that in order to simulate a single nanosecond of “real-time” MD, the integrator must be solved one million times. This can be slow for very large systems, leading to an interest in coarse-grained simulations that result in fewer particles to determine the forces for, but also enable to use of larger timesteps,<sup>121</sup> for example, the use of a MARTINI potential model allows for an up to twenty times increase in the timestep compared to an all-atom model.

The above discussion ignored two aspects that are necessary to run an MD simulation, both of which are associated with the original configuration of the system; the original particle positions and velocities. The particle positions are usually taken from some library, for example for the simulation of a protein, often the protein data bank<sup>122</sup> is a useful resource. Small molecules may be configured by hand using graphical programs such as Jmol.<sup>123</sup> These small molecules may be built into complex, multicomponent structures using software such as the Packmol package.<sup>124</sup> The importance of this initial structure cannot be overstated, for example, if the initial structure in an MD simulation is unrepresentative of the equilibrium structure, it may take a large amount of simulation time before the equilibrium structure is obtained.<sup>125</sup>

The initial particle velocities are obtained in a much more general fashion. They are selected randomly, and then scaled such that the kinetic energy,  $E_k$ , of the system agrees with a defined temperature,  $T$ ,

$$E_k = \sum_{i=1}^N \frac{m_i |\mathbf{v}_i|^2}{2} = \frac{3}{2} N k_B T, \quad (2.70)$$

where,  $m_i$  and  $\mathbf{v}_i$  are the masses and velocities of the particles,  $N$  is the number of particles, and  $k_B$  is the Boltzmann constant.

The above algorithm details a simulation that makes use of an NVE ensemble.<sup>126</sup> However, this is not the only simulation ensemble that is available, within this work two other ensembles have been used extensively,

- the NVT (canonical) ensemble; this is similar to the NVE ensemble except the simulation temperature is controlled via a thermostat,
- the NPT (isothermal-isobaric); this ensemble is similar to the NVT ensemble, however, the system volume is allowed to vary while the overall system pressure is held constant using a barostat.

Thermostating involves controlling the kinetic energy of the particles<sup>127</sup> such that the simulation temperature is kept at a predefined value. There are a variety of methods for thermostating a MD simulation, such as the Andersen, Nosé-Hoover, or Berendsen methods.<sup>128</sup> However, the most straightforward to describe, and that

<sup>120</sup> femtoseconds.

<sup>121</sup> R. E Rudd *et al.* *Phys. Rev. B* 58.10 (1998), R5893–R5896; E. Brini *et al.* *Soft Matter* 9.7 (2013), pp. 2108–2119, so fewer, faster integration steps must be solved.

<sup>122</sup> RCSB PDB: Protein Data Bank. URL: <http://www.rcsb.org> (Accessed 2018-1-28).

<sup>123</sup> Jmol: An Open-Source Java Viewer for Chemical Structures in 3D. URL: <http://www.jmol.org/> (Accessed 2018-1-28).

<sup>124</sup> L. Martínez *et al.* *J. Comput. Chem.* 30.13 (2009), pp. 2157–2164.

<sup>125</sup> This can be much longer than could be reasonably simulated.

<sup>126</sup> A simulation where the number of particles (N), the volume of the system (V), the energy of the system (E) are all kept constant.

<sup>127</sup> Using Equation 2.70.

<sup>128</sup> H. C. Andersen. *J. Chem. Phys.* 72.4 (1980), pp. 2384–2393; S. Nosé. *J. Chem. Phys.* 81.1 (1984), pp. 511–519; H. J. C. Berendsen *et al.* *J. Chem. Phys.* 81.8 (1984), pp. 3684–3690; W. G. Hoover. *Phys. Rev. A* 31.3 (1985), pp. 1695–1697.

Tool for Classical Atomistic Simulation";

A. R. McCluskey *et al.* *Arm61/Pylj:*

*Pylj-1.1.0.* 2018. URL: <http://doi.org/10.5281/zenodo.1403828>.

Discussed in detail in Chapter 6.

<sup>130</sup> G. Bussi *et al.* *J. Chem. Phys.* 126(1) (2007), p. 014101.

implemented in the *pylj* software,<sup>129</sup> is a velocity rescaling.<sup>130</sup> This is where the velocities for a random subset of the particles,  $\mathbf{v}_i$  are adapted based on the following relation,

$$\mathbf{v}_i \leftarrow \mathbf{v}_i \sqrt{\frac{T_{\text{target}}}{\bar{T}}} \quad (2.71)$$

where,  $T_{\text{target}}$  is the target temperature, and  $\bar{T}$  is the average simulation temperature.

The use of a barostat to control the simulation pressure usually involves varying the simulation cell parameters and the distances between the particles. In a similar way to thermostating, where the simulation dimensions are scaled by a value in an effort to control the pressure. The barostating methods are similar to the thermostating methods with Andersen, Nosé-Hoover, and Berendsen methods. However, there is also the Parrinello-Rahman barostat which allows for independent control of the different cell dimensions giving control of stress in addition to pressure.<sup>131</sup>

These optimisation and sampling methods were used in a variety of different applications within this work, firstly DE optimisation and MCMC sampling are used in Chapter 3 in the study of a chemically-consistent modelling approach to X-ray and neutron reflectometry analysis. MD simulation is investigated as a possible tool to assist in the analysis of reflectometry in Chapter 4. Finally, the PSO is applied for the efficient determination of a micelle structure for fitting SAS data in Chapter 5.

<sup>131</sup> M. Parrinello *et al.* *J. Appl. Phys.* 52(12) (1981), pp. 7182–7190.

# 3

## CHEMICALLY CONSISTENT MODELLING OF X-RAY AND NEUTRON REFLECTOMETRY

### ABSTRACT

The work discussed in this chapter is the first example of the use of a chemically-consistent reflectometry model to co-refine X-ray reflectometry<sup>1</sup> measurements at different surface pressures.<sup>2</sup> This was coupled with a differential evolution<sup>3</sup> optimisation and Markov chain Monte Carlo<sup>4</sup> sampling methodology in order to rationalise the model inverse uncertainties and correlations. This chemically-consistent modelling approach was applied to the study of phospholipid monolayers at the air-deep eutectic solvent<sup>5</sup> interface, which required that the head and tail group volumes were not constrained. By co-refining multiple experimental datasets, it was possible to accurately model the experimental data without these constraints present.

<sup>1</sup> Abbreviated to XRR.

<sup>2</sup> Abbreviated to SPs.

<sup>3</sup> DE.

<sup>4</sup> MCMC.

<sup>5</sup> Abbreviated to DES.

### CONTEXT

This project offers a severe method of coarse-graining for the analysis of neutron reflectometry<sup>6</sup> and XRR data. The system is coarse-grained to represent a head group and a tail group of a phospholipid species. There is a chemical constraint present in the model, such that the number of head groups must be equal to the number of pairs of tail groups. However, there is no potential model considered beyond this “bonded” interaction. Additionally, this modelling approach is applied again in Chapter 4, as an example of the cutting edge of traditional modelling, against which the classical simulation-driven methods are compared. The specific application of this modelling approach grew from a collaboration with experimental colleagues working on self-assembly in DES. Therefore, this chemical system will be briefly introduced in Section 3.1. However, the main focus of this chapter will be the modelling methodology.

<sup>6</sup> NR.

### 3.1 INTRODUCTION

#### 3.1.1 Deep eutectic solvents

<sup>7</sup> E. L. Smith *et al.* *Chem. Rev.* 114.21 (2014), pp. 11060–11082; Y. Dai *et al.* *Anal. Chim. Acta* 766 (2013), pp. 61–68.

<sup>8</sup> O. S. Hammond *et al.* *Green Chem.* 18.9 (2016), pp. 2736–2744; O. S. Hammond *et al.* *J. Phys. Chem. B* 121.31 (2017), pp. 7473–7483; C. F. Araujo *et al.* *Phys. Chem. Chem. Phys.* 19.27 (2017), pp. 17998–18009.

<sup>9</sup> A. Pandey *et al.* *Phys. Chem. Chem. Phys.* 16.4 (2014), pp. 1559–1568.

<sup>10</sup> Smith *et al.*, see n. 7.

<sup>11</sup> S. Zahn *et al.* *ChemPhysChem* 17.21 (2016), pp. 3354–3358.

<sup>12</sup> B. D. Ribeiro *et al.* *ACS Sustain. Chem. Eng.* 3.10 (2015), pp. 2469–2477; D. J. G. P. van Osch *et al.* *Green Chem.* 17.9 (2015), pp. 4518–4521.

<sup>13</sup> A. Sanchez-Fernandez *et al.* *Phys. Chem. Chem. Phys.* 18.48 (2016), pp. 33240–33249; T. Arnold *et al.* *Langmuir* 31.47 (2015), pp. 12894–12902; Y.-T. Hsieh *et al.* *Langmuir* 34.35 (2018), pp. 10270–10275; M. K. Banjare *et al.* *RSC Adv.* 8.15 (2018), pp. 7969–7979.

<sup>14</sup> S. J. Bryant *et al.* *Soft Matter* 12.6 (2016), pp. 1645–1648; S. J. Bryant *et al.* *Langmuir* 33.27 (2017), pp. 6878–6884; M. G. Gutiérrez *et al.* *Langmuir* 25.10 (2009), pp. 5509–5515.

<sup>15</sup> L. Sapir *et al.* *J. Phys. Chem. A* 120.19 (2016), pp. 3253–3259.

<sup>16</sup> A. Sanchez-Fernandez *et al.* *Phys. Chem. Chem. Phys.* 19.13 (2017), pp. 8667–8670.

<sup>17</sup> Mohwald, see n. ??; Kewalramani *et al.*, see n. ??.

<sup>18</sup> Dai *et al.*, see n. 7; Hammond *et al.*, see n. 8.

<sup>19</sup> F. Graner *et al.* *J. Phys. II France* 5.2 (1995), pp. 313–322.

<sup>20</sup> S. P. Weinbach *et al.* *J. Phys. Chem.* 97.20 (1993), pp. 5200–5203; O. M. Magnussen *et al.* *Nature* 384 (1996), pp. 250–252; H. Kraack *et al.* *Science* 298.5597 (2002), pp. 1404–1407.

DES are a class of green, sustainable liquids that may be obtained from the combination of ionic species with compounds capable of acting as hydrogen bond donors, such as sugar, alcohols, amines, and carboxylic acids.<sup>7</sup> The resulting extensive hydrogen bonding network is capable of stabilising both species, such that the eutectic mixture will remain liquid at room temperature.<sup>8</sup> Using different precursor materials can allow for the ability to tune the resulting solvent's physicochemical properties, such as polarity,<sup>9</sup> viscosity and surface tension,<sup>10</sup> network charge,<sup>11</sup> and hydrophobicity.<sup>12</sup> Recently DES have also been shown to exhibit a "solvophobic" effect through the promotion of surfactant micelle formation,<sup>13</sup> phospholipid bilayer formation,<sup>14</sup> and the ability to stabilise non-ionic polymer<sup>15</sup> and protein conformations.<sup>16</sup>

Phospholipid monolayers at the air/water interface have been widely studied as simplistic models for biological membranes. As such, they have been used to gain insight into many biological processes that are technologically and medically relevant. For example, investigations at the air/salt-water interface have identified the importance that interactions between charged phospholipid head groups and ions present in solution have on the structure, monolayer packing and stability.<sup>17</sup> However, the native environment for phospholipids in-vivo is far from simple aqueous solutions. In fact, it has been suggested<sup>18</sup> that DES might form within the crowded cellular environment and could assist in solubilizing biological species in an intermediate environment between that of the hydrophobic phospholipid tail groups and the highly polar water-rich regions, thereby assisting survival under extreme conditions such as freezing temperatures and drought where the water content of the cells is restricted.

This chapter presents the first observation of phospholipid monolayer at an air-DES interface. Furthermore, this is one of a few examples of a phospholipid monolayer at the interface between air and a non-aqueous solvent, with only formamide noted previously.<sup>19</sup> Langmuir monolayers of non-phospholipidic surfactant molecules have also been noted at air-formamide and air-mercury interfaces.<sup>20</sup> In these previous works, the authors noted that the non-aqueous surface had an effect on the overall structure of the monolayer, but little was said about the underlying mechanism.

#### 3.1.2 Optimisation and sampling in reflectometry analysis

The analysis of reflectometry data usually involves the use of some model-dependent methodology. Therefore it is necessary to optimise the difference between our model, and the experimental dataset. Analytical methods, such as the gradient descent method

are not usually suitable for application to the optimisation of a reflectometry model, as these are only capable of optimisation to local minima, which would require accurate prior knowledge of the model structure.<sup>21</sup> Despite the analytical nature of the Maximum entropy (MaxEnt) optimisation method, this showed some success in the optimisation of reflectometry models, due in part to the ability of this optimisation to produce of a large number of solutions.<sup>22</sup> However, this approach is computationally intensive and therefore it is unusual to apply it instead of other more efficient methods. Some aspects of the MaxEnt methods were replicated in the work of Sivia *et al.*, which employed Bayesian probability theory to rationalise the model selection.<sup>23</sup>

The use of analytic methods became less favoured as more frequently stochastic methods were used, these offer a more pragmatic solution to the local minima problem. Stochastic methods are those that utilise inherently random behaviour to determine a global minimum. The groove tracking method of Zhou and Chen,<sup>24</sup> was one of the first examples of a stochastic optimisation process applied to the analysis of reflectometry data. This randomly varied the SLD of the layers in the model using a Monte Carlo approach. A similar approach used a simulated annealing approach, with a “temperature” factor that decreased as the number of iterations increased,<sup>25</sup> however, this approach is still subject to the local minima problem as the probability of move acceptance decreased over time.

Both these Monte Carlo based approaches and the analytic methods previously discussed make the same perturbations to the fitted parameters during processing. However, this is the cause of the propensity to converge to a local minimum. This led to the application of genetic algorithm-derived methods for the optimisation of reflectometry models, beginning with the works of de Haan and Drikkonen<sup>26</sup> and Dane *et al.*<sup>27</sup> These methods are designed to stochastically sample an entire search-space, and therefore are more able to overcome the local minima issues and determine the vicinity of a global minimum. Following this initial application, genetic algorithms were used frequently in the optimisation of reflectometry models<sup>28</sup> In particular, the work of Wormington *et al.*<sup>29</sup> showed the applicability of the DE method towards reflectometry model optimisation, which resulted in the inclusion of such methods in many common reflectometry analysis software packages.<sup>30</sup>

The use of MCMC methods to probe the probability distribution functions<sup>31</sup> of the fitting parameters of a reflectometry model have also grown in popularity.<sup>32</sup> This is due to the inclusion of MCMC methods in common analysis software packages such as ReflID.<sup>33</sup> These methods enable the user to better understand the inverse uncertainties of the model. Additionally, they enable the quantification of the correlation between parameters, important in ensuring that the model applied is suitably constrained such as to

<sup>21</sup> M. R. Lovell *et al.* *Curr. Opin. Colloid Interface Sci.* 4.3 (1999), pp. 197–204.

<sup>22</sup> M. Geoghegan *et al.* *Phys. Rev. E* 53.1 (1996), pp. 825–837; D. G. Bucknall *et al.* *Physica B* 241–243 (1997), pp. 1071–1073.

<sup>23</sup> Geoghegan *et al.*, see n. 22; D. S. Sivia *et al.* *Physica D* 66.1–2 (1993), pp. 234–242; D. S. Sivia *et al.* *Physica B* 248.1–4 (1998), pp. 327–337; D. S. Sivia *et al.* *Physica B* 173.1–2 (1991), pp. 121–138.

<sup>24</sup> Xiao-Lin Zhou *et al.* *Physical Review E* 47.5 (1993), pp. 3174–3190; X.-L. Zhou *et al.* *Phys. Rep.* 257 (1995), pp. 223–348.

<sup>25</sup> K. Kunz *et al.* *Macromolecules* 26.16 (1993), pp. 4316–4323.

<sup>26</sup> V.-O. de Haan *et al.* *Physica B* 198.1–3 (1994), pp. 24–26.

<sup>27</sup> A. D. Dane *et al.* *Physica B* 253.3–4 (1998), pp. 254–268.

<sup>28</sup> A. Ulyanenkov *et al.* *Physica B* 283 (2000), pp. 237–241; A. Ulyanenkov *et al.* *J. Phys. D* 38 (10A 2005), A235–A238; E. Politsch *et al.* *J. Appl. Crystallogr.* 35.3 (2002), pp. 347–355; Wormington *et al.*, “Characterization of Structures from X-Ray Scattering Data Using Genetic Algorithms”; Characterization of Structures from X-Ray Scattering Data Using Genetic Algorithms”.

<sup>30</sup> Björck, “Fitting with Differential Evolution: An Introduction and Evaluation”; Björck *et al.*, “GenX: An Extensible X-Ray Reflectivity Refinement Program Utilizing Differential Evolution”; Nelson, “Genetic Algorithms for Multiple-Configured Neutron/X-Ray Reflectivity Data Using Monte-Carlo PEPE”; “Reflex: Neutron and X-Ray Reflectivity Analysis in Python”; Ott, *Sigal Reflx*; Kienzle, “NMR Reflectometry Software”; *NMR Reflectometry*.

<sup>33</sup> Kienzle *et al.*, see n. 94.

<sup>34</sup> Nelson, "Co-Refinement of Multiple-Contrast Neutron/X-Ray Reflectivity Data Using MOTOFIT", see n. 24.

<sup>35</sup> Storn *et al.*, see n. 51; E. Jones *et al.* *SciPy: Open Source Scientific Tools for Python*. URL: <http://www.scipy.org> (Accessed 2019-3-4).

<sup>36</sup> D. Foreman-Mackey *et al.* *Publ. Astron. Soc. PAc.* 125.925 (2013), pp. 306–312.

<sup>37</sup> The history of this modelling is introduced well in Campbell *et al.*, see n. 26.

<sup>38</sup> K. Wojciechowski *et al.* *Langmuir* 32.35 (2016), pp. 9064–9073; K. Wojciechowski *et al.* *BBA - Biomembranes* 1858.2 (2016), pp. 363–373.

<sup>39</sup> F. Foglia *et al.* *Langmuir* 30.30 (2014), pp. 9147–9156; G. Bello *et al.* *BBA - Biomembranes* 1858.2 (2016), pp. 197–209.

<sup>40</sup> Campbell *et al.*, see n. 26.

<sup>41</sup> C. W. McConlogue *et al.* *Langmuir* 13.26 (1997), pp. 7158–7164; D. Small. *J. Lipids Res.* 25 (1984), pp. 1490–1500.

<sup>42</sup> As is discussed in Section 3.3.

<sup>43</sup> C. M. Hollinshead *et al.* *Langmuir* 25.7 (2009), pp. 4070–4077.

reduce the cross-correlation present between parameters.<sup>34</sup>

This work applies a DE algorithm<sup>35</sup> to the optimisation of the reflectometry model. The search-space, available within the experimental uncertainty of the data is then sampled using MCMC, as implemented in emcee,<sup>36</sup> to understand the parameter probability distributions and quantify the inter-parameter correlations.

### 3.1.3 Chemically-consistent modelling

The use of chemically-consistent modelling is common in the fitting of XRR and NR measurements from phospholipid monolayers to obtain structural insights.<sup>37</sup> Chemically-consistent modelling, in reflectometry analysis, is the reparameterisation of the layer-model into rational chemical terms, such as molecular volume and the elemental scattering lengths. This reparameterisation allows for chemically-realistic constraints to be applied to the model. While it is possible to model the NR from a phospholipid monolayer with a single layer model,<sup>38</sup> the use of at least two layers; representing the head and tail groups is more commonplace.<sup>39</sup> Even when two layers are utilised, it is often the case that the volumes of the head and tail groups,  $V_i$ , are used as constraints in the modelling process, as the SLD of a layer may be determined as follows,

$$\text{SLD}_i = \frac{b_i}{V_i} (1 - \phi_i) + \text{SLD}_s(\phi_i) \quad (3.1)$$

where,  $b_i$  is the scattering length of the head or tail,  $\phi_i$  is the volume fraction of solvation by the solvent,  $\text{SLD}_s$  is the solvent scattering length density, and  $i$  indicates either the head or tail layer. The scattering length for the head or tail of the phospholipid can be found based on the different atom types present in each. Therefore, the volume and solvation fraction are variables that may be fitted to give value for the SLD that may be used in Equation 2.34. However, as noted by Campbell *et al.*,<sup>40</sup> this method often fails to account for the compaction of the carbon chains under elevated SPs,<sup>41</sup> which may lead to a volume reduction of up to ~15 %. Furthermore,<sup>42</sup> the use of a constrained head group volume may also influence the result of the modelling process in situations where the volume is poorly defined.

Equation 3.1 enables the use of chemical-inference in the modelling approach for reflectometry data. This allows for the co-refinement of NR data where different isotopic-contrasts of the phospholipid species or solvent have been used. This is possible based on the expectation that the effect of contrast variation on the structure and chemistry of the monolayer will be negligible, and therefore the same values of all parameters in the fitting, except  $b_i$  and  $\text{SLD}_s$  may be constrained between the different measurements.<sup>43</sup> In this work, similar logic was applied, with the assumption that the volume of the head and tail groups is constant across different SPs, while the phospholipid phase is the same.

This means that, for this system, all of the fitted parameters may be constrained aside from the tail thickness, head solvation, and interfacial roughness, across the different SP measurements. This is the first time that such a methodology has been applied to the analysis of XRR and NR, additionally, it is believed that this ability to co-refine XRR measurements enables a greater understanding of the structure than that possible from a single measurement.

### 3.2 EXPERIMENTAL

My role in this work was entirely on the analysis of the measurements and the development of the chemically-consistent model.<sup>44</sup> However, it is necessary to briefly discuss the materials and experimental methods used to enable a complete understanding of the context of the work.

#### 3.2.1 Materials

Choline chloride (99 %, Sigma-Aldrich), glycerol (99 %, Sigma-Aldrich), d<sub>9</sub>-choline chloride (99 %, 98 % D, CK Isotopes), and d<sub>8</sub>-glycerol (99 %, 98 % D, CK Isotopes) were used in the preparation of the DES. This is achieved by mixing a 1:2 ratio of choline-chloride and glycerol and heating at 80 °C until a homogeneous, transparent liquid is formed.<sup>45</sup> This was then stored under a dry atmosphere to reduce the amount of water dissolved in the solvent.

The limited availability of deuterated precursors lead to only a fully protonated and a partially deuterated<sup>46</sup> being prepared and used in the NR measurements. The partially deuterated subphase was prepared using the following mixtures of precursors: 1 mol of 0.38 mol fraction of h-choline-chloride/0.62 mol fraction of d-choline-chloride; and 2 mol of 0.56 mol fraction of h-glycerol/0.44 mol fraction of d-glycerol.

The water content of the DES was assessed before and after each experiment by Karl-Fisher titration (Mettler Toledo DL32 Karl-Fischer Coulometer, Aqualine Electrolyte A, Aqualine Catholyte CG A) and found to be always below 0.3 wt%. This was taken to be a negligible amount and would not have a considerable impact on the DES characteristics.<sup>47</sup>

1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (C<sub>16</sub> tails, >99 %), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (C<sub>14</sub> tails, >99 %), and the sodium salt of 1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol) (C<sub>14</sub> tails, >99 %)<sup>48</sup> were obtained from Avanti Polar Lipids and 2-dilauroyl-*sn*-glycero-3-phosphocholine (C<sub>12</sub> tails, >99 %)<sup>49</sup> was obtained from Sigma Aldrich and all were used without further purification. Deuterated versions of DPPC (d<sub>62</sub>-DPPC, >99 %, deuterated tails-only) and DMPC (d<sub>54</sub>-DMPC, >99 %, deuterated tails-only) were obtained from Avanti Polar Lipids and used without further purification. These phospholipids

<sup>44</sup> The experimental measurements were designed and conducted by Drs Tom Arnold, Andrew Jackson, Adrian Sanchez-Fernandez, and Prof. Karen Edler, with the assistance of Dr Richard Campbell.

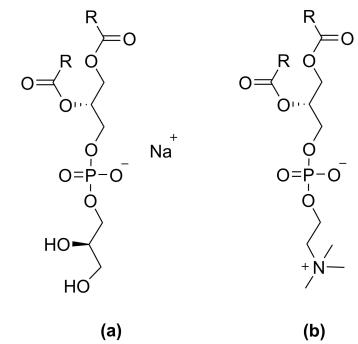


Figure 3.1: The two phospholipid forms investigated in this work, where R indicates the hydrocarbon tail; (a) phosphatidylglycerol (PG), (b) phosphocholine (PC).

<sup>45</sup> Smith *et al.*, see n. 7.

<sup>46</sup> Abbreviated to hDES and hdDES respectively.

<sup>47</sup> Hammond *et al.*, see n. 8; Hammond *et al.*, see n. 8.

<sup>48</sup> Abbreviated to DPPC, DMPC, and DMPG respectively.

<sup>49</sup> Abbreviated to DLPC.

<sup>50</sup> PC indicates where the phospholipid molecule contains a phosphocholine head group, while PG indicates a phosphatidylglycerol head group, the chemical structures of these can be seen in Figure 3.1.

<sup>51</sup> Abbreviated to PTFE

were dissolved in chloroform solution ( $0.5 \text{ mg mL}^{-1}$ ) at room temperature.<sup>50</sup>

In the XRR experiment, the sample was prepared in-situ using the standard method for the spreading of insoluble monolayers on water. A small amount of the phospholipid solution was spread on the liquid surface. Following the evaporation of the chloroform, it is assumed that the resulting system is a subsurface of solvent with a monolayer of the phospholipid at the interface. The surface concentration is then controlled by opening and closing the polytetrafluoroethylene<sup>51</sup> barriers of a Langmuir trough. To reduce the volume used in the NR experiments, a small Delrin adsorption trough was used that did not have controllable PTFE barriers. Therefore, although the surface concentration was nominally the same as for the XRR, the lack of precise control meant that it was determined to be inappropriate to co-refine the XRR and NR contrasts together.

### 3.2.2 Methods

The XRR measurements were carried out at the I07 beamline at the Diamond Light Source, with a photon energy of 12.5 keV using the double-crystal deflector system.<sup>52</sup> The reflected intensity was measured for a  $q$  range of  $0.018\text{--}0.7 \text{ \AA}^{-1}$ . The data were normalised with respect to the incident beam and the background was measured from off-specular reflection and subsequently subtracted. All of the samples were allowed at least one hour to equilibrate and preserved under an argon atmosphere. XRR data were collected for each of the phospholipids, DLPC, DMPC, DPPC, and DMPG at four SPs each,<sup>53</sup> as measured with an aluminium Wilhelmy plate; measurements were conducted at 7 and 22 °C. An aluminium Wilhelmy plate was used over a traditional paper due to the low wettability of paper by the DES.

The NR experiments were performed on the FIGARO instrument at the Institut Laue-Langevin using time-of-flight methods.<sup>54</sup> Data were collected at two incident angles; 0.62 and 3.8°, providing a  $q$  range from  $0.005\text{--}0.18 \text{ \AA}^{-1}$ . Two SPs for each phospholipid and contrast were measured.<sup>55</sup> As with the XRR measurements, the samples were given at least one hour to equilibrate, kept under an inert atmosphere. All measurements were conducted at 22 °C.

## 3.3 DATA ANALYSIS

XRR and NR methods have a well documented history for the analysis of the structure of phospholipid monolayers at the air-water interface.<sup>56</sup> Typically these have involved using a model-dependent analysis method, however, the modelling approaches have varied significantly in the number of layers used, the shape of the layers, the use of interfacial roughness, the parameterisation of constraints

<sup>52</sup> T. Arnold *et al.* *J. Synchrotron Radiat.* 19.3 (2012), pp. 408–416.

<sup>53</sup> DLPC: 20, 25, 30 and  $35 \text{ mN m}^{-1}$ , DMPC: 20, 25, 30 and  $40 \text{ mN m}^{-1}$ , DPPC: 15, 20, 25 and  $30 \text{ mN m}^{-1}$ , DMPG: 15, 20, 25 and  $30 \text{ mN m}^{-1}$ .

<sup>54</sup> R. A. Campbell *et al.* *Eur. Phys. J. Plus* 126.11 (2011), p. 107.

<sup>55</sup> DMPC: 20 and  $25 \text{ mN m}^{-1}$ , DPPC: 15 and  $20 \text{ mN m}^{-1}$ .

<sup>56</sup> Mohwald, see n. ??; Kewalramani *et al.*, see n. ??; T. M. Bayerl *et al.* *Biophys. J.* 57.5 (1990), pp. 1095–1098; S. J. Johnson *et al.* *Biophys. J.* 59.2 (1991), pp. 289–294; L. A. Clifton *et al.* *Phys. Chem. Chem. Phys.* 14.39 (2012), p. 13569; C. A. Helm *et al.* *EPL* 4.6 (1987), pp. 697–703; J. Daillant *et al.* *EPL* 12.8 (1990), pp. 715–720.

employed, and even the method by which the reflectometry profile was calculated from the model. Recently, an evaluation of the applicability of different models for surfactant and phospholipid monolayers using NR outlined a view of “best practice”.<sup>57</sup> However, frequently the constraints employed in the modelling process include the head and tail volume for the phospholipid head and tail groups. These values are taken from a variety of other techniques, some examples are shown in Table 3.1.

Table 3.1 provides a general consensus that the volume of the PC head group is 320–360 Å<sup>3</sup>, while the PG head group is 289–291 Å<sup>3</sup>. However, these values were all determined from experiments<sup>58</sup> or simulations<sup>59</sup> where the head group was interacting with water molecules. It is not clear if this will influence the volume that it occupies, and if that volume will change in the presence of a non-aqueous solvent.<sup>60</sup> The charged nature of the zwitterionic and anionic phospholipid head groups may have different interactions with the polar, but neutral water and the charged DES.<sup>61</sup> Additionally, it is known that, on water, increased SPs and the associated Liquid-Expanded to Liquid-Condensed phase transition will lead to a compression of the phospholipid tail volume, compared to the values in Table 3.1,<sup>62</sup> and that this compaction has not necessarily been accounted for in the literature.<sup>63</sup>

These factors meant that it was necessary to develop a model that was appropriate for the phospholipid chemistry while applying as much of the “best practice” from Campbell *et al.*<sup>64</sup> as possible, and ensuring that the head and tail group volumes were not constrained parameters. The lack of having these normally constrained parameters meant that it was necessary to consider methods by which the reflectometry measurements could be co-refined, in a similar fashion to contrast variation co-refinement in NR. This could be achieved by the co-refinement of reflectometry measurements at different SPs, as the model was appropriate for the phospholipid chemistry, and the different SPs were in the same phase.<sup>65</sup> Therefore the head and tail group volumes remain constant, and only the surface concentration and tail thickness vary.

The chemically-constrained model that has been used in this work was implemented in the Python library `refnx`.<sup>66</sup> The software enables the inclusion of custom model classes that feed parameters into the Abelès model.<sup>67</sup> Our chemically-consistent model class can be seen in Code Block 3.1, and is shared under a CC BY-SA 4.0 license in the ESI for the associated publication.<sup>68</sup> In order to ensure that the phospholipid chemistry was consistent both within the phospholipid molecule and across the different SPs, Code Block 3.2 was implemented.

The chemically-consistent model<sup>69</sup> consisted of two layers that define the phospholipid monolayer; the head layer at the interface with the solvent and the tail layer at the air interface. The head groups have a scattering length that can be calculated from a sum-

<sup>57</sup> Campbell *et al.*, see n. 26.

<sup>58</sup> W.-J. Sun *et al.* *Phys. Rev. E* 49.5 (1994), pp. 4665–4676; N. Kučerka *et al.* *Eur. Biophys. J.* 33.4 (2004), pp. 328–334; P. Balgavý *et al.* *Acta Phys. Slovaca* 51.1 (2001), pp. 53–68; J. Pan *et al.* *BBA - Biomembranes* 1818.9 (2012), pp. 2135–2148.

<sup>59</sup> R. S. Armen *et al.* *Biophys. J.* 75.2 (1998), pp. 734–744; N. Kučerka *et al.* *J. Phys. Chem. B* 116.1 (2012), pp. 232–239.

<sup>60</sup> Such as the DES considered herein.

<sup>61</sup> A. Sanchez-Fernandez *et al.* *Soft Matter* 14.26 (2018), pp. 5525–5536.

<sup>62</sup> D. Marsh. *Chem. Phy. Lipids* 163.7 (2010), pp. 667–677; Small, see n. 41.

<sup>63</sup> Campbell *et al.*, see n. 26.

<sup>64</sup> Campbell *et al.*, see n. 26.

<sup>65</sup> Liquid-Condensed (LC) for DPPC and Liquid-Expanded (LE) for DMPC, DLPC, and DMPG.

<sup>66</sup> Nelson *et al.*, see n. 29; A. R. J. Nelson *et al.* *Refnx v0.1.2*. 2019. URL: <http://doi.org/10.5281/zenodo.2552023>.

<sup>67</sup> Abelès, see n. 40; Parratt, see n. 40, discussed in detail in Section 2.2.6.

<sup>68</sup> A. R. McCluskey *et al.* *Lipids\_at\_airdes (Version 1.0)*. 2019. URL: <http://doi.org/10.5281/zenodo.2577796>.

<sup>69</sup> That is outlined in Code Block 3.1.

Table 3.1: Phospholipid component volumes extracted from different literature sources.  $V_t$  corresponds to the total phospholipid volume,  $V_h$  to the tail group volume,  $V_h'$  to the head group volume, MD to molecular dynamics simulations, WAXS to wide-angle X-ray scattering, NB to neutral buoyancy, and DVTD to differential vibrating tube densimetry. The values for DPPC are from R. S Armen *et al.* *Biophys. J.* 75.2 (1998), pp. 734–744, W.-J. Sun *et al.* *Phys. Rev. E* 49.5 (1994), pp. 4665–4676, and N Kučerka *et al.* *Eur. Biophys. J.* 33.4 (2004), pp. 328–334; P. Balagový *et al.* *Acta Phys. Slovaca* 51.1 (2001), pp. 53–68 respectively, the values for DMPC are from R. S Armen *et al.* *Biophys. J.* 75.2 (1998), pp. 734–744 and N. Kučerka *et al.* *Eur. Biophys. J.* 33.4 (2004), pp. 328–334; P. Balagový *et al.* *Acta Phys. Slovaca* 51.1 (2001), pp. 53–68 respectively, the values for DLPC are from R. S Armen *et al.* *Biophys. J.* 75.2 (1998), pp. 734–744 and N. Kučerka *et al.* *Eur. Biophys. J.* 33.4 (2004), pp. 328–334; P. Balagový *et al.* *Acta Phys. Slovaca* 51.1 (2001), pp. 53–68 respectively, the values for DMPG are from J. Pan *et al.* *BBA - Biomembranes* 1818.9 (2012), pp. 2135–2148, and the values for POPG are from N. Kučerka *et al.* *J. Phys. Chem. B* 116.1 (2012), pp. 232–239.

| Phospholipid        |                   | DPPC         |                   | DMPC              |                   | DLPC              |                   | DMPG   |      | POPG |  |
|---------------------|-------------------|--------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------|------|------|--|
|                     |                   |              |                   |                   |                   |                   |                   |        |      |      |  |
| $V_l/\text{\AA}^3$  | 1237.3 $\pm$ 25.5 | 1148 $\pm$ 2 | 1268.2 $\pm$ 32.1 | 1172.5 $\pm$ 25.1 | 1155.4 $\pm$ 30.0 | 1057.7 $\pm$ 24.7 | 1046.6 $\pm$ 28.0 | 1011.4 | 1203 |      |  |
| $V_t/\text{\AA}^3$  | 965.4 $\pm$ 5.4   | 829 $\pm$ 4  | 924.7 $\pm$ 17.6  | 851.5 $\pm$ 5.0   | 815.9 $\pm$ 15.5  | 736.8 $\pm$ 4.6   | 707.1 $\pm$ 13.5  | 720.4  | 914  |      |  |
| $V_h/\text{\AA}^3$  | 320.9 $\pm$ 20.1  | 319 $\pm$ 6  | 339.5 $\pm$ 14.5  | 320.9 $\pm$ 20.1  | 339.5 $\pm$ 14.5  | 320.9 $\pm$ 20.1  | 339.5 $\pm$ 14.5  | 291.0  | 289  |      |  |
| Method              | MD                | WAXS         | NB                | MD                | NB                | MD                | NB                | DVTD   | MD   |      |  |
| T/ $^\circ\text{C}$ | 50                | 24           | 30                | 50                | 30                | 50                | 30                | 20     | 20   |      |  |

```

import numpy as np
from refnx.analysis import possibly_create_parameter as pcp
from refnx.analysis import Parameters
from refnx.reflect import Component

class VolMono(Component):
    def __init__(self, vol, b, d_h, c_length, name=""):
        super(VolMono, self).__init__()
        t_t = 1.54 + 1.265 * c_length
        self.vol = [
            pcp(vol[0], "{}-V_h".format(name)),
            pcp(vol[1], "{}-V_t".format(name)),
        ]
        self.realb = [
            pcp(b[0].real, name="{}-b_h".format(name)),
            pcp(b[1].real, name="{}-b_t".format(name)),
        ]
        self.imagb = [
            pcp(b[0].imag, name="{}-ib_h".format(name)),
            pcp(b[1].imag, name="{}-ib_t".format(name)),
        ]
        self.d = [
            pcp(d_h, name="{}-d_h".format(name)),
            pcp(t_t * 0.8, name="{}-d_t".format(name)),
        ]
        self.phi = [
            pcp(0.5, name="{}-phi_h".format(name)),
            pcp(0.0, name="{}-phi_t".format(name)),
        ]
        self.sigma = pcp(3.0, name="{}-sigma".format(name))
        self.name = name

    @property
    def slabs(self):
        layers = np.zeros((2, 5))
        layers[0, 0] = self.d[1]
        layers[0, 1] = self.realb[1] * 1.0e16 / self.vol[1]
        layers[0, 2] = self.imagb[1] * 1.0e16 / self.vol[1]
        layers[0, 3] = self.sigma
        layers[0, 4] = self.phi[1]
        layers[1, 0] = self.d[0]
        layers[1, 1] = self.realb[0] * 1.0e16 / self.vol[0]
        layers[1, 2] = self.imagb[0] * 1.0e16 / self.vol[0]
        layers[1, 3] = self.sigma
        layers[1, 4] = self.phi[0]
        return layers

```

mation of the X-ray or neutron atomic scattering lengths,  $b_h$ , and a volume,  $V_h$ . These groups make up a layer of a given thickness,  $d_h$ , which has some interfacial roughness,  $\sigma_h$ , within which some volume fraction of solvent may penetrate,  $\phi_h$ . Therefore, the SLD<sub>i</sub> in Equation 3.1 may be found and used in Equation 2.34, while the layer thickness may be used in Equation 2.36 to calculate the reflected intensity from a given layer. The tail layer is defined in the same way, however, the tail thickness,  $d_t$ , is constrained such that it can be no greater than the maximum extended length for the phospholipid tail,<sup>70</sup> which is given in Table 3.2, and that no solvent may penetrate into the layer.<sup>71</sup> Therefore, the SLD may be determined as discussed in Equation 3.1. Based on the work of Campbell *et al.*,<sup>72</sup> a single value for the interfacial roughness was fitted for all of the interfaces, including the subphase,<sup>73</sup> as there is only a single phospholipid molecule type present in each monolayer. Therefore, any capillary wave roughness at the air-DES interface is carried conformally through the layers. The interfacial roughness was constrained to be greater than 3.3 Å, in agreement with previous work.<sup>74</sup>

The constraints implemented in Code Block 3.2 involved two aspects. The first was to ensure that the number density of head groups and pairs of tail groups was kept the same. This was achieved with the following relation,<sup>75</sup>

$$\phi_h = 1 - \left( \frac{d_t V_h}{V_t d_h} \right). \quad (3.2)$$

**Code Block 3.1:** The chemically-consistent model class that was implemented in `refnx` A. R. J. Nelson *et al.* *J. Appl. Crystallogr.* 52.1 (2019), pp. 193–200; A. R. J. Nelson *et al.* *Refnx v0.1.2*. 2019. URL: <http://doi.org/10.5281/zenodo.2552023>. The input variables are `vol` which is an array of floats containing the initial values for the head and tail group volumes, `b` which is the calculated scattering length for the head and tail groups, `d_h` which is the initial value for the thickness of the head group region, `c_length` which is the number of carbon atoms in the phospholipid tail, `str` which is the name to be given to the object. The `slabs` function will return an array of floats representing the scattering length density profile.

<sup>70</sup> This is defined as the Tanford length,  $t_t$  from C. Tanford. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. 1980.

<sup>71</sup> Such at  $\phi_t = 0$ .

<sup>72</sup> Campbell *et al.*, see n. 26.

<sup>73</sup> i.e.  $\sigma_t = \sigma_h = \sigma_s$ .

<sup>74</sup> Sanchez-Fernandez *et al.*, see n. 13.

<sup>75</sup> L. Braun *et al.* *Adv. Colloid Interface Sci.* 247 (2017), pp. 130–148.

Code Block 3.2: The `set_constraints` that was used to impose chemical-consistency on the phospholipid monolayer structure.. The input variables are `lipids` and `structures` which are `refnx` objects that contain information about the phospholipids and monolayers, and `hold_tails`, `hold_rough`, and `hold_phih` are Boolean switches to constrain the tail layer thickness, the interfacial roughness, and the volume fraction of solvent across the different measurements, in this work these were all kept as `False`.

```
import numpy as np
import periodictable as pt

def set_constraints(
    lipids,
    structures,
    hold_tails=False,
    hold_rough=False,
    hold_phih=False,
):
    i = 0
    lipids[i].phi[0].constraint = 1 - (
        lipids[0].vol[0] / lipids[0].vol[1]
    ) * (lipids[i].d[1] / lipids[i].d[0])
    lipids[i].sigma.constraint = structures[i][-1].rough
    for i in range(1, len(lipids)):
        lipids[i].vol[0].constraint = lipids[0].vol[0]
        lipids[i].vol[1].constraint = lipids[0].vol[1]
        lipids[i].d[0].constraint = lipids[0].d[0]
        if hold_tails:
            lipids[i].d[1].constraint = lipids[0].d[1]
        lipids[i].phi[0].constraint = 1 - (
            lipids[i].vol[0] / lipids[i].vol[1]
        ) * (lipids[i].d[1] / lipids[i].d[0])
        lipids[i].phi[1].constraint = lipids[0].phi[1]
        if hold_rough:
            structures[i][-1].rough.constraint = structures[0][
                -1
            ].rough
        lipids[i].sigma.constraint = structures[i][-1].rough
        if hold_phih:
            lipids[i].phi[0].constraint = lipids[0].phi[0]
    return lipids
```

The second aspect was to enforce chemically-consistent constraints across the measurements that were conducted at different SPs. This was achieved by constraining the head and tail group volumes and the head layer thickness such that they do not vary between the different SP measurements.

The justification for constraining the tail volume is built on the assumption that the phospholipids remain in the same phase. On water, this may be demonstrated with a Langmuir isotherm. However, it was not possible to collect consistent Langmuir isotherm measurements.<sup>76</sup> Instead, grazing incidence X-ray diffraction was used to confirm the phases of DMPC and DPPC at 30 mN m<sup>-1</sup>. Figure 3.2 shows the grazing-incidence X-ray diffraction<sup>77</sup> data from different phospholipids at different temperatures. Unfortunately, all the patterns show a weak artefact due to scattering from the Teflon trough. However, there are clear (2, 0) diffraction peaks in the GIXD pattern for DPPC at 22 °C and DMPC at 7 °C indicating that both phospholipids are in the LC phase. This peak was also

<sup>76</sup> Due to the high viscosity of the DES.

<sup>77</sup> Abbreviated to GIXD.

Table 3.2: The invariant parameters within the chemically-consistent model. Values for  $t_t$  were taken from the Tanford formula (C. Tanford. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. 1980), and the SLD values for the DES from A. Sanchez-Fernandez *et al. Phys. Chem. Chem. Phys.* 18.48 (2016), pp. 33240–33249.

| Component            | $b_t/\text{fm}$ | $b_h/\text{fm}$ | $t_t/\text{\AA}$ | $\text{SLD}/10^{-6} \text{\AA}^{-2}$ |
|----------------------|-----------------|-----------------|------------------|--------------------------------------|
| X-ray                |                 |                 |                  |                                      |
| DPPC                 | 6827            | 4635            | 20.5             | –                                    |
| DMPC                 | 5924            | 4635            | 18.0             | –                                    |
| DLPC                 | 5021            | 4635            | 15.5             | –                                    |
| DMPG                 | 5924            | 4694            | 18.0             | –                                    |
| Air                  | –               | –               | –                | 0                                    |
| DES                  | –               | –               | –                | 10.8                                 |
| Neutron              |                 |                 |                  |                                      |
| $d_{54}\text{-DMPC}$ |                 |                 | 18.0             | –                                    |
| $d_{62}\text{-DPPC}$ |                 |                 | 20.5             | –                                    |
| h-DES                | –               | –               | –                | 0.43                                 |
| hd-DES               | –               | –               | –                | 3.15                                 |

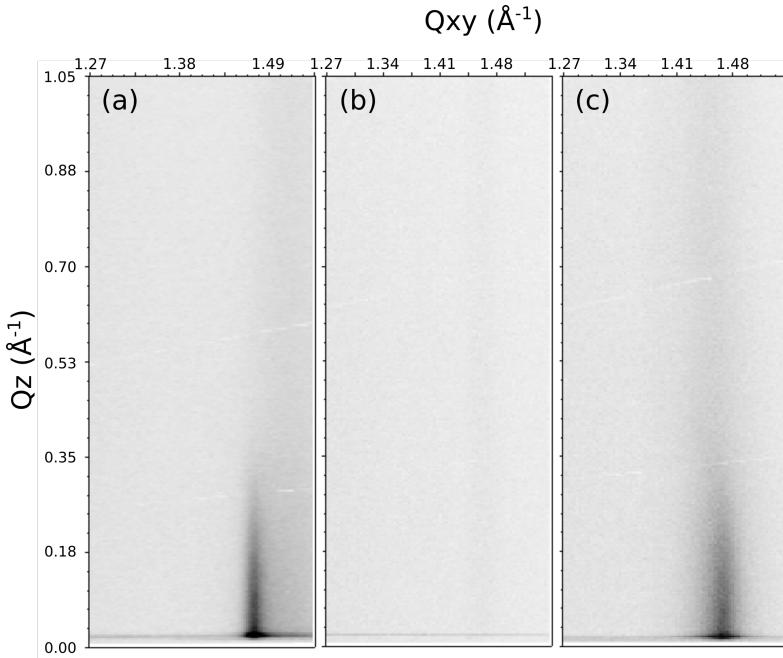


Figure 3.2: The GIXD patterns, where  $Q_z$  is the scattering vector normal to the interface and  $Q_{xy}$  is that in the plane of the interface; (a) DPPC at  $30 \text{ mN m}^{-1}$  and  $22^\circ\text{C}$ , (b) DMPC at  $30 \text{ mN m}^{-1}$  and  $22^\circ\text{C}$ , and (c) DMPC at  $30 \text{ mN m}^{-1}$  and  $7^\circ\text{C}$ . Note that  $Q$  is equivalent to  $q$ .

present at other SPs (data not shown). The peak position corresponded well with that found for DPPC in water.<sup>78</sup> DMPC at  $22^\circ\text{C}$  showed no evidence of a diffraction peak indicating the presence of the LE phase. It was assumed that DLPC and DMPG were also in the LE phase as there is no reason for the phase behaviour of these systems to differ significantly from that of DMPC at room temperature.

Initially, this chemically-consistent modelling approach was applied only to the XRR data. The tail layer thickness and interfacial roughness were allowed to vary independently across the SPs, while the other parameters were constrained as discussed above or held constant to the values given in Table 3.2. For each co-refinement of four XRR measurements, there were, in total, eleven degrees of freedom. Throughout all of the analyses, the reflected intensity scale factor was allowed to vary freely, while the background was constrained to the reflected intensity at the largest  $q$ -value.

Following this, the head and tail group volumes, and the head layer thickness that were found from the XRR analysis were used as fixed variables for the refinement of the NR measurements. This reduced the number of fitted parameters in the NR data to two, namely the thickness of the tail layer,  $d_t$ , and the interfacial roughness,  $\sigma_{t,h,s}$ , for the co-refinement of two datasets. Table 3.2 also presents details of the scattering lengths and SLDs used for the NR refinement. Again, the intensity scale factor was allowed to vary freely and the background constrained to the intensity at the largest  $q$ -value.

In both the XRR and the NR analysis, the refinement of the

<sup>78</sup> E. B. Watkins *et al.* *Phys. Rev. Lett.* 102.23 (2009), pp. 238101–238104.

<sup>79</sup> Gerelli, "Aurore: New Software for Neutron Reflectivity Data Analysis. Corrigendum", see n. ??.

<sup>80</sup> Jones *et al.*, see n. 35.

<sup>81</sup> J. Goodman *et al.* *Comm. App. Math. Comp. Sci.* 5 (2010), pp. 65–80.

<sup>82</sup> Foreman-Mackey *et al.*, see n. 36.

<sup>83</sup> S. S. Shapiro *et al.* *Biometrika* 52.3-4 (1965), pp. 591–611, this is a common test to assess the normality of a distribution.

<sup>84</sup> Therefore, they do not account for systematic uncertainty in the measurement technique.

<sup>85</sup> K. Pearson. *Proc. Royal Soc. Lond.* 58 (1895), pp. 240–242.

<sup>86</sup> An increase in one variable is associated with a decrease in the other.

<sup>87</sup> An increase in one variable is associated with a similar increase in the other.

chemically-consistent model to the experimental data involved the transformation of the reflectometry calculated from the model and the data into  $Rq^4$ -space, such that the contribution of the Fresnel decay was removed.<sup>79</sup> The model was then optimised using the DE method that is available within the `scipy` library.<sup>80</sup> This refined the parameters to give the best fit to the data. MCMC was then used to probe the search-space available to each parameter, given the experimental uncertainty of the data. The MCMC sampling method used was Goodman & Weare's Affine Invariant Ensemble<sup>81</sup> as implemented in the `emcee` package.<sup>82</sup> This enabled the determination of the probability distribution for each of the parameters, and therefore the quantification of their inverse uncertainty, given the uncertainty in the experimental data. A Shapiro-Wilk test<sup>83</sup> was used to determine if the PDF fitted to a normal distribution and therefore could be considered to have symmetric confidence intervals. If the PDF failed the test the value was quoted with asymmetric confidence intervals, compared with the symmetric confidence intervals given for those that passed the Shapiro-Wilk test. It is important to note that the PDFs and therefore the determined confidence intervals are not true confidence intervals, and account only for the uncertainty that is present in the data.<sup>84</sup> In addition to determining parameter confidence intervals, it was also possible to use these probability distributions to understand the correlations present between the parameters and the impact this has on the fitting process. The correlation was quantified using the Pearson correlation coefficient,<sup>85</sup> a common statistical definition for the level of correlation present between two variables. The Pearson correlation coefficient can have values that range from  $-1$  to  $1$ , with a value of  $-1$  corresponding to a complete negative correlation<sup>86</sup>, while a value of  $1$  corresponds to a complete positive correlation<sup>87</sup>, a value of  $0$  indicates no correlation between the two variables. The MCMC sampling involved 200 walkers that were used for 1000 iterations, following a burn-in of 200 iterations.

### 3.4 RESULTS & DISCUSSION

#### 3.4.1 X-ray reflectometry

The chemically-consistent model was co-refined across XRR measurements at all four SPs for each phospholipid. The resulting XRR profiles and associated SLD profiles are shown in Figure 3.3. Table 3.3 gives the parameters for each of the phospholipids at the second-highest SP measured, as well as the details of  $\phi_h$ , as determined from Equation 3.2.<sup>88</sup>

Following the structural determination of the monolayer from the XRR measurements. NR was used to confirm the values of the head and tail group volumes that had been determined. The resulting NR profiles and associated SLD profiles, at both SPs measured,

<sup>88</sup> The parameters at the other SPs may be found in Appendix B.

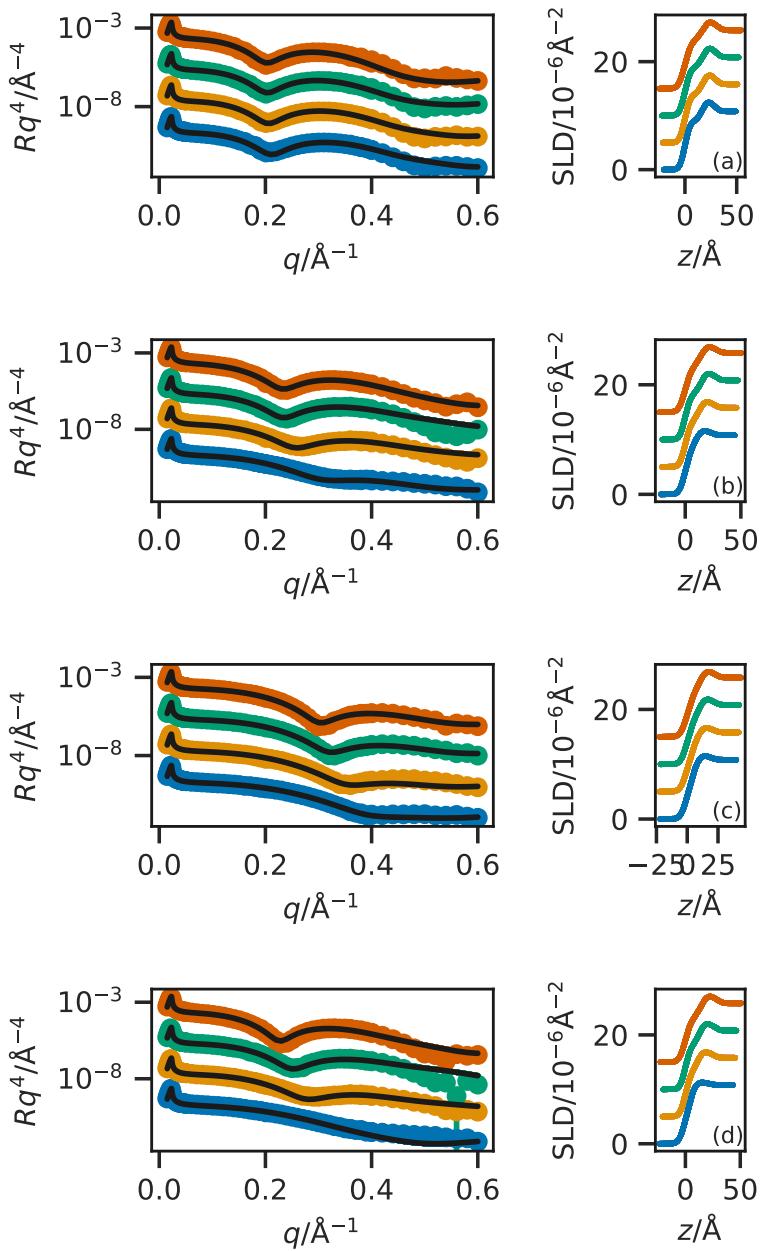


Figure 3.3: The XRR profiles (left) and SLD profiles (right) for each of the four phospholipids; (a) DPPC, (b) DMPC, (c) DLPC, and (d) DMPG, at the four measured SPs; increasing in SP from blue, orange, green to red, with the black lines representing the best fit. The different SP XRR profiles have been offset in the  $y$ -axis by two orders of magnitude and the SLD profiles offset in the  $y$ -axis by  $5 \times 10^{-6} \text{\AA}^{-2}$ , for clarity.

Table 3.3: The best-fit values, and associated 95 % confidence intervals for each of the varying parameters for each phospholipid at the second highest SP measured from XRR. The values of  $\phi_h$  were obtained from the appropriate use of Equation 3.2. SP has been abbreviated to SP.

| Phospholipid<br>SP/mN m <sup>-1</sup> | DPPC                    |                         |                   |                   |
|---------------------------------------|-------------------------|-------------------------|-------------------|-------------------|
|                                       | 25                      | 30                      | 30                | 25                |
| $V_t/\text{\AA}^3$                    | $765.31 \pm 0.38$       | $718.76 \pm 0.54$       | $625.50 \pm 3.70$ | $734.00 \pm 0.58$ |
| $V_h/\text{\AA}^3$                    | $322.00 \pm 0.24$       | $339.53 \pm 0.27$       | $331.38 \pm 0.81$ | $329.96 \pm 0.34$ |
| $d_h/\text{\AA}$                      | $12.70_{-0.03}^{+0.03}$ | $13.21_{-0.04}^{+0.04}$ | $11.00 \pm 0.13$  | $13.95 \pm 0.03$  |
| $d_t\text{\AA}$                       | $16.83 \pm 0.01$        | $13.72 \pm 0.01$        | $9.51 \pm 0.06$   | $12.24 \pm 0.01$  |
| $\sigma_{t,h,s}/\text{\AA}$           | $4.31 \pm 0.00$         | $3.86 \pm 0.00$         | $4.16 \pm 0.04$   | $3.81 \pm 0.00$   |
| $\phi_h/\times 10^{-2}$               | $44.23 \pm 0.22$        | $50.94 \pm 0.25$        | $54.20 \pm 1.10$  | $60.57 \pm 0.17$  |

can be found in Figure 3.4. Table 3.4 gives the parameters as determined from the NR measurements, along with  $\phi_h$  as determined from Equation 3.2.

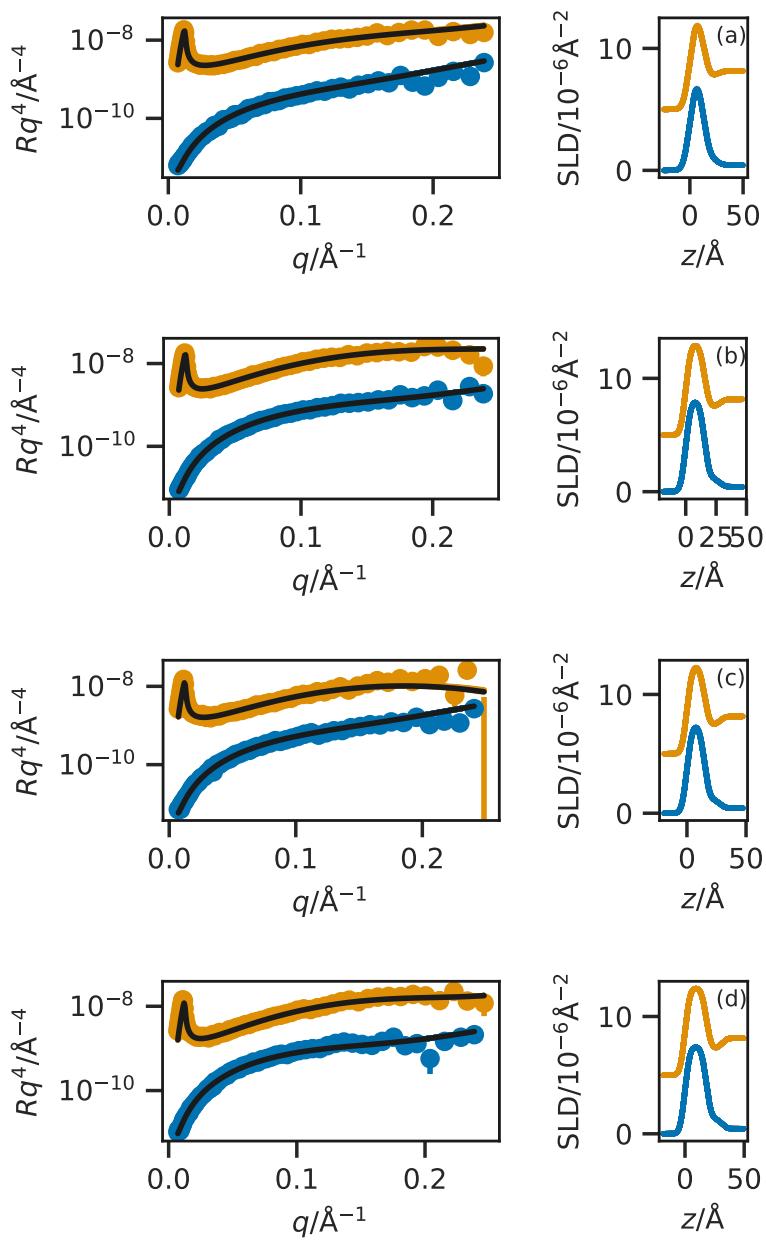


Figure 3.4: The NR profiles (left) and SLD profiles (right) for each of the four phospholipids; (a) DPPC at  $15 \text{ mN m}^{-1}$ , (b) DPPC at  $20 \text{ mN m}^{-1}$ , (c) DMPC at  $20 \text{ mN m}^{-1}$ , and (d) DMPC at  $25 \text{ mN m}^{-1}$ , where the blue data indicates the h-DES contrast, while the orange is the hd-DES, with the black lines representing the fit. The different SP NR profiles have been offset in the y-axis by an order of magnitude and the SLD profiles offset in the y-axis by  $5 \times 10^{-6} \text{ Å}^{-2}$ , for clarity.

### 3.4.2 Effect of compression on the monolayer thickness

From Tables B.1, B.2, 3.3, B.3, and 3.4, it can be seen that, as expected and shown in previous work at the air-water interface,<sup>89</sup> the thickness of the tail layer increases as the number of carbon atoms in the tail chain increases. Furthermore, the thickness of the tail layers determined here agrees well with values found for

<sup>89</sup> Mohwald, see n. ??; D. Vaknin *et al.* *Biophys. J.* 59.6 (1991), pp. 1325–1332.

Table 3.4: The best-fit values, and associated 95 % confidence intervals for each of the varying parameters for each phospholipid at the each SP from NR. The values of  $\phi_h$  were obtained from the appropriate use of Equation 3.2.

| Phospholipid<br>SP/mN m <sup>-1</sup> | DPPC<br>15       | DPPC<br>20             | DMPC<br>20       | DMPC<br>25              |
|---------------------------------------|------------------|------------------------|------------------|-------------------------|
| $d_t \text{Å}$                        | $12.67 \pm 0.13$ | $15.43 \pm 0.08$       | $14.81 \pm 0.13$ | $17.98^{+0.01}_{-0.03}$ |
| $\sigma_{t,h,s} / \text{Å}$           | $4.77 \pm 0.16$  | $3.31^{+0.04}_{-0.01}$ | $3.47 \pm 0.15$  | $3.30^{+0.02}_{-0.00}$  |
| $\phi_h / \times 10^{-2}$             | $58.04 \pm 0.44$ | $48.86 \pm 0.26$       | $47.06 \pm 0.48$ | $35.71^{+0.11}_{-0.03}$ |

water-analogues;  $13.72 \pm 0.01 \text{ Å}$  at  $30 \text{ mN m}^{-1}$  in DES compared with  $15.8 \text{ Å}$  at  $30 \text{ mN m}^{-1}$  in water for DMPC, and  $16.91 \pm 0.01 \text{ Å}$  at  $30 \text{ mN m}^{-1}$  in DES compared with  $16.7 \text{ Å}$  at  $40 \text{ mN m}^{-1}$  in water for DPPC.

The variation of the tail layer thickness in the models with SP is given for each phospholipid in Figure 3.5. For all of the phospholipids, as the SP increases, the thickness of the tail layer also increases to a point before plateauing; for DPPC this occurs at  $20 \text{ mN m}^{-1}$ , DMPC at  $30 \text{ mN m}^{-1}$ , and for DMPG and DLPC can be assumed to be at higher pressures than those studied. This relationship of increasing tail layer thickness with increasing SP has been noted previously for DMPC<sup>90</sup> and DPPC<sup>91</sup> at the air-water interface. This can be easily understood as the angle of the tail group with respect to the surface normal decreasing as the SP increases.

### 3.4.3 Effect of compression on solvent fraction

In Figure 3.5, it is clear that for all four phospholipids, as the SP is increased there is a corresponding reduction in the volume fraction of solvent in the phospholipid head layer. This can be rationalised by considering that when the SP is increased, there is a corresponding increase in surface concentration, hence the free volume available to the solvent is less. A similar effect has been observed when increasing the SP from  $11 \text{ mN m}^{-1}$  to  $31 \text{ mN m}^{-1}$  for a mixed DMPC/DMPG monolayer at the air-water interface.<sup>92</sup>

### 3.4.4 Effect of compression on the phospholipid tail component volumes

It can be seen from comparing Table 3.1 with Tables B.1, B.2, 3.3, and B.3 that the volumes of the phospholipid tails are substantially lower in the current measurements than found previously, by other techniques. It is unlikely that this is a result of the DES subphase, due to the solvophobic nature of these tail groups. However, a similar reduction has been shown previously,<sup>93</sup> where it was rationalised by the compaction of the monolayer at elevated SP. In that work, the optimal value for the tail group volume of DPPC was found to be  $772 \text{ Å}^3$  at a SP of  $35 \text{ mN m}^{-1}$ , which agrees well with the value of  $765.31 \pm 0.38 \text{ Å}^3$  found in this work at SPs of 15, 20, 25 and  $30 \text{ mN m}^{-1}$ . The reduction in tail volume was found to be between 8-12 % for DPPC, DMPC, DLPC when compared with the literature sources at 24 and  $30^\circ\text{C}$ . This is close to the maximum

<sup>90</sup> Bayerl *et al.*, see n. 56.

<sup>91</sup> Campbell *et al.*, see n. 26.

<sup>92</sup> Bayerl *et al.*, see n. 56.

<sup>93</sup> Campbell *et al.*, see n. 26.

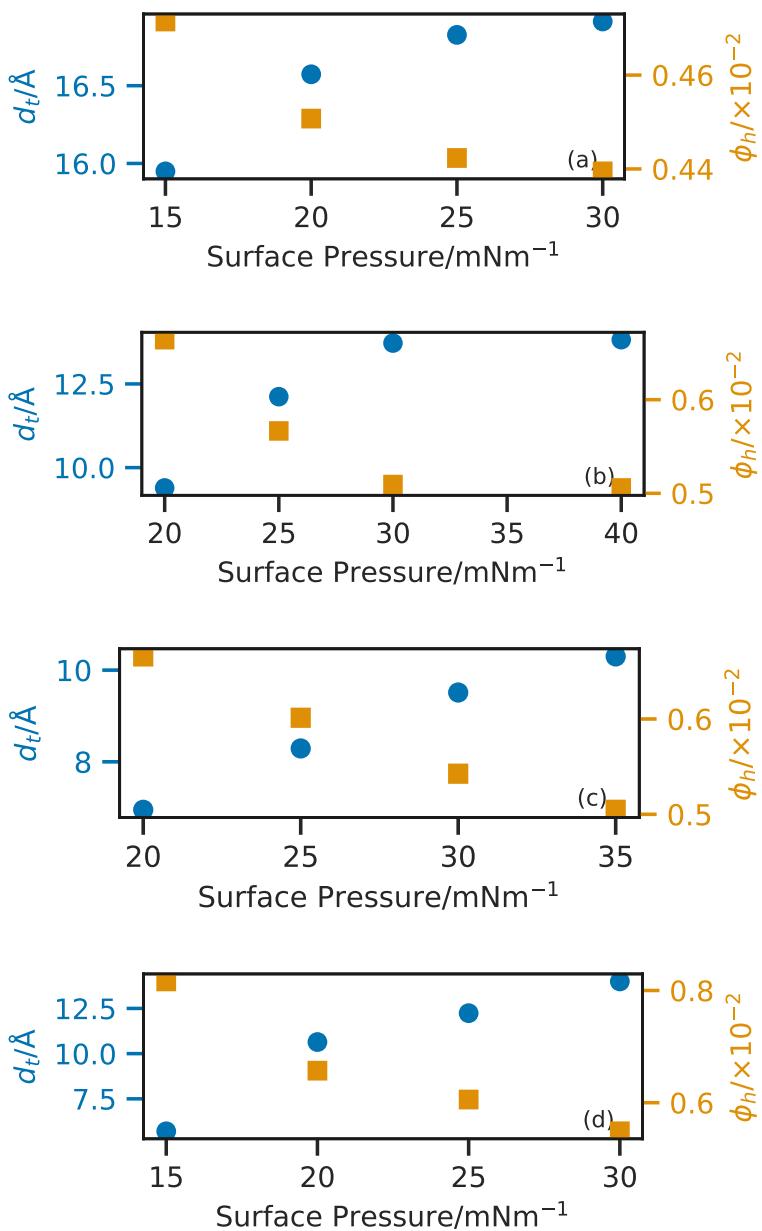


Figure 3.5: The variation of tail layer thickness,  $d_t$ , (blue circles) and head group solvation,  $\phi_h$ , (orange squares) with SPs for each of the four phospholipids; (a) DPPC, (b) DMPC, (c) DLPC, and (d) DMPG.

<sup>94</sup> Small, see n. 41.

compression of 15 % noted by Small.<sup>94</sup> DMPG shows a small increase in the tail volume when compared with the literature value, albeit at a higher temperature. However, this value agrees well with that found for DMPC which shares the same tail structure.

### 3.4.5 Solvent effect on the phospholipid head group volume

Tables B.1, B.2, 3.3, and B.3 give the best-fit values for the head group volumes for each of phospholipids investigated. The three phospholipids with the PC head group are consistent, giving values of  $\sim 330 \text{ \AA}^3$ , regardless of the tail group. This agrees well with the values found for the same head component in water, shown in Table 3.1. Interestingly, the head group volume determined for the PG containing phospholipid is similar to that for the PC head group, with a value of  $329.96 \pm 0.34 \text{ \AA}^3$ . The PG head group volume in water from either DMPG using DVT<sup>95</sup> or POPG using MD simulations,<sup>96</sup> is noticeably smaller. This indicates that there may be some effect arising from the solvation of the PG component in the choline chloride:glycerol DES. However, this has only been shown for a single PG-lipid at the air-DES interface.

The major difference between the two head groups of the phospholipid is that the PG is present as a sodium salt, whereas the PC is zwitterionic. When in solution the anionic PG head is expected to associate with cations in solution, as it does in water<sup>97</sup> where such interactions depend on a variety of factors including the ionic strength. In the case of a DES, the environment is inherently ionic and therefore the interaction of an anionic phospholipid head may be more complex. As well as interacting with this sodium, the head is likely to interact with the choline cations, similar to behaviour reported previously for surfactant micelles.<sup>98</sup> The extent of interaction with each of the cations is unclear, but regardless it seems likely that the solvation of the PG head is improved in the DES relative to water. This better solvation would explain the apparent increase in the volume of the PG head since it would result in swelling of this group through its strong interactions with the solvent. In the case of PC, the proximity of a local cation within the molecule results in the same folding of the head group seen in water because this interaction is less transient than the equivalent interactions with the solvent.

### 3.4.6 Analysis of neutron reflectometry

The ability to fit NR data in Figure 3.4 indicated that the values found for the head and tail groups are consistent between the pair of measurements for the same systems. It is clear that again stable monolayers of the phospholipids are forming at the air-DES interface and that the volumes determined by XRR measurements are robust enough to be used in the modelling of NR data. Furthermore, as shown in Table 3.4, the trends observed with increasing

<sup>95</sup> Pan *et al.*, see n. 58.<sup>96</sup> Kučerka *et al.*, see n. 59.<sup>97</sup> D. Grigoriev *et al.* *J. Phys. Chem. B* 103.6 (1999), pp. 1013–1018.<sup>98</sup> A. Sanchez-Fernandez *et al.* *Phys. Chem. Chem. Phys.* 20.20 (2018), pp. 13952–13961.

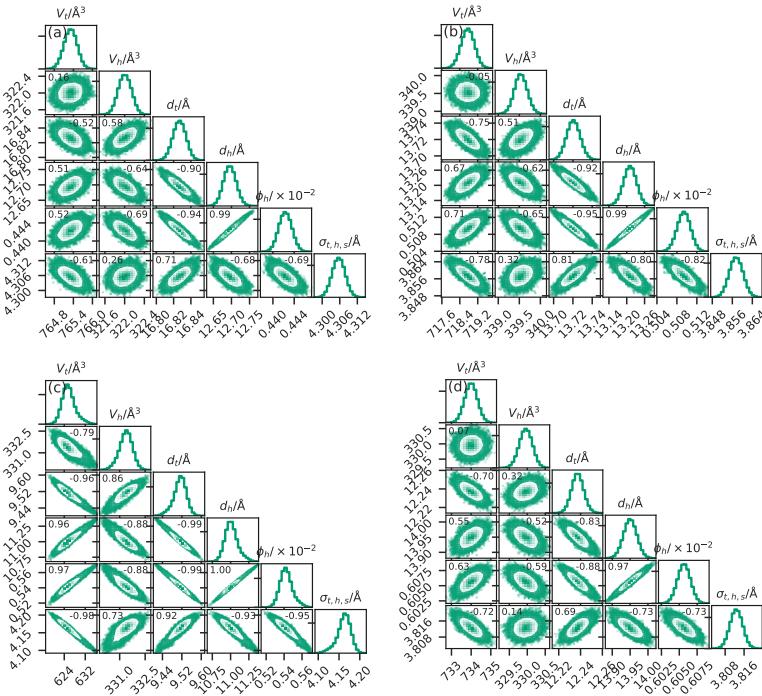


Figure 3.6: The PDFs from the chemically-consistent modelling of (a) DPPC at 25 mN m<sup>-1</sup>, (b) DMPC at 30 mN m<sup>-1</sup>, (c) DLPC 30 mN m<sup>-1</sup>, and (d) DMPG at 25 mN m<sup>-1</sup>. The Pearson correlation coefficient for each pair of parameters is given in the top corner of each two-dimensional PDF.

SP in the XRR models, pertaining to a responsive increase in tail thickness and a decrease in solvent concentration in the head layer are consistent with that found with the NR analysis.

### 3.4.7 Utility of MCMC sampling

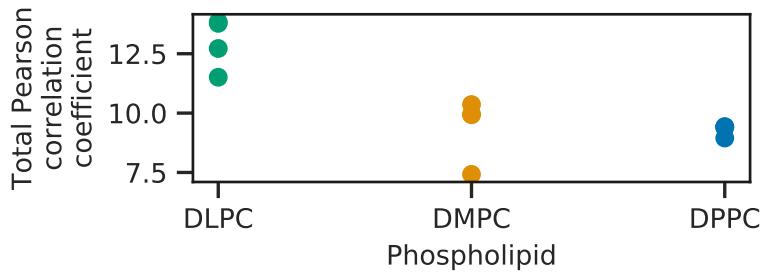
The use of MCMC sampling enabled the inverse uncertainties for each of the fitted parameters to be determined as a confidence interval from the PDFs. The PDFs for each lipid at the second-highest SPs are shown in Figure 3.6.<sup>99</sup> These confidence intervals are useful for understanding the probability of the given value, however as discussed in Section 3.3 these intervals only represent the uncertainty in the experimental data and do not account for systematic uncertainty present in the measurement method.

The sums of the magnitudes of the Pearson correlation coefficients for each PC-containing phospholipids at each SPs are given in Figure 3.7. From this figure, it is clear that there is a relationship between the phospholipid and the correlation present in the reflectometry model due to the skewed two-dimensional probability distribution. This is the effect of the reduction in the phospholipid length from DPPC to DLPC, and that a corresponding decrease is not observed for the interfacial roughness. Therefore, the boundary between phospholipid head and tail layers is less well defined.<sup>100</sup> Furthermore, the magnitude of the Pearson correlation between the head and tail thicknesses increases with increasing tail length; from -0.89 for DPPC, to -0.92 for DMPC, to -0.99 for DLPC, each at 30 mN m<sup>-1</sup>. Indicating that as the phospholipid tail length de-

<sup>99</sup> Those for the other phospholipid-XRR and NR measurements can be found in Appendix B.

<sup>100</sup> As can be observed by investigating the SLD profiles in Figure 3.3.

Figure 3.7: The sum of the magnitudes for each of the Pearson correlation coefficients for each of the PC-containing lipids.



creases, the negative correlation between the head and tail layers increases to the point for DLPC where the two variables are almost completely correlated and the boundary between the head and the tail is nearly nonexistent due to the magnitude of the interfacial roughness.

There is also a substantial positive correlation present for all of the datasets between the phospholipid head layer thickness and the volume fraction of solvent in the head layer. This correlation can be rationalised as a result of the SLD of the solvent and the head layer<sup>101</sup> being similar, and therefore the boundary between the head layer and the solvent is also poorly defined. A significant correlation such as this is unavoidable, without considering the use of many neutron contrasts for both the phospholipid and the solvent, due to the highly solvophilic nature of the head groups.

### 3.5 CONCLUSIONS

Stable PC and PG lipid monolayers were observed and characterised on an ionic solvent surface. Until the emergence of ionic liquids and DES, only a limited number of molecular solvents exhibited the ability to promote self-assembly and only water and formamide among those had previously demonstrated the formation of phospholipid monolayers at the air-liquid interface.<sup>102</sup>

<sup>101</sup> Mohwald, see n. ??; Graner *et al.*, see n. 19.

For the first time, a physically and chemically-consistent reflectometry modelling approach was used to co-refine XRR measurements at different SPs. This enabled modelling without the need to constrain the head and tail group volumes, enabling these parameters to vary freely to account for any variation occurring due to the elevated SPs used or the presence of a non-aqueous solvent, compared to the commonly applied literature values. This allows a significant difference in the PG head group volume to be observed; having a larger volume than observed for the same system in water. This suggests that the transfer of phospholipids to a DES is not just a simple substitution of the subphase. In this specific case, an explanation based on the dissociation of the PG head group salt and the subsequent interaction with the DES has been proposed.

Finally, MCMC sampling was used to understand the inverse

uncertainties present in the modelling parameters, enabling a confidence interval to be quoted alongside the most probable value. The use of MCMC sampling also allowed the quantification of the correlations between the parameters of the chemically-consistent model. This shows the significant correlations between the head layer thickness and the volume fraction of solvent, and the head layer thickness and the tail layer thickness that becomes more prominent for short-tailed phospholipids. The quantification of these correlations gives us a better understanding of the uncertainties on the parameters and can be rationalised based on the chemistry of the monolayers.



# 4

## APPLYING ATOMISTIC AND COARSE-GRAINED SIMULATION TO REFLECTOMETRY ANAL- YSIS

### ABSTRACT

The use of molecular simulation to aid in the analysis of neutron reflectometry<sup>1</sup> measurements has become commonplace. However, reflectometry is a tool to probe large-scale structures, and therefore the use of all-atom simulation may be irrelevant. This work presents the first direct comparison between the reflectometry profiles obtained from different all-atom and coarse-grained molecular dynamics<sup>2</sup> simulations and the reflectometry profiles from a chemically-consistent layer-based modelling method. We find that systematic limitations reduce the efficacy of the MARTINI potential model, while the Berger united-atom and Slipids all-atom potential models agree similarly well with the experimental data. The chemically-consistent layer model gives the best agreement, however, the higher resolution simulation-dependent methods produce an agreement that is comparable.

<sup>1</sup> NR.

<sup>2</sup> MD.

### CONTEXT

This chapter builds on the previous chapter, by using the traditional, highly coarse-grained, chemically-consistent layer-based model as a point of comparison with classical molecular simulations that use a variety of simulation grain-sizes, from the coarse-grained MARTINI potential model to the all-atom Slipids. Therefore, the analysis in this chapter applies significantly more constraints on the system, however, the constraints are built on substantial underlying chemistry, given their grounding in the potential models. It is hoped that this work will provide an advisory document to those interested in applying classical simulation to the analysis of their NR experiments. Furthermore, the simulations are used to advise on ways that the chemically-consistent models may be improved in the future. It is noted again that the focus of

this chapter is the methodological developments, rather than the particular chemical system to which they are applied.

#### 4.1 INTRODUCTION

The use of a traditional, layer-based model approach<sup>3</sup> is a powerful tool to understand the structure of complex systems such as biomimetic bacterial membranes<sup>4</sup> and polymeric energy materials.<sup>5</sup> These layers structures are typically defined by the underlying chemistry of the system. However, there has been growing interest in the use of MD simulations to inform the development of these layer structures. This is due to the fact that the equilibrium structures for soft matter interfaces, that are often of interest in reflectometry studies, are accessible on all-atom simulation timescales.<sup>6</sup> However, there has been no work that directly compares different levels of simulation coarse-graining in order to assess the required resolution for the accurate reproduction of a given NR profile.

Simulation-driven multi-modal analysis has been applied previously, either by the calculation of the SLD profile from the simulation by the full determination of the reflectometry profile. In the former case, the calculated SLD profile may be compared with the SLD profile determined from the use of a traditional analysis method. Bobone *et al.* used such a method to study the antimicrobial peptide trichogin GA-IV within a supported phospholipid bilayer.<sup>7</sup> A four-layer model consisting of the hydrated SiO<sub>2</sub> layer, an inner phospholipid head-region, a phospholipid tail-region, and an outer phospholipid head region was used in the Abelès matrix formalism. The SLD profile from the MD simulations agreed well with that fitted to the reflectometry data from the layer model.

The reflectometry profile was determined explicitly from the classical simulation in the works of Miller *et al.* and Anderson and Wilson.<sup>8</sup> In these studies, an amphiphilic polymer at the oil-water interface was simulated by Monte Carlo and MD respectively, and the NR profile was found by splitting the simulation cell into a series of small layers and treating these layers with the Abelès formalism. There was good agreement between the experimental and calculated reflectometry, for low interfacial coverages of the polymer. Another study that has made a direct comparison between the atomistic simulation-derived reflectometry data and those measured experimentally includes that of Darré *et al.*<sup>9</sup> In this work, NeutronRefTools was developed to produce the NR profile from an MD simulation. The particular system studied was a supported DMPC phospholipid bilayer, with good agreement found between the simulation-derived profile and the associated experimental measurements. However, the nature of the support required that a correction for the head group hydration be imposed to achieve this agreement.

Koutsioubas used the MARTINI coarse-grained representation of a DPPC phospholipid bilayer to compare with experimental reflectometry.<sup>10</sup> This work shows that the parameterisation of the MARTINI water bead was extremely important in the reproduction

<sup>3</sup> As is outlined in Chapter 3.

<sup>4</sup> R. D. Barker *et al.* "Neutron Reflectivity as a Tool for Physics-Based Studies of Model Bacterial Membranes". In: *Biophysics of Infection*. 2016, pp. 261–282.

<sup>5</sup> S. Khodakarimi *et al.* *J. Mater. Sci. Mater. Electron.* 27.1 (2016), pp. 182–190.

<sup>6</sup> Scoppola *et al.*, see n. 31.

<sup>7</sup> S. Bobone *et al.* (*BBA*) - *Biomembranes* 1828.3 (2013), pp. 1013–1024.

<sup>8</sup> Miller *et al.*, see n. 37; Anderson *et al.*, see n. 37.

<sup>9</sup> L. Darré *et al.* *J. Chem. Theory Comput.* 11.10 (2015), pp. 4875–4884.

<sup>10</sup> Koutsioubas, see n. 41.

<sup>11</sup> A. V. Hughes *et al.* *Acta Crystallogr. D* 72.12 (2016), pp. 1227–1240.

of the reflectometry data, as the non-polarisable water bead would freeze into crystalline sheets resulting in artefacts in the reflectometry profiles calculated. The work of Hughes *et al.* studied, again, a DPPC phospholipid bilayer system,<sup>11</sup> albeit an all-atom representation, that was compared with a supported DPPC phospholipid bilayer system measured with polarised NR. The SLD profile found from MD simulation was varied to better fit the experimental measurement, resulting in good agreement. Additionally, the ability to vary the SLD profile was used to remove an anomalous difference present in the SLD, that arose when the MD simulations were merged with an Abelès layer model. This was done to account for regions present in the experiment that were not modelled explicitly.

In all the examples discussed so far, there is no direct comparison between the reflectometry profile determined from simulation and that from the application of a traditional modelling approach. Indeed, the only example,<sup>12</sup> where a direct comparison was drawn is the work of Dabkowska *et al.*<sup>13</sup> This work compares the reflectometry profile from a DPPC monolayer at the air-water interface containing dimethyl sulfoxide molecules with a similar MD simulation using the CHARMM potential model. The use of multimodal analysis allowed the determination of the position and orientation of dimethyl sulfoxide molecules at a particular region within the monolayer.

The previously mentioned work of Koutsoubas involved the use of the MARTINI coarse-grained potential model to simulate the DPPC bilayer system.<sup>14</sup> The use of atomistic simulation for soft matter systems, such as a phospholipid bilayer, is undesirable as this requires a huge number of atoms to be simulated, due to the large lengths scales involved. The purpose of simulation coarse-graining is to reduce the number of particles over which the forces must be integrated, additionally, by removing the higher frequency bond vibrations, the simulation timestep can also be increased.<sup>15</sup> Together, these two factors enable an increase in both simulation size and length. The use of the MARTINI 4-to-1 coarse-grained and the Berger united atom<sup>16</sup> potential models are particularly pertinent for the application to phospholipid simulations as both were developed with this specific application in mind.<sup>17</sup>

The MARTINI potential model involves integrating the interactions of every four heavy atoms<sup>18</sup> into beads of different chemical nature. This potential model attempts to simplify the interactions of phospholipid and protein molecules significantly by allowing for only eighteen particle types, defined by their polarity, charge, and hydrogen-bond acceptor/donor character, which are discussed in detail in the work of Marrink *et al.*<sup>19</sup> This coarse-grained potential model was initially developed for the simulation of a phospholipid bilayer, and proteins held within and therefore is parameterised well under these conditions. It has successfully been used to sim-

<sup>12</sup> To the author's knowledge.

<sup>13</sup> Dabkowska *et al.*, see n. 40.

<sup>14</sup> Koutsoubas, see n. 41.

<sup>15</sup> Pluhackova *et al.*, see n. 89.

<sup>16</sup> Where hydrogen atoms are integrated into the heavier atoms to which they are bound.

<sup>17</sup> Marrink *et al.*, see n. 43; O. Berger *et al.* *Biophys. J.* 72.5 (1997), pp. 2002–2013.

<sup>18</sup> Larger than hydrogen.

<sup>19</sup> Marrink *et al.*, see n. 43.

ulate a wide range of systems, such as DNA nucleotides,<sup>20</sup> the micellisation of zwitterionic and nonionic surfactants,<sup>21</sup> and the self-assembly of ionic surfactants.<sup>22</sup>

Increasing the simulation resolution gives an united-atom potential model, where all of the hydrogen atoms are integrated into the heavier atoms to which they are bound. One of the most popular united-atom potential models for phospholipid simulations is that developed by Berger *et al.*<sup>23</sup> The Berger parameters were optimised to reproduce phospholipid density and area per phospholipid, the latter of which is often an important parameter for the understanding of reflectometry profiles. Since its inception, this potential model has proven one of the most commonly used and resilient sets of phospholipid parameters, with the original paper being cited 1500 times at the time of writing. Applications of this potential model have mostly been focussed on the simulation of membrane-bound proteins in a phospholipid bilayer.<sup>24</sup>

The Slipid<sup>25</sup> potential model was developed in 2012 by Jämbeck and Lyubartsev,<sup>26</sup> where the potential model was again designed to reproduce the structure of a phospholipid bilayer. The authors optimised the average area per phospholipid, the thermal expansivity, and contractivity, among other structural and thermodynamic parameters. This included comparing the X-ray reflectometry profiles of the phospholipid bilayers with those measured experimentally. In later work, additional parameters were optimised to agree well with experimental values.<sup>27</sup> Similar to the application of the Berger potential model, the Slipid potential model has been applied to the study of membrane-protein bound systems, such as the modulation of ion transfer.<sup>28</sup> However, it has also been used for the study of water diffusion within phospholipid membranes.<sup>29</sup>

It is clear that there is substantial interest in the use of classical simulation, and coarse-graining, for the analysis of NR data. However, there has been no work to investigate whether the use of atomistic simulations gives more detail than is required to reproduce the reflectometry profile accurately or to assess whether the application of a coarse-grained representation is suitable to aid in analysis. This chapter presents the comparison of three MD simulations of different potential models, with different degree of coarse-graining; namely the Slipid all-atom,<sup>30</sup> Berger united-atom,<sup>31</sup> and MARTINI coarse-grained potential models.<sup>32</sup> This comparison offers a fundamental insight into the simulation resolution that is necessary to reproduce experimental NR measurements. Furthermore, the highest resolution simulations are used to suggest possible adjustments that may be made to the traditional, layer models that are commonly used to analyse these measurements.

<sup>20</sup> J. J. Uusitalo *et al.* *J. Chem. Theory Comput.* 11.8 (2015), pp. 3932–3945.

<sup>21</sup> S. A. Sanders *et al.* *J. Chem. Phys.* 132.11 (2010), p. 114902.

<sup>22</sup> S. Wang *et al.* *Langmuir* 31.4 (2015), pp. 1262–1271.

<sup>23</sup> Berger *et al.*, see n. 17.

<sup>24</sup> D. P. Tieleman *et al.* *J. Phys. Condens. Matter* 18.28 (2006), S1221–S1234; A. Cordoní *et al.* *J. Chem. Theory Comput.* 8.3 (2012), pp. 948–958.

<sup>25</sup> A shortening of Stockholm Lipids, after the University at which the potential model was developed.

<sup>26</sup> J. P. M. Jämbeck *et al.* *J. Phys. Chem. B* 116.10 (2012), pp. 3164–3179.

<sup>27</sup> J. P. M. Jämbeck *et al.* *J. Chem. Theory Comput.* 8.8 (2012), pp. 2938–2948; J. P. M. Jämbeck *et al.* *J. Chem. Theory Comput.* 9.1 (2013), pp. 774–784.

<sup>28</sup> E. Segala *et al.* *J. Med. Chem.* 59.13 (2016), pp. 6470–6479.

<sup>29</sup> Y. von Hansen *et al.* *Phys. Rev. Lett.* 111.11 (2013), pp. 118103–118105.

<sup>30</sup> Jämbeck *et al.*, “Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids”, see n. 26.

<sup>31</sup> Berger *et al.*, see n. 17.

<sup>32</sup> Marrink *et al.*, see n. 43.

Table 4.1: The different contrasts of phospholipid monolayer and water investigated.

| Shorthand                         | Phospholipid contrast | Water contrast   |
|-----------------------------------|-----------------------|------------------|
| h-D <sub>2</sub> O                | h-DSPC                | D <sub>2</sub> O |
| d <sub>13</sub> -D <sub>2</sub> O | d <sub>13</sub> -DSPC | D <sub>2</sub> O |
| d <sub>13</sub> -ACMW             | d <sub>13</sub> -DSPC | ACMW             |
| d <sub>70</sub> -D <sub>2</sub> O | d <sub>70</sub> -DSPC | D <sub>2</sub> O |
| d <sub>70</sub> -ACMW             | d <sub>70</sub> -DSPC | ACMW             |
| d <sub>83</sub> -D <sub>2</sub> O | d <sub>83</sub> -DSPC | D <sub>2</sub> O |
| d <sub>83</sub> -ACMW             | d <sub>83</sub> -DSPC | ACMW             |

## 4.2 METHODS

### 4.2.1 Neutron reflectometry measurements

<sup>33</sup> Hollinshead *et al.*, see n. 43, full details of the experimental methods used can be found in that publication.

<sup>34</sup> Abbreviated to DSPC.

<sup>35</sup> The different lipid constants are abbreviated to h-DSPC, d<sub>13</sub>-DSPC, d<sub>70</sub>-DSPC, and d<sub>83</sub>-DSPC respectively.

<sup>36</sup> Abbreviated to D<sub>2</sub>O and ACMW respectively, where ACMW is a mixture of D<sub>2</sub>O and H<sub>2</sub>O such that the SLD is zero.

<sup>37</sup> SPs.

The NR measurements analysed in this chapter were published previously by Hollinshead *et al.*.<sup>33</sup> These measurements concern the study of a monolayer of 1,2-distearoyl-sn-phosphatidylcholine<sup>34</sup> at the air-water interface. The NR measurements were conducted on seven isotopic contrasts of the phospholipid and water. These contrasts were made up from four phospholipid types; fully-hydrogenated phospholipid, head-deuterated phospholipid, tail-deuterated phospholipid, and fully-deuterated phospholipid,<sup>35</sup> were paired with two water contrasts; fully-deuterated water and air-contrast matched water.<sup>36</sup> The pairing of the fully-hydrogenated phospholipid with ACMW was not performed, due to the lack of scattering available from such a system. Measurements were conducted at four different surface pressures,<sup>37</sup> 20, 30, 40 and 50 mN m<sup>-1</sup>. Table 4.1 outlines the shorthands used to refer to the different contrast pairings in this work.

### 4.2.2 Molecular dynamics simulations

<sup>38</sup> Jämbeck *et al.*, “Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids”, see n. 26.

<sup>39</sup> H. J. C. Berendsen *et al.* *J. Phys. Chem.* 91.24 (1987), pp. 6269–6271, abbreviated to SPC.

<sup>40</sup> S. Miyamoto *et al.* *J. Comput. Chem.* 13.8 (1992), pp. 952–962; B. Hess. *J. Chem. Theory Comput.* 4.1 (2008), pp. 116–122.

<sup>41</sup> Berger *et al.*, see n. 17.

<sup>42</sup> It is noted that these timesteps are shorter than those typically used for both potential models and that timesteps of up to 2 fs have been applied previously.

<sup>43</sup> Marrink *et al.*, see n. 43.

<sup>44</sup> S. O. Yesylevskyy *et al.* *PLoS Comput. Biol.* 6.6 (2010). Ed. by M. Levitt, e1000810.

<sup>45</sup> Koutsoubas, see n. 41.

The DSPC monolayer simulations were made up of phospholipid molecules modelled with three potential models, each with a different degree of coarse-graining. The Slipids potential model is an all-atom representation of the phospholipid molecules,<sup>38</sup> which was used alongside the single point charge water model,<sup>39</sup> with a timestep of 0.5 fs, the SHAKE, RATTLE, and PLINCS methods were used to constrain the C–H bond.<sup>40</sup> The Berger potential model is obtained by the integration of the hydrogen atoms into the heavy atoms to which they are bound, producing a united-atom potential model;<sup>41</sup> again the SPC water molecules were used. This potential model was simulated with an increased timestep of 1 fs.<sup>42</sup> Finally, the lowest resolution potential model used was the MARTINI<sup>43</sup> alongside the polarisable MARTINI water model,<sup>44</sup> to avoid the freezing issues observed previously.<sup>45</sup> The MARTINI 4-to-1 heavy atom beading allows for the use of a 20 fs timestep. For the Slipids and Berger potential model simulations a short-range cut-off of 10 Å was used, while for the MARTINI potential model simulations the cut-off was extended to 15 Å. All simulations were conducted

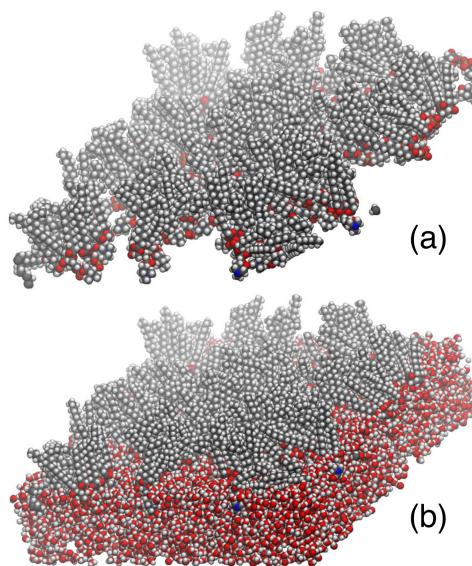


Figure 4.1: The DSPC monolayer; (a) without water layer and (b) with water layer, visualised using VMD (W. Humphrey *et al.* *J. Mol. Graph.* 14.1 (1996), pp. 33–38).

with temperature coupling to a heat bath at 300 K and a leapfrog integrator, and run using GROMACS 5.0.5<sup>46</sup> on 32 cores of the STFC Scientific Computing resource SCARF. The simulations were of monolayers, therefore the Ewald 3DC correction was applied to allow for the use of  $x/y$ -only periodic boundary condition.<sup>47</sup> A close-packed “wall” of non-interacting dummy atoms was placed at each side of the simulation cell in the  $z$ -direction, to ensure that the atoms could not leave the simulation cell.

The starting simulation structure was generated using the molecular packing software Packmol.<sup>48</sup> This was used to produce a monolayer of 100 DSPC molecules, with the head group oriented to the bottom of the simulation cell. A 6 Å layer of water was then added such that it overlapped the head groups, this was achieved with the solvate functionality in GROMACS 5.0.5. Examples, of the dry and wet monolayer for the Berger potential model, can be seen in Figure 4.1.

A general protocol was then used to relax the system at the desired surface coverage, reproducing the effects of a Langmuir trough *in-silico*. This involved subjecting the system to a semi-isotropic barostat, with compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  of the Slurids and Berger simulations and  $3.0 \times 10^{-4} \text{ bar}^{-1}$  for the MARTINI simulations. The pressure in the  $z$ -dimension was kept constant at 1 bar, while it was increased in the  $x$ - and  $y$ -dimensions isotropically. This allowed the surface area of the interface to reduce, as the phospholipid molecules have a preference to stay at the interface, while the total volume of the system stayed relatively constant, as the water molecules move down to relax the pressure in the  $z$ -dimension. When the  $xy$ -area is reached that is associated with the area per molecule<sup>49</sup> for each SP, described by the experimental SP-isootherm shown in Figure 4.2 and given in Table 4.2, the coordinates were saved and used as the starting structure for the

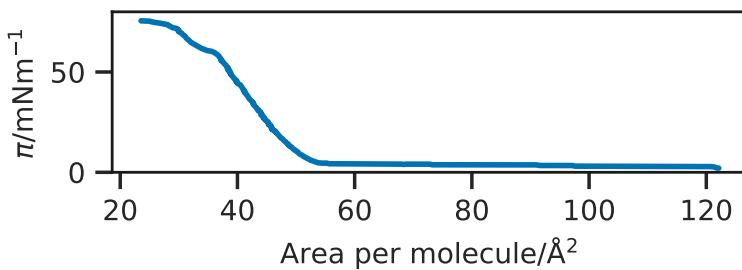
<sup>46</sup> H. J. C. Berendsen *et al.* *Comput. Phys. Commun.* 91.1-3 (1995), pp. 43–56; E. Lindahl *et al.* *J. Mol. Model.* 7.8 (2001), pp. 306–317; D. van der Spoel *et al.* *J. Comput. Chem.* 26.16 (2005), pp. 1701–1718; B. Hess *et al.* *J. Chem. Theory Comput.* 4.3 (2008), pp. 435–447.

<sup>47</sup> I.-C. Yeh *et al.* *J. Chem. Phys.* 111.7 (1999), pp. 3155–3162.

<sup>48</sup> Martínez *et al.*, see n. 124.

<sup>49</sup> Abbreviated to APM.

Figure 4.2: The experimental SP-isotherm for DSPC, taken from I. Kubo *et al.* *Thin Solid Films* 393.1-2 (2001), pp. 80–85.



<sup>50</sup> Constant number of particles, volume, and temperature.

equilibration simulation. This equilibration simulation involved continuing the use of the semi-isotropic barostat, with the *xy*-area of the box fixed, allowing the system to relax at a pressure of 1 bar in the *z*-dimension. Following the application of the pair of semi-isotropic barostats, the thickness of the water layer was typically in the region of 30 Å. The equilibration period was 1 ns, following which the 50 ns NVT<sup>50</sup> ensemble production simulations were run, on which all analyses were conducted.

### 4.3 DATA ANALYSIS

#### 4.3.1 Traditional layer-model analysis

In order to provide a point of comparison for the simulation-derived methods, the chemically-consistent reflectometry model developed in Chapter 3 was used for the analysis of the experimental data. Two modifications were made to the methodology described in Chapter 3. The first was that the volume of the phospholipid tail group,  $V_t$  was constrained based on the APM,<sup>51</sup>

$$V_t = d_t \text{APM}, \quad (4.1)$$

where  $d_t$  is the tail layer thickness. The result of this constraint is that both the monolayer model and the simulation-derived models were constrained equally by the measured surface coverage. The second modification was to constrain the head group volume to a value of 339.5 Å, in agreement with the work of Kučerka *et al.*<sup>52</sup> and Balgavý *et al.*<sup>53</sup> This constraint was possible on this occasion as the monolayer was at the air-water interface, compared to the air-DES interface previously. A uniform background<sup>54</sup> and a scale factor were then determined using `refnx` to offer the best agreement

<sup>52</sup> Kučerka *et al.*, see n. 58.

<sup>53</sup> Balgavý *et al.*, see n. 58.

<sup>54</sup> This was limited to lie within 10 % of the highest *q*-vector reflected intensity.

Table 4.2: The areas per molecule (APMs) associated with particular SPs and the size of the *x*- and *y*-cell dimension for a simulation of 100 phospholipid molecules.

| $\pi / \text{mNm}^{-1}$ | APM / Å <sup>2</sup> | <i>xy</i> -cell length / Å |
|-------------------------|----------------------|----------------------------|
| 20                      | 47.9                 | 69.1                       |
| 30                      | 46.4                 | 68.1                       |
| 40                      | 45.0                 | 67.1                       |
| 50                      | 44.6                 | 66.0                       |

| Contrast  | d <sub>13</sub> -DSPC | d <sub>70</sub> -DSPC | d <sub>83</sub> -DSPC | h-DSPC |
|---|-----------------------|-----------------------|-----------------------|--------|
| $b_{\text{head}} 10 \times 10^{-4} \text{ \AA}$ | 19.54                 | 11.21                 | 24.75                 | 6.01   |
| $b_{\text{tail}} 10 \times 10^{-4} \text{ \AA}$ | -3.58                 | 69.32                 | 69.32                 | -3.58  |

Table 4.3: The different scattering lengths of the head and tail phospholipid components.

between the calculated reflectometry profile and that measured experimentally.

The experimental data from all seven contrasts were co-refined to a single monolayer model at each SP, where the head thickness, tail thickness, and interfacial roughness were allowed to vary. The co-refinement of measurements at different SPs was not required, as there was a substantial number of contrasts measured. The values of the head and tail scattering lengths are given in Table 4.3, while the SLD of the super and subphase were taken as  $0 \text{ \AA}^{-1}$ , and  $6.35 \text{ \AA}^{-1}$  and  $0 \text{ \AA}^{-1}$  respectively. For each co-refinement of seven NR measurements, there were a total of five degrees of freedom in the fitting process, and the fitting was performed using a differential evolution algorithm. As with the work of Chapter 3, Markov chain Monte Carlo sampling was used to obtain experimental uncertainties on the fitted model, enabled by the `emcee` package.<sup>55</sup> The same protocol for this sampling was used herein.

<sup>55</sup> Foreman-Mackey *et al.*, see n. 36.

### 4.3.2 Simulation-derived analysis

A custom-class, `md_simulation`, was developed for `refnx`<sup>56</sup> that enabled the determination of a reflectometry profile from simulation, using a similar method to that employed in previous work, such as Dabkowska *et al.*<sup>57</sup> The Abelès layer model formalism is applied to layers, the SLD of which is drawn directly from the simulation, and the thickness of which is defined. The layer thickness used was  $1 \text{ \AA}$  for the Slipid and Berger potential model simulations, with an interfacial roughness between these layers of  $0 \text{ \AA}$ . For the MARTINI potential model, a layer thickness of  $4 \text{ \AA}$  was used, with an interfacial roughness of  $0.4 \text{ \AA}$ .<sup>58</sup> Each of the 50 ns production simulations were analysed each 0.1 ns, and the SLD profiles were determined by summing the scattering lengths,  $b_j$ , for each of the atoms in a given layer.

$$\text{SLD}_n = \frac{\sum_j b_j}{V_n}, \quad (4.2)$$

where,  $V_n$  is the volume of the layer  $n$ , obtained from the simulation cell parameters in the plane of the interface and the defined layer thickness. A uniform background<sup>59</sup> and a scale factor were then determined using `refnx` to offer the best agreement between the calculated reflectometry profile and that measured experimentally.

<sup>56</sup> Nelson *et al.*, see n. 29; Nelson *et al.*, see n. 66.

<sup>57</sup> Dabkowska *et al.*, see n. 40.

<sup>58</sup> The motivation for this is discussed in Section 4.4.2.

<sup>59</sup> This was limited to lie within 10 % of the highest  $q$ -value reflected intensity.

### 4.3.3 Comparison between monolayer model and simulation-derived analysis

In order to assess the agreement between the model from each method, the following goodness-of-fit metric was used, following the transformation of the data into  $Rq^4$  space,

$$\chi^2 = \sum_{i=1}^{N_{\text{data}}} \frac{[R_{\text{exp}}(q_i) - R_{\text{sim}}(q_i)]^2}{[\delta R_{\text{exp}}(q_i)]^2}, \quad (4.3)$$

where,  $q_i$  is a given  $q$ -vector,  $R_{\text{exp}}(q_i)$  is the experimental reflected intensity,  $R_{\text{sim}}(q_i)$  is the simulation-derived/traditionally-developed reflected intensity, and  $\delta R_{\text{exp}}(q_i)$  is the resolution function of the experimental data.

<sup>60</sup> Abbreviated to wph.

The number of water molecules per head group<sup>60</sup> was also compared between the different methods. This was obtained from the monolayer model by considering the solvent fraction in the head-layer,  $\phi_h$ , the volume of the head group,  $V_h$ , and taking the volume of a single water molecule to be  $29.9 \text{ \AA}^3$ ,<sup>61</sup>

$$\text{wph} = \frac{\phi_h V_h}{29.9 - 29.9\phi_h}. \quad (4.4)$$

<sup>61</sup> Found from the density of water as  $997 \text{ kg m}^{-3}$ .

<sup>62</sup> The phospholipid heads, phospholipid tails, and water.

In MD simulations, the number densities, in the  $z$ -dimension, for each of the three components<sup>62</sup> may be obtained directly from the trajectory. In order to determine the number of water molecules per headgroup from the MD simulations, a head-layer region was defined as that which contained the middle 60 % of the phospholipid head number density. The ratio between the water density and the phospholipid head density was then found within the head-layer region.

### 4.3.4 Simulation trajectory analysis

In order to use the MD trajectory to guide the future development of the chemically-consistent layer model, it was necessary to investigate the solvent penetration into the head group regions of the phospholipids, the roughness of each interface and the phospholipid tail length. The solvent penetration was determined using the intrinsic surface approach, as detailed by Allen *et al.*<sup>63</sup> The intrinsic surface approach enables the calculation of the solvent penetration without the effect of the monolayer roughness. This involves taking the  $z$ -dimension position of each atom with respect to an anchor point, in this work the anchor point was taken as the phosphorus atom of the phospholipid head that was closest to the atom in the  $xy$ -plane. The roughness was probed by investigating the variation in positions from the start, middle and end of each of the head and tail groups. The start of the phospholipid head was defined as the nitrogen atom, the middle the phosphorus and the end the tertiary carbon, while the start of the phospholipid tail was defined as the carbonyl carbon atom, the middle the ninth carbon in the tail, and

<sup>63</sup> D. T. Allen *et al.* *Phys. Chem. Chem. Phys.* 18.44 (2016), pp. 30394–30406; S. A. Pandit *et al.* *J. Chem. Phys.* 119.4 (2003), pp. 2199–2205.

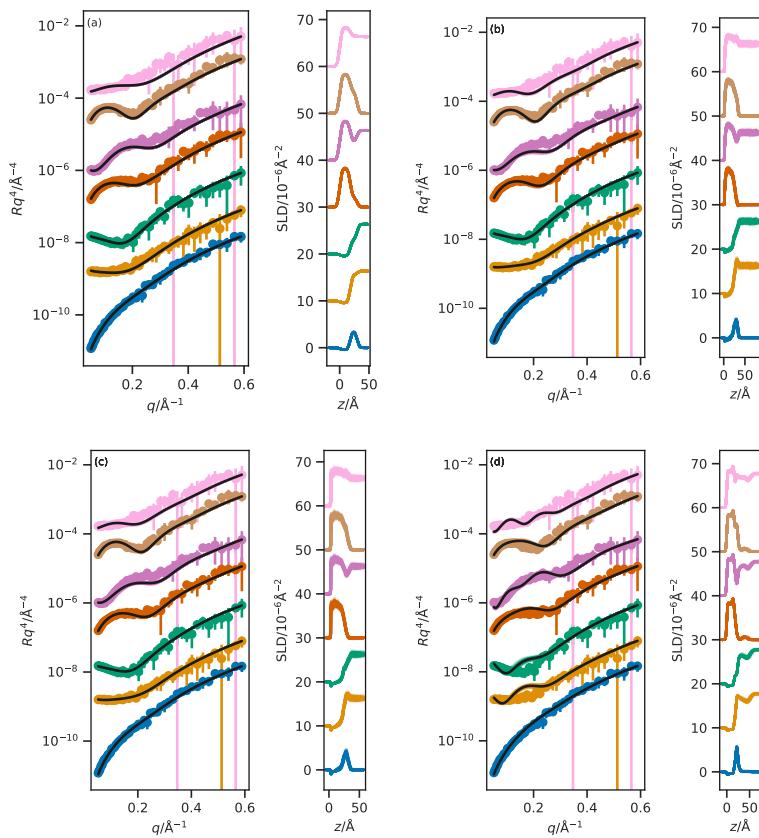


Figure 4.3: The NR profiles (left) and SLD profiles (right) determined at an APM associated with a SP of  $30 \text{ mN m}^{-1}$  for; (a) the chemically-consistent model, (b) the Lipid all-atom potential model simulations, (c) the Berger united-atom potential model simulations, and (d) the MARTINI coarse-grained potential model simulations. From top-to-bottom the contrasts are as follows;  $d_{83}\text{-D}_2\text{O}$ ,  $d_{83}\text{-ACMW}$ ,  $d_{70}\text{-D}_2\text{O}$ ,  $d_{70}\text{-ACMW}$ ,  $\text{h-D}_2\text{O}$ ,  $d_{13}\text{-D}_2\text{O}$ ,  $d_{13}\text{-ACMW}$ . The different contrast NR profiles have been offset in the y-axis by an order of magnitude and the SLD profiles offset in the y-axis by  $10 \times 10^{-6} \text{ Å}^{-2}$ , for clarity.

the end the final carbon in the tail. The distribution of each of these atom types was determined by finding the 95 % quantile for the position in the  $z$ -dimension and comparing the spread of the mean and the upper quantile. Finally, the tail length distance,  $t_t$ , was found as the distance from the carbonyl carbon atom to the final primary carbon atom of the phospholipid tail. All of these analyses used the MDAnalysis package.<sup>64</sup>

#### 4.4 RESULTS & DISCUSSION

Figure 4.3 compared the reflectometry and SLD profiles from each of the different methods at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ .<sup>65</sup> In addition, the  $\chi^2$  between each of the models and the experimental data for each contrast at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ , the average  $\chi^2$  and standard deviation for each method are given in Table 4.4.<sup>66</sup>

##### 4.4.1 Traditional analysis

The chemically-consistent model was used to determine the structure of the phospholipid monolayer, Table 4.5 gives the optimum values for the parameters that were varied in the model. It is clear

<sup>64</sup> R. Gowers *et al.* In: Python in Science Conference. Austin, Texas, 2016, pp. 98–105; N. Michaud-Agrawal *et al.* *J. Comput. Chem.* 32.10 (2011), pp. 2319–2327.

<sup>65</sup> The data for the other APMs can be found in Appendix C, however, the trends are similar at all SPs.

<sup>66</sup> The same data for other APMs is available Appendix C.

Table 4.4: The  $\chi^2$  values for each of the reflectometry models at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ .

| Contrast                          | Chemically-consistent | Slipids             | Boger               | MARTINI               |
|-----------------------------------|-----------------------|---------------------|---------------------|-----------------------|
| h-D <sub>2</sub> O                | 39.57                 | 166.19              | 119.72              | 1489.50               |
| d <sub>13</sub> -D <sub>2</sub> O | 84.34                 | 235.50              | 90.35               | 2042.12               |
| d <sub>13</sub> -ACMW             | 82.05                 | 92.21               | 85.29               | 165.94                |
| d <sub>70</sub> -D <sub>2</sub> O | 148.67                | 659.32              | 583.49              | 1882.23               |
| d <sub>70</sub> -ACMW             | 155.62                | 83.02               | 92.85               | 362.66                |
| d <sub>83</sub> -D <sub>2</sub> O | 314.99                | 343.03              | 424.55              | 3141.62               |
| d <sub>83</sub> -ACMW             | 102.84                | 144.51              | 315.74              | 737.54                |
| Average                           | $132.58 \pm 83.22$    | $246.25 \pm 249.21$ | $244.57 \pm 238.79$ | $1403.09 \pm 1044.80$ |

from this Table, that as the SP is increased, as expected, and as found previously,<sup>67</sup> the overall thickness of the monolayer increases. The thickness increase for the phospholipid tails may be associated with the straightening of the tails with respect to the interface normal, while the thickness increase of the head groups has been noted previously for DSPC.<sup>68</sup>

It would be anticipated that as the SP increases, there would be a corresponding decrease in the volume fraction of solvent in the head group.<sup>69</sup> However, for DSPC, the volume fraction of the solvent appears to be constant (or even increase slightly) with increasing SP. This is due to the decision to constrain the volume of the phospholipid head, which may decrease with increasing SP.

Hollinshead *et al.*<sup>70</sup> suggest a tail volume of 972 Å<sup>3</sup> from the density data. However, the values found in this work are substantially lower, at ~850 Å<sup>3</sup>. This reduction, of ~12 %, agrees well with the work of Campbell *et al.*<sup>71</sup> and Small,<sup>72</sup> which suggest that under the SP investigated in this work a reduction of the tail volume of up to 15 % may be observed. The model layer structure from the chemically-consistent method provides a satisfactory description of the monolayer structure. However, the use of an MD-driven analysis method may provide greater insight into the chemical nature of the monolayer.

#### 4.4.2 MARTINI

Initially, the MARTINI coarse-grained simulations were analysed with a layer thickness of 1 Å and an interfacial roughness of 0 Å, in a similar fashion to the other potential models. However, as can be seen in Figure 4.4 there is a clear ordering effect present in the MARTINI water, despite the use of the polarised water model. The effect of this ordering on the SLD profile, and therefore the reflectometry profile, can be reduced by using a larger layer thickness and introducing an interfacial roughness. Therefore, in the results discussed below the MARTINI potential model simulation were analysed using a layer thickness of 4 Å and an interfacial roughness of 0.4 Å. It is noted that this structuring may be reduced through the use of a less ordered wall<sup>73</sup> at the extreme of the simulation cell, however, the aim was to reproduce the experimental conditions using off-the-shelf tools and this would require custom modifications not easily available. Alternatively, it may be possible to effect the presence of this structuring through the inclusion of ~10 % of antifreeze MARTINI beads alongside the normal MARTINI water. However, this method has been noted to also give structuring effects in the presence on an ordered wall.<sup>74</sup>

It can be seen from Table 4.4 and Figure 4.3 that even with the larger layer thickness and adding interlayer roughness, the MARTINI potential model simulations do not effectively reproduce the reflectometry profile. Furthermore, it is noted that the agreement

<sup>67</sup> Mohwald, see n. ??; Vaknin *et al.*, see n. 89.

<sup>68</sup> Hollinshead *et al.*, see n. 43.

<sup>69</sup> Bayerl *et al.*, see n. 56.

<sup>70</sup> Hollinshead *et al.*, see n. 43.

<sup>71</sup> Campbell *et al.*, see n. 26.

<sup>72</sup> Small, see n. 41.

<sup>73</sup> Koutsoubas, see n. 41.

<sup>74</sup> S. J. Marrink *et al.* *Phys. Chem. Chem. Phys.* 12.9 (2010), p. 2254.

Table 4.5: The values for the parameters allowed to vary in the fitting of the chemically-consistent model, at each SP measured.

| SP/<br>mN m <sup>-1</sup> | $d_h/\text{\AA}$ | $d_t/\text{\AA}$ | $\sigma_{t,h,s}/\text{\AA}$ | $\phi_h \times 10^{-2}$ | $V_t/\text{\AA}^3$ |
|---------------------------|------------------|------------------|-----------------------------|-------------------------|--------------------|
| 20                        | $9.60 \pm 0.38$  | $17.89 \pm 0.20$ | $3.31^{+0.03}_{-0.01}$      | $16.69 \pm 4.28$        | $857.17 \pm 9.40$  |
| 30                        | $11.65 \pm 0.44$ | $17.99 \pm 0.22$ | $3.31^{+0.04}_{-0.01}$      | $30.38^{+3.48}_{-3.64}$ | $834.86 \pm 10.15$ |
| 40                        | $13.88 \pm 0.52$ | $17.85 \pm 0.21$ | $3.31^{+0.03}_{-0.01}$      | $44.18 \pm 2.99$        | $803.38 \pm 9.37$  |
| 50                        | $14.12 \pm 0.47$ | $19.01 \pm 0.19$ | $3.31^{+0.06}_{-0.01}$      | $45.55^{+2.33}_{-2.45}$ | $848.00 \pm 8.34$  |

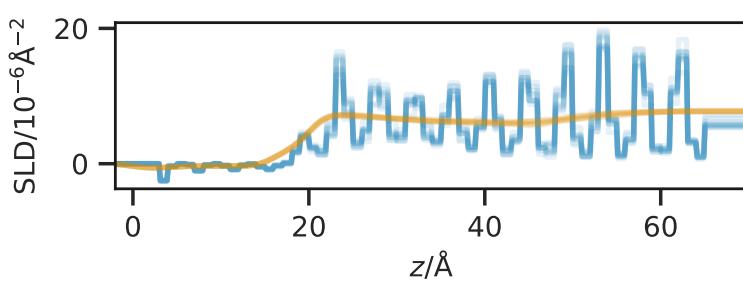


Figure 4.4: A comparison of the scattering length density profile for the MARTINI potential model simulations at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ ; the blue line shows the data where the layer thickness was  $1 \text{ \AA}$  and no interfacial roughness and the orange line shows that with a layer thickness of  $4 \text{ \AA}$  and a roughness of  $0.4 \text{ \AA}$ .

with the contrasts containing  $\text{D}_2\text{O}$  is particularly poor. This is most likely an artefact of the structuring effect mentioned above which cannot be completely removed.

However, the agreement for the samples where the contrast uses ACMW, where the water is effectively removed from the SLD profile is also poor. This indicates that there are other artefacts limiting the applicability of the MARTINI potential model. One such artefact is clear from investigating the calculated length of the hydrocarbon tail from the MARTINI simulation, at an APM associated with an SP of  $30 \text{ mN m}^{-1}$ , which was found to be  $16.60^{+1.65}_{-1.88} \text{ \AA}$ , significantly less than the  $24.3 \text{ \AA}$  estimated by the Tanford equation.<sup>75</sup> This is due to the nature of the MARTINI's 4-to-1 beading process, as DSPC has a hydrocarbon tail consisting of 18 carbon atoms, and it is not possible to bead such a chain accurately with the MARTINI potential model. In this work, a MARTINI phospholipid molecule was used with 4 MARTINI beads making up the chain.<sup>76</sup> Applying the Tanford equation to a hydrocarbon chain of such a length results in an anticipated length of  $18.7 \text{ \AA}$ , which agrees better with that found from the simulation.

In addition to the disagreement from the tail beading process, there is also a clear problem with respect to the solvation of the head layer by polarisable water beads. It can be seen that the number of water molecules per head group in the MARTINI potential model is typically  $1.34 \pm 0.35$ , this is the value at an APM associated with an SP of  $30 \text{ mN m}^{-1}$ . The chemically-consistent model, however, gives a value of  $5.49^{+0.86}_{-0.84}$ . It is clear that the 4-to-1 beading present in the MARTINI potential model is creating water molecules that are too large to intercalate into the head layer structure, causing a reduction in the number of waters per head group.

The requirement for a 4-to-1 beading strategy for the MARTINI potential model is a significant weakness. A better method may be limiting experiments to a system that can be modelled exactly or the use of a different beading model. However, I am not aware of an off-the-shelf coarse-grained potential model that would easily offer the exact beading of DSPC.

<sup>75</sup> Tanford, see n. 70, which has the form  $t_t = (1.5 + 1.265)n_c \text{ \AA}$ , where  $t_t$  is the length and  $n_c$  is the number of carbon atoms.

<sup>76</sup> This corresponds to an all-atom hydrocarbon chain of 16 atoms.

#### 4.4.3 Comparison with other simulations

Table 4.4 shows that both the Slipid and Berger potential model simulations agree well with the experimental data, with the Slipid potential model offering a slight improvement over the Berger. The quality of agreement between these higher-resolution potential models and the chemically-consistent model is relatively similar. However, the chemically-consistent model still offers a better fit to the experimental data than those determined from MD simulation.

The result that the chemically-consistent model offers better agreement with the data than those from even all-atom simulations is to be expected by considering the level of constraint present implicitly when determining the reflectometry profile directly from the simulation. While the chemically-consistent model constrains the layer model to ensure that the number of phospholipid head groups is the same as the pairs of tail groups, those from MD simulation have more realistic chemical constraints present from the potential model.<sup>77</sup> The quality of the agreement from this multi-modal approach is sufficient for such a method to be applied regularly to the analysis of NR.

Both the Slipid and Berger potential model simulations produced values for the tail length that were in better agreement with that from the Tanford equation than the MARTINI potential model simulations. For the Slipid potential model, with simulations at an APM associated with an SP of  $30\text{ mN m}^{-1}$  the tail length was found to be  $20.17^{+1.41}_{-7.39}\text{ \AA}$ , while for the Berger potential model, at the same APM, a value of  $19.80^{+1.59}_{-8.17}\text{ \AA}$  was obtained. Neither is quite as large as the  $24.3\text{ \AA}$  from the Tanford equation.<sup>78</sup>

Using the MD simulations and the chemically-consistent model, it is possible to compare the number of water molecules per head group. From the Slipid and Berger potential model simulations, the number of water molecules per head group at an APM associated with an SP of  $30\text{ mN m}^{-1}$  was found to be  $6.41^{+1.63}_{-0.76}$  and  $5.49^{+0.68}_{-0.53}$  respectively. These are in good agreement with the value of  $5.49^{+0.86}_{-0.84}$  found from the chemically-consistent model, using Equation 4.4.

It should be noted that to obtain the 50 ns production run simulation using the all-atom Slipid potential model required over 13 days of using 32 cores of the SCARF computing resource. This is non-trivial and therefore not necessarily applicable to all NR measurements. However, the use of a 2 fs timestep could reduce this time significantly. Additionally, Figure 4.5 shows the results from the first 5 ns of the Slipid potential model simulations, at an APM associated with an SP of  $30\text{ mN m}^{-1}$ , and already good agreement with the data is apparent. It is important to acknowledge that the length of simulation required may be extremely system specific. Furthermore, recent developments of MD simulations of graphical processing units<sup>79</sup> may allow for significant speed up of the

<sup>77</sup> Such as the bonding between atoms, and the non-bonded potentials.

<sup>78</sup> It should be noted that this value is considered a theoretical maximum for a fully extended carbon chain, which is unlikely to occur, in a liquid phase monolayer, in reality, due to entropic considerations.

<sup>79</sup> Abbreviated to GPUs.

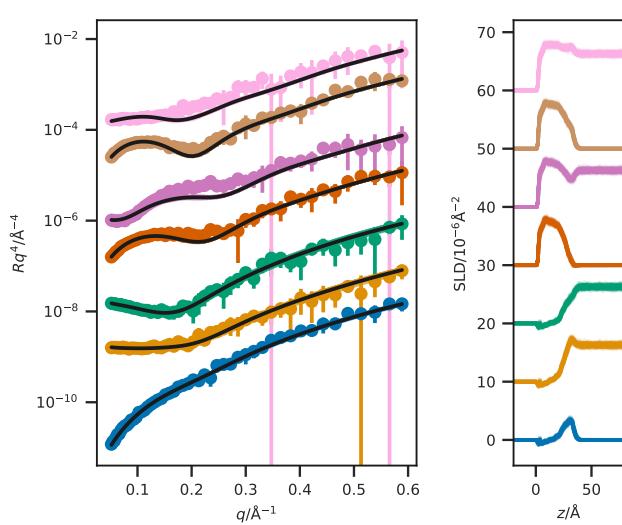


Figure 4.5: The reflectometry and SLD profiles obtained from the first 5 ns of the Slipid potential model simulation, at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ . From top-to-bottom the contrasts are as follows;  $d_{83}-\text{D}_2\text{O}$ ,  $d_{83}-\text{ACMW}$ ,  $d_{70}-\text{D}_2\text{O}$ ,  $d_{70}-\text{ACMW}$ ,  $\text{h}-\text{D}_2\text{O}$ ,  $d_{13}-\text{D}_2\text{O}$ ,  $d_{13}-\text{ACMW}$ . The different contrast reflectometry profiles have been offset in the  $y$ -axis by an order of magnitude and the SLD profiles offset in the  $y$ -axis by  $10 \times 10^{-6}\text{\AA}^{-2}$ , for clarity.

simulations. The nearly as accurate Berger potential model simulations, only approximately 2 days of the same compute resource was required. This suggests that by using a larger timestep, shorter simulations, and the power of GPU-based MD engines, it may be possible to run these simulations alongside experiments at large facilities to aid interpretation and analysis.

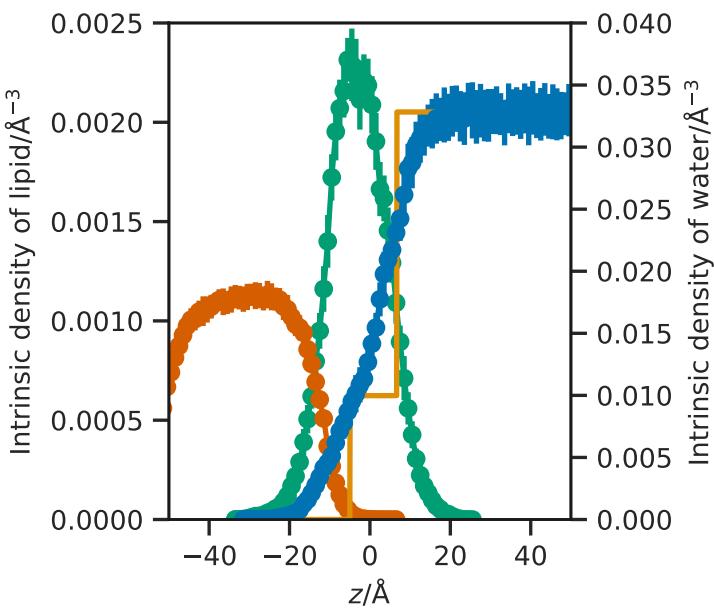
#### 4.4.4 Using the Slipid potential model simulations to improve the mono-layer model

Despite the chemically-consistent model offering a small improvement in agreement over the Slipid potential model simulation, it is possible to use the MD simulations to improve this model. A possible improvement can be found from considering Figure 4.6, which shows the solvent penetration of the phospholipid heads, using the intrinsic surface approach to remove the effect of the interfacial roughness. It is clear that the plot is not step-wise as is obtained from the uniform solvation model that is commonly used in traditional layer models. Nor is the distribution sigmoidal, as there is a small deviation in the region of the ester group of the phospholipid heads. This is either due to the hydrophilic interaction of the carbonyl moiety or from pockets of water forming at the air-water interface. Regardless of the mechanism, this suggests that a different solvation model should be considered for a realistic description of the solvent penetration.

Figure 4.6 also shows that without the presence of the roughness, the distribution of the head groups is relatively normal. This agrees well with the model used previously to fit the experimental data by Hollinshead *et al.*,<sup>80</sup> where Gaussian functions were used to describe the phospholipid head and tail groups. However, the tail group distribution is not Gaussian and this previous method

<sup>80</sup> Hollinshead *et al.*, see n. 43.

Figure 4.6: The simulation time-averaged intrinsic density profile of the water molecules (blue dots), and phospholipid components (head groups: green dots, tail groups: red dots) at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ , where the phosphorus atoms of the phospholipid heads create the intrinsic surface at  $z = 0\text{\AA}$ , and the equivalent number density from the chemically-consistent model (orange line).



<sup>81</sup> I. V. Kozhevnikov. *Crystallogr. Rep.* 57.4 (2012), pp. 490–498; Campbell *et al.*, see n. 26.

<sup>82</sup> Similar data at other SPs can be found in Appendix C.

<sup>83</sup> At the head group.

failed to include any additional factors to account for interfacial roughness. Previous work has suggested that when only a single phospholipid type is present, the roughness between the layers should be conformal in nature, that is it should be carried uniformly through the layers.<sup>81</sup> However, from the investigation of the SLD profiles in Figure 4.3(b), it appears that the roughness between the phospholipid tails and the air is dramatically different from that at the phospholipid head-water interface. In an effort to quantify the interfacial roughness in the simulations, the method outlined in Section 4.3.4 was used. The values for the mean, 95 % quantile, and the spread between these for the  $z$ -dimension position for atoms representative of the start, middle, and end of each of the phospholipid head and tails are given in Table 4.6, for each the APM associated with a SP of  $30 \text{ mN m}^{-1}$ .<sup>82</sup> From this table, it is clear that at the very start of the phospholipid molecule<sup>83</sup> the roughness is very large with a value of  $\sim 10 \text{ \AA}$  for the nitrogen

Table 4.6: The mean, 95 % quantile, and their spread for the  $z$ -dimension position of atoms representative of difference parts of the phospholipid, at an APM associated with a SP of  $30 \text{ mN m}^{-1}$ .

| Position     | Mean/\text{\AA} | 95 % quantile/\text{\AA} | Spread/\text{\AA} |
|--------------|-----------------|--------------------------|-------------------|
| Start-Head   | 66.6            | 76.6                     | 10.1              |
| Mid-Head     | 67.7            | 76.6                     | 9.0               |
| End-Head     | 70.8            | 79.3                     | 8.6               |
| Start-Tail 1 | 72.2            | 80.3                     | 8.1               |
| Start-Tail 2 | 73.0            | 81.7                     | 8.6               |
| Mid-Tail 1   | 80.9            | 87.1                     | 6.2               |
| Mid-Tail 2   | 82.3            | 87.9                     | 5.6               |
| End-Tail 1   | 91.1            | 93.3                     | 2.2               |
| End-Tail 2   | 92.4            | 93.5                     | 1.1               |

atom. However, this decreases slightly within the phospholipid head, reaching a value of 8.6 Å for the end of the head group. There is then a substantial decrease noted in the phospholipid tail, going from ~8.5 Å at the start of the tail to ~1.5 Å at the end. This indicates the presence of a highly non-conformal roughness in the phospholipid monolayer of a single phospholipid type and therefore in future, it is important to consider this possibility in the use of model layer structure method.

#### 4.5 CONCLUSIONS

This chapter presents a direct comparison between a traditional method for the analysis of NR measurements with an analysis method using simulations from a range of all-atom and coarse-grained MD potential models. It was shown that the MARTINI potential model could not accurately reproduce the experimental NR data, likely, due to the limitations of the 4-to-1 beading system when applied to a carbon chain of 18 atoms.

The Berger united atom and Slipids all-atom potential models both showed good agreement with the experimental data, however, the best agreement was obtained from the traditional chemically-consistent layer model. This would be expected given that the chemically-consistent model contains many more “degrees of freedom” than the simulations which are severely chemically-constrained by the potential model.

Finally, some points from the highest resolution, Slipid, simulations were noted that may be used to improve the traditional monolayer model. For example, it is desirable to model non-uniform solvation of the head group region which would enable a more accurate modelling of the lipid monolayer and the use of a conformal roughness may not be the best constraint to apply. Application of these improvements may enable the more accurate modelling of phospholipid monolayers from NR.



# 5

## USING PARTICLE SWARM METHODS FOR SMALL ANGLE SCATTERING ANALYSIS

### ABSTRACT

The simulation of micellar species is a non-trivial task, and currently many simulations of micelles still underestimate the micelle size and are unable to model micelle-micelle interactions. This chapter represents an attempt to apply a high-performance computing compatible optimisation methodology to the generation of micellar structures, consisting of hundreds of molecules, based on the experimental small angle scattering<sup>1</sup> alone. The software `fitoog` was developed to perform this, using a severe coarse-graining method to describe the molecules within the optimisation, while the scattering profile was calculated with a MARTINI style coarse-graining. This software is open-source and well-parallelised enabling it to be run on high-performance computing resources. While `fitoog` was capable of optimising a simple test case, the PSO alone could not overcome the high dimensionality of the parameter space necessary to produce a multi-micellar system.

<sup>1</sup> SAS.

### CONTEXT

The use of classical simulation as a tool to assist SAS analysis is popular in the biomolecular community.<sup>2</sup> However, their application in the study of other soft matter species, such as micelles, is limited due to the significant computational cost of obtaining a starting structure that is similar to the experimental system. While it is possible to produce a series of simulations of various micelle sizes and compare these with experimental data, one of the targets for such experimental work is the inter-micelle interactions and hence the large system sizes mean that these methods quickly become unfeasible. Herein, the generation of a starting structure for a molecular dynamics<sup>3</sup> simulation from scattering data is treated as an optimisation problem and a particle swarm optimisation<sup>4</sup> algorithm is applied to attempt to resolve it. The aim is to obtain a starting structure that is representative of the experimental system,

<sup>2</sup> S. J. Perkins *et al.* *J. Appl. Crystallogr.* 49.6 (2016), pp. 1861–1875; Hub, see n. 68.

<sup>3</sup> MD.

<sup>4</sup> PSO.

including scattering from inter-micelle interactions.

## 5.1 INTRODUCTION

SAS is a popular technique for the structural investigation of surfactant micelles.<sup>5</sup> Typically, coarse shape-based modelling<sup>6</sup> is used for the analysis, which allows for the classification of the micelle shape and interactions. However, as with reflectometry, there has been a growing interest in the use of atomistic simulations as a multi-modal analysis tool for solution scattering methods,<sup>7</sup> to provide greater detail about the structure.

The use of atomistic simulation as an analysis method in the study of biomolecules has been popular for many years.<sup>8</sup> These have built on the success of biomolecular simulation and crystallography, using structural information from the protein data bank<sup>9</sup> and applying popular all-atom potential models. Typically, this is used for the study of systems where the solution state differs significantly from that present in the crystal, such as flexible protein multimers. The benefit of MD is that an ensemble of structures can be represented in a single simulation trajectory.<sup>10</sup>

The ability for MD simulations to represent an ensemble of structures has led to the application of interesting aspects from probability theory. In particular, Bayesian inference to understand the presence and population of different structures in solution. For example, in the work of Bowerman *et al.*,<sup>11</sup> accelerated MD simulations<sup>12</sup> were performed on an all-atom representation of the protein multimer, tri-ubiquitin. The scattering profile was calculated from the simulation and compared with experimental data, and the agreement between the simulation and experiment assessed in a Bayesian fashion, with a uniform prior probability. This methodology showed that the presence of a two-state ensemble was more likely than the single state that would be obtained from, for example, a crystallographic study. Therefore, giving a more realistic understanding of the solution structure of the system.

The simulation of a surfactant micelle is inherently more complex than for a protein ensemble, due to the greater number of states<sup>13</sup> available under thermodynamic conditions. This means that there is rarely a suitable starting structure for an MD simulation, instead, it is necessary to simulate the system from a random configuration or artificially create a micelle structure to start from. Early work on the simulation of surfactant aggregation involved the 4.5 ns simulation of 42 sodium dodecyl sulfate molecules.<sup>14</sup> These simulations started with a random solution of surfactant molecules and resulted in two small surfactant aggregates consisting of 17 and 25 molecules each. The aggregates simulated were much smaller than those measured experimentally by small angle neutron scattering,<sup>15</sup> which consist of  $79 \pm 1$  molecules.<sup>16</sup> However, this deviation may be due to the fact that the simulations were performed with the simulation thermostat at 60 °C, as the SDS micelle size is noted to reduce with increasing temperature.<sup>17</sup>

<sup>5</sup> Sanchez-Fernandez *et al.*, see n. 13.

<sup>6</sup> Such as that introduced in Section 2.2.7.

<sup>7</sup> Ivanović *et al.*, see n. 36.

<sup>8</sup> S. J. Perkins *et al.* *J. Mol. Biol.* 221 (1991), pp. 1345–1366; M. O. Mayans *et al.* *Biochem. J.* 331 (1995), pp. 283–291; Hub, see n. 68.

<sup>9</sup> RCSB PDB: Protein Data Bank, see n. 122.

<sup>10</sup> P.-C. Chen *et al.* *Biophys. J.* 107.2 (2014), pp. 435–447; Bowerman *et al.*, see n. 33.

<sup>11</sup> Bowerman *et al.*, see n. 33.

<sup>12</sup> Similar to traditional MD but with use of a “boost” potential to improve sampling.

<sup>13</sup> For example, micelles of different sizes.

<sup>14</sup> M. Tarek *et al.* *J. Mol. Liq.* 78.1-2 (1998), pp. 1–6.

<sup>15</sup> SANS.

<sup>16</sup> Hassan *et al.*, see n. 25.

<sup>17</sup> S. Hayashi *et al.* *J. Phys. Chem.* 84.7 (1980), pp. 744–751.

<sup>18</sup> However, it is possible that the application of a coarse-grained potential model may also achieve this.

<sup>19</sup> J.-B. Maillet *et al.* *Phys. Chem. Chem. Phys.* 1.23 (1999), pp. 5277–5290.

<sup>20</sup> Between 15 and 20 molecules

<sup>21</sup> Tarek *et al.*, see n. 14; Maillet *et al.*, see n. 19.

<sup>22</sup> M. Jorge. *Langmuir* 24.11 (2008), pp. 5714–5725.

<sup>23</sup> Like the Berger potential model from Chapter 4, where all hydrogen atoms are integrated into the atoms to which they are bound.

<sup>24</sup> Tarek *et al.*, see n. 14.

<sup>25</sup> Sanders *et al.*, see n. 21.

<sup>26</sup> A. Malliaris *et al.* *J. Phys. Chem.* 89.12 (1985), pp. 2709–2713; N. Kamenka *et al.* *Langmuir* 11.9 (1995), pp. 3351–3355.

<sup>27</sup> J. C. Shelley *et al.* *Curr. Opin. Colloid Interface Sci.* 5.1-2 (2000), pp. 101–110.

<sup>28</sup> A. Vishnyakov *et al.* *J. Phys. Chem. Lett.* 4.5 (2013), pp. 797–802.

<sup>29</sup> Abbreviated to cmc.

This increased temperature is used to improve sampling and more efficiently reach equilibrium.<sup>18</sup> Maillet *et al.* studied the mechanism of micelle formation with a 3 ns simulation of the cationic *n*-nonyltrimethylammonium chloride; observing micelle formation, fragmentation, and monomer exchange.<sup>19</sup> The micelles that were formed were smaller than would be expected,<sup>20</sup> however, this work suggested that the initial stages of the micellisation process are dominated by collisions between aggregates, whereas monomer exchanges are more frequent closer to equilibrium.

These early examples of the simulation of a micellisation event both used all-atom potential models,<sup>21</sup> however, it is clear that in order to obtain an experimentally realistic micelle from simulation a much longer simulation would be required. In Section 2.3.3, the use of potential model coarse-graining as a method to increase the “real-time” length of an MD simulation was introduced. This can be used as a tool to simulate micellisation events from a random solution of surfactants. For example, in the work of Jorge,<sup>22</sup> a united-atom potential model<sup>23</sup> for *n*-decyltrimethylammonium bromide was used to simulate micellisation at 80 °C. This work showed that the use of a united-atom potential model allows for more efficient simulation, but also resulted in a larger mass average cluster size. However, again a value of only ~25 was reached after 14 ns, as with the work of Tarek *et al.*,<sup>24</sup> these small micelles might be realistic at these elevated temperatures, however, there appears to be no experimental structural data, such as SANS measurements, for *n*-decyltrimethylammonium bromide micelles at 80 °C. Therefore, it is not possible to assess the accuracy of these results. Sanders and Panagiotopoulos attempted to use the MARTINI coarse-grained potential model to simulate the micellisation of the zwitterionic dodecylphosphocholine.<sup>25</sup> This was simulated at 97 °C for 1.8 μs resulting in a trajectory where the cluster size mode was 41, the authors noted that experimentally at 25 °C a micelle size of 56 ± 5 is expected. Again, it was also noted that this variation may be due to the increased simulation temperature.<sup>26</sup> Furthermore, in each of these examples, it was only possible to simulate the formation of a single, or possibly a pair of micelles, which would not be effective for the modelling of inter-micelle interactions.

The use of mesoscale simulation techniques, such as dissipative particle dynamics, are common for the simulation of surfactant molecules in solution.<sup>27</sup> DPD simulations are similar to coarse-grained MD simulations, however, with the inclusion of additional dissipative and random forces. These serve both as a thermostat and to make up for degrees of freedom lost to coarse-graining, meaning that sites within a DPD model typically account for more atoms than in even a MARTINI coarse-grained MD simulation. An example of the use of DPD to study micellisation is that of Vishnyakov *et al.*,<sup>28</sup> where the critical micelle concentration<sup>29</sup> of nonionic surfactants was investigated. This work found cmc values

and mean micelle aggregation numbers that agreed qualitatively with experimental measurements. While mesoscopic techniques, such as DPD, are an interesting tool for the study of rough micelle structure, these methods lack the structural detail that would allow for greater insight into the system over SAS alone.

It is clear from the discussion above that the simulation of experimentally realistic micelles from a random structure requires very long simulations. Therefore a different approach is necessary, essentially this can take one of two routes; building a micelle-like starting structure based on a simple analysis of the experimental data or tackling the problem as an optimisation challenge, where the aim is to optimise the atomistic, or near-atomistic structure to the experimental data. The former method was that applied by Ivanovic *et al.* in their recent work,<sup>30</sup> where simulations were performed and compared with experimental scattering from micelles of *n*-dodecyl- $\beta$ -D-maltoside and *n*-decyl- $\beta$ -D-maltoside. This work used two approaches to determine the size of the micelle that should be simulated; the first used the forward scattering<sup>31</sup> to determine the density and therefore the aggregation number of the micelle,<sup>32</sup> while the second involved the simulation of a series of micelles of different sizes and the calculation of scattered intensity at a single scattering vector. These methods gave good agreement with the size of the micelle, and the scattering profiles calculated from 50 ns MD simulations offered reasonable agreement with those measured experimentally at a series of temperatures. The experimental scattering profiles were then used as an energetic restraint on the simulation improving the agreement between experiment and simulation.

The work of Ivanovic *et al.* benefited from the monodispersity of the particular micellar system chosen. However, if a more polydisperse system were being studied, this approach may require significantly more computation, as many more aggregation numbers of surfactants would need to be considered. Additionally, in order to perform a “realistic” simulation of a micelle solution it is often necessary to simulate multiple micelles in order to obtain the structure factor<sup>33</sup> present in experimental data. In this chapter, I discuss a truly model-free approach for the analysis of micellar SAS data, by applying a global-optimisation process to a near-atomistic system. This involved using a PSO algorithm to fit an experimental scattering profile. It was believed that the application of a population-based optimisation method would result in a series of suitable structures, allowing for a more realistic understanding of the ensemble structures that are present for micelles in solution. The aim of this approach is to create a starting structure for future MD simulations based on the scattering profile, this would ease the ability for experimental users to perform simulations of the system under study. Furthermore, by applying a severely coarse-grained description of the molecular species, it may be possible to generate

<sup>30</sup> Ivanović *et al.*, see n. 36.

<sup>31</sup> This is the scattering at  $q = 0$ .

<sup>32</sup> J. Lipfert *et al.* *J. Phys. Chem. B* 111, 12427–12438 (2007), pp. 12427–12438.

<sup>33</sup> Which arises from inter-micelle interactions.

a starting structure for molecular simulations consisting of multiple micelles.

## 5.2 METHODS

The computational methodology described herein has been implemented in the open-source C\_MPI program<sup>34</sup> *fitoog*.<sup>35</sup>

### 5.2.1 Simulation Methodology

The *fitoog* software takes a series of input files that define the molecules, intramolecular interactions, and scattering lengths for the constituent particles. The molecule input file is a space-separated file consisting of an index, particle name, *x*-coordinate, *y*-coordinate, *z*-coordinate, and scattering length. When *fitoog* reads in a molecular input file, a *differences* object is created, this stores the differential between the atomistic, or near-atomistic<sup>36</sup> description and the more severe coarse-grained description that is used in the PSO.

Inspired by the coarse-graining methodology applied to directional colloid self-assembly by Law *et al.*,<sup>37</sup> a severe coarse-graining methodology was developed for use on surfactant molecules in *fitoog*. This allowed for a significantly reduced parameter dimensionality to which the PSO<sup>38</sup> could be applied. The severe coarse-graining reduced the surfactant to a “director” description; where each surfactant molecule is defined by a position and a direction, shown pictorially in Figure 5.1. This reduced the parameter dimensionality to just six variables per molecule; three of which described the centre-of-mass position of the molecule and three that describe the angular orientation of the surfactant in space.<sup>39</sup> Additionally, if the molecule was a single particle in length, as is the case for the  $\text{NO}_3^-$  anion in the Section 5.3.3, then the dimensionality is just three, as the direction is arbitrary.

The methodology for the particle swarm in this work followed the implementation discussed in Section 2.4.2. Where the parameters to be optimised were the position and orientation of the input molecules.<sup>40</sup> Following each PSO iteration, the molecule director is expanded from the position variable<sup>41</sup>. This representation is then rotated based on a rotation matrix, in this work the rotation matrix was constructed by first rotating the rotation axis by  $-\phi$  and  $-\omega$ , then rotating by  $\kappa$  in around the *z*-axis, before rotating the axis back to the original position by  $\omega$  and  $\phi$ .<sup>42</sup> Following the expansion and reorientation, the scattering profile is then calculated using the Debye equation,<sup>43</sup> this was used over the more efficient Golden Vectors<sup>44</sup> or Fibonacci sequence<sup>45</sup> methods as the aim of this work was to assess the application of the PSO method and efficiency was not the initial goal. The agreement between the calculated scattering profile and the experimental input scattering was used as a

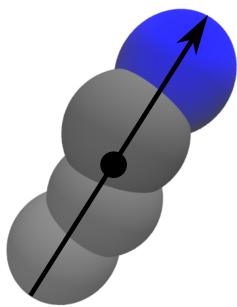


Figure 5.1: A graphical description of the severe coarse-graining applied to the MARTINI description of the *n*-decyltrimethylammonium surfactant molecule for the use of the particle swarm algorithm.

<sup>36</sup> Such as the MARTINI-style beadng use in this work.

<sup>37</sup> C. Law *et al.* *J. Chem. Phys.* 145.8 (2016), p. 084907.

<sup>38</sup> Which is described in Section 2.4.2.

<sup>39</sup> Given the symbols, *a*, *b*, *c* for the positions and  $\phi$ ,  $\omega$ , and  $\kappa$  for the angles.

<sup>40</sup> The *a*, *b*, *c*,  $\phi$ ,  $\omega$ , and  $\kappa$  mentioned above.

<sup>41</sup> Using the *differences* object mentioned above.

<sup>42</sup> P. R. Evans. *Acta Crystallogr. D* 57.10 (2001), pp. 1355–1359, Figure 5.2 defines these angles.

<sup>43</sup> Debye, see n. 62.

<sup>44</sup> Watson *et al.*, see n. 64.

<sup>45</sup> Svergun, see n. 64.

figure of merit,  $\zeta$ , that was to be optimised by the particle swarm algorithm. This  $\zeta$  was a simple  $\chi^2$  value found as follows,

$$\zeta = \chi^2 = \sum \frac{(I_{\text{exp}}(q) - I_{\text{calc}}(q))^2}{dI_{\text{exp}}(q)}, \quad (5.1)$$

where  $I_{\text{exp}}(q)$  is the experimental scattering intensity,  $I_{\text{calc}}(q)$  is the calculated scattering intensity, and  $dI_{\text{exp}}(q)$  is the uncertainty in the experimental scattering intensity, all at a given  $q$ -vector.

Throughout this work, an inertia weight for the PSO of 0.4 was used. Generally, values of between 0 and 2 are used for the global and personal acceleration coefficients<sup>46</sup>. For the test case discussed below a value of 2 was used for both acceleration coefficients, however, in the real case discussed later, the global acceleration coefficient was reduced to 1. This was chosen to reduce the acceleration toward the global best and improve the ability for the PSO to search the parameter space available in this much larger problem.

### 5.2.2 Parallelisation

The use of a population-based optimisation method, such as the PSO, allowed for easy access to highly parallel simulation. Parallelisation was achieved by spreading the population evenly across the cores that were available to the simulation. Inter-core messaging was performed using the MPI libraries, and to ensure efficiency only the figures of merit and the best possible structure were shared across the cores. This means that during a given `fitoog` run, the only serial component was the determination of the lowest figure of merit. The efficiency of the parallelisation was defined by considering the strong and weak scaling of the software. It was possible to determine the percentage of serial,  $s$ , and parallel,  $p$ , components of the software by fitting the speedup<sup>47</sup> with Amdahl's law,<sup>48</sup>

$$\text{speedup} = \frac{1}{s + p/N}, \quad (5.2)$$

where,  $N$  is the number of cores in the parallel job, and  $s + p = 1$ . While more sophisticated methodologies could be used to further reduce the serial component, such as having a core-level best population that is only occasionally communicated with the entire swarm, this implementation was shown to be highly parallelised and useful for assessing the utility of the PSO method.

## 5.3 RESULTS & DISCUSSION

The aim of this work was to produce a well-parallelised software capable of quickly producing starting structures for later MD simulations of multiple micellar species from SAS data.

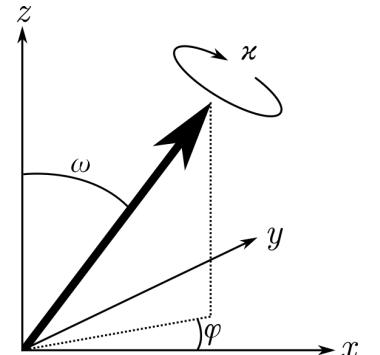


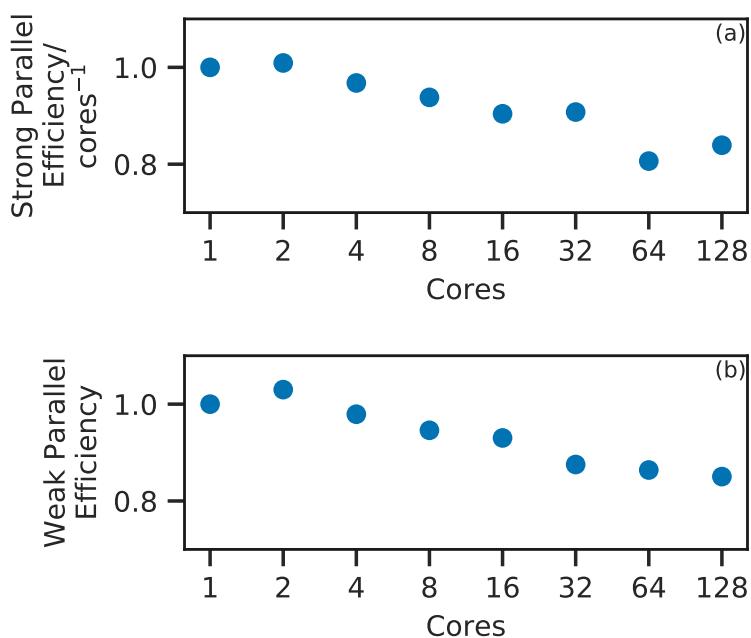
Figure 5.2: The definition of the polar angles used in the coarse grained representation of the surfactant molecule.

<sup>46</sup> These are standard values for the PSO method following P. Sun *et al.* In: *2010 2nd International Conference on Information Engineering and Computer Science*. 2010 2nd International Conference on Information Engineering and Computer Science (ICIECS). Wuhan, China, 2010, pp. 1–4.

<sup>47</sup> The time taken for a job to run on a single core divided by the time taken on multiple cores.

<sup>48</sup> G. M. Amdahl. In: *AFIPS Conference Proceedings*. AFIPS Conference. 1967.

Figure 5.3: The (a) strong and (b) weak scaling relationships of `fitoog` running on up to 128 cores of the SCARF cluster. The slight increase between 1 and 2 cores is most likely due to small difference in the speed of the different processors.



### 5.3.1 Parallelisation

<sup>49</sup> Central processing unit

The parallelisation of a software package is commonly assessed using two metrics, strong and weak scaling. These assess the CPU-bound<sup>49</sup> efficiency and memory-bound efficiency of the software respectively. A perfectly parallelised software would present a strong and weak scaling efficiency of 1 regardless of the number of processors.

In order to determine the strong scaling relationship for `fitoog`, a system was set up with a population size of 128 and was run for 5000 iterations. This system was run on a range of processor counts, from 1 to 128,<sup>50</sup> on the SCARF cluster of STFC. Figure 5.3(a) shows the strong scaling relationship for `fitoog` running on up to 128 cores. The weak scaling was probed by increasing the population size alongside the number of processors, both were varied in the same range as for the strong scaling, e.g. a population of 1 on a single core up to a population of 128 over 128 cores. The weak scaling relationship is shown in Figure 5.3(b).

It can be seen that both the strong and weak parallel efficiency of `fitoog` are relatively good, with the efficiency not dropping below 80% even when spread over 128 cores. The speedup of `fitoog` is shown in Figure 5.4,<sup>51</sup> from Amdahl's law<sup>52</sup> it is possible to find that the parallel component of a given `fitoog` run makes up 99.8% of the computation. This suggests that the parallelisation methodology currently implemented is successful, and it would not be advantageous to utilise a more sophisticated parallelisation method.

<sup>51</sup> This is generated from the strong scaling relationship.

<sup>52</sup> Amdahl, see n. 48.

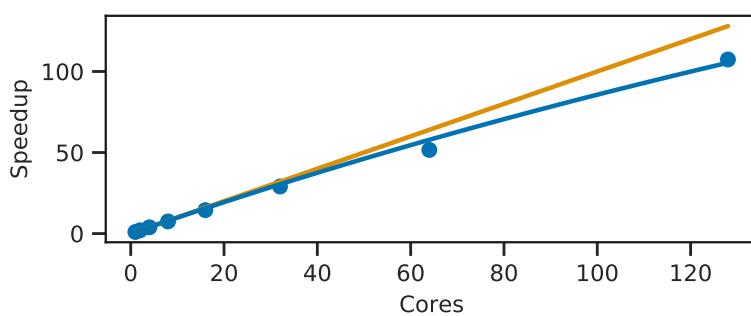


Figure 5.4: The speed up of `fitoog` running on upto 128 cores of the SCARF cluster, the blue dots show the speedup at different compute size, the orange line indicates the theoretical maximum, and the blue line shows the fitting Amdahl's law to the measured speedup.

The high efficiency in both strong and weak scaling regimes and the indication that the serial component of the computation is very small indicate that it is sensible to utilise high-performance computing resources for this software package. In the real data application of this work, the `fitoog` software was run on 48 cores of the SCARF cluster. This was chosen as it would spread the computation over exactly two nodes of the SCARF cluster,<sup>53</sup> therefore both nodes are being used to their full capacity.

<sup>53</sup> Each node contains 24 cores.

### 5.3.2 Test system

In order to assess the PSO implementation, a simple test system was defined. This consisted of a coordinate cell that contained four surfactant molecules at four corners of a 20 Å cube, each orientated in a different direction, see Figures 5.5(a-c). The scattering intensity was calculated from the cell, with the blue beads given a scattering length of 100 fm and the grey beads a scattering length of 20 fm, to ensure the presence of intense scattering. The scattering was calculated using the Debye equation<sup>54</sup> for values of  $q$  in a range from 0.3-1.5 Å<sup>-1</sup> with 100 data points, this profile is shown in Figure 5.5(d).

`fitoog` was used to fit the “experimental” data; a population size of 100 was iterated over 5000 steps. Ten repetitions of the `fitoog` run were performed,<sup>55</sup> taking around two and a half minutes per run on a workstation computer with four cores. Figure 5.6 shows the optimised scattering profile obtained from each of the runs and compares with the “experimental” data. It is clear that some of the runs agree well with the data, in particular runs 1 and 2, the resulting coordinate cell for these profiles are also shown in Figure 5.6.

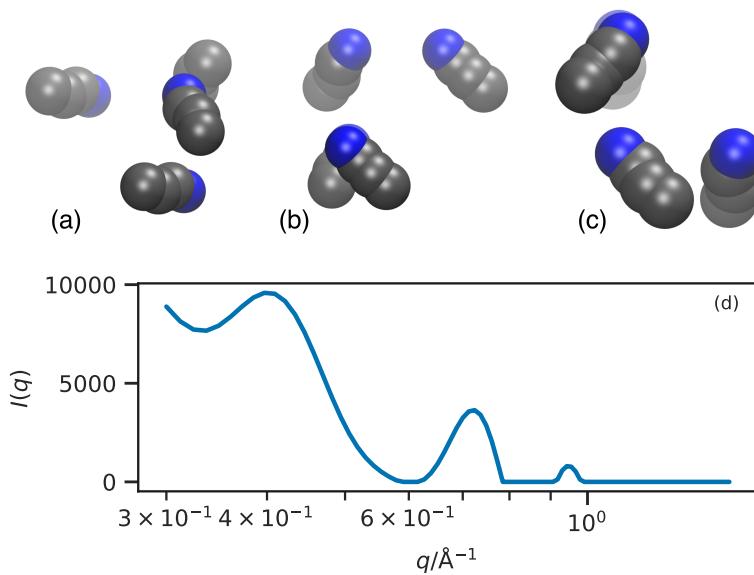
This agreement with the “experimental data” in the test case is a positive result, and the resulting cells are shown in Figure 5.6 appear to show quantitative agreement<sup>56</sup> with the coordination cell from which the “experimental” data was found. In particular, the agreement between the simulated and “experimental” data sufficient to consider that in the more disordered example it may be

<sup>54</sup> Debye, see n. 62.

<sup>55</sup> The random seed and therefore the initial starting configuration varied between each run.

<sup>56</sup> The molecules are in similar locations.

Figure 5.5: Test system coordinate cell observed down the (a) x-, (b) y, and (c) z-axis, and the calculated scattering data from the Debye equation.



possible to form a realistic starting structure. Therefore, the use of a PSO method was continued and applied to the real experimental data.

### 5.3.3 Real data

This real experimental data consisted of a single SANS profile for the hydrogenated *n*-decyltrimethylammonium micelle, with nitrate counter ions<sup>57</sup> counter ions in D<sub>2</sub>O. It was assumed that this data was completely background subtracted such that the scattering present was a result of the micelles alone. Figure 5.7 shows the scattering profile that was being modelled.

The aim of the application to real data was to attempt to quickly produce a system with multiple micellar species. Therefore, the simulation cell was substantially larger; containing 500 MARTINI coarse-grained C<sub>10</sub>TA<sup>+</sup> and NO<sub>3</sub><sup>-</sup> molecules, the available cell was a cube with a side 177 Å in length.<sup>58</sup> In order to assess the utility for the PSO, two fitoog runs were performed.<sup>59</sup> The PSO method was compared with a random method, where at each iteration a new random population was generated. Figure 5.8 shows the variation in the figure of merit,  $\zeta$ , in each of these optimisations. Figure 5.8(a-c) shows the best structure that was obtained from the fitoog runs, notably it is from the randomisation based run, which shows no evidence of the formation of micelle-like species. From this, there is no clear benefit to the use of the PSO method over simply selecting random structures.

This inability for the PSO method to optimise the structure of the micellar species could be due to a wide variety of reasons, some which could be acted on and others that could not. While it

<sup>57</sup> Abbreviated to C<sub>10</sub>TA<sup>+</sup> and NO<sub>3</sub><sup>-</sup> respectively.

<sup>58</sup> No solvent was included in the box as the scattering was considered to have arisen from the micelle scattering alone. The concentration of the solution was  $\sim 0.15 \text{ moldm}^{-3}$ , which is nearly three times the cmc for C<sub>10</sub>TA<sup>+</sup> from J. L. Rodríguez *et al.* *J. Colloid Interface Sci.* 314.2 (2007), pp. 699–706.

<sup>59</sup> Each with five repetitions, and a population size of 96 over 500 iterations.

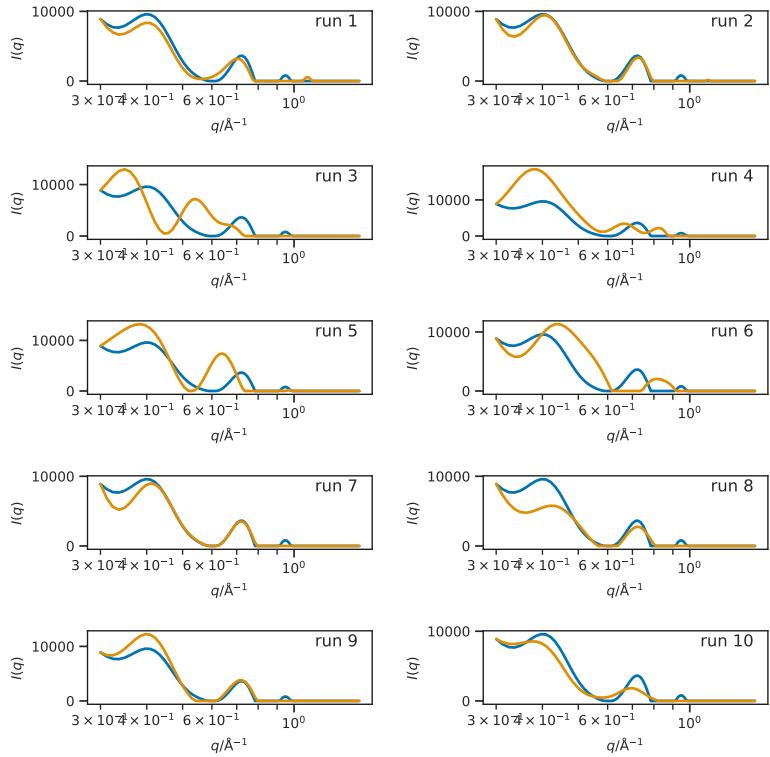


Figure 5.6: The best fit from the `fitoog` run (orange line) is compared with the “real experimental” data (blue line) for each of the ten runs. The result of runs 1 (a, b, and c) and 2 (d, e, and f) along each axis for the test system coordinated cell.

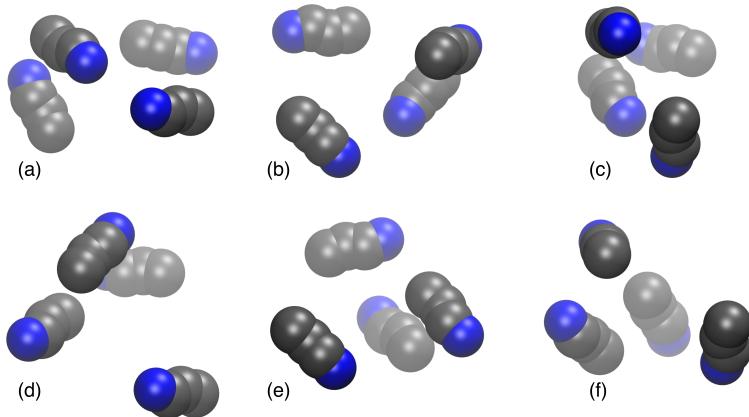


Figure 5.7: The experimental SANS data, from a solution of hydrogenated  $\text{C}_{10}\text{TANO}_3$  in  $\text{D}_2\text{O}$  at a concentration of  $\sim 0.15 \text{ mol dm}^{-3}$ , to which the real fitoog run was attempting to fit.

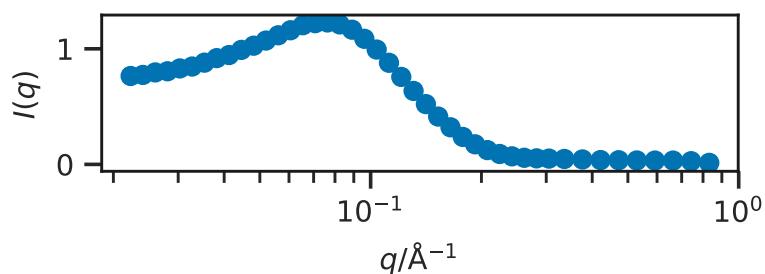
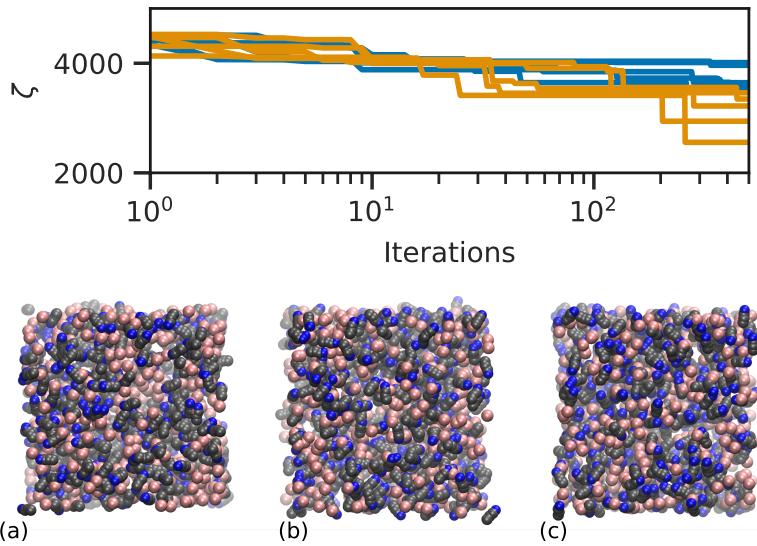


Figure 5.8: The quality of agreement between the fitoog model and the experimental data; where the blue lines are the different PSO runs and the orange are the randomisation runs. Result of the best outcome from the real data example, observed down the (a) x-, (b) y, and (c) z-axis.



is common to use values for the acceleration coefficients that are typically in the range of 0-2, it may be necessary to optimise these values.<sup>60</sup> However, it may be the case that the parameter space was too large to be optimised using the PSO alone, this is very likely considering that the dimensionality of the parameter space was 4500.<sup>61</sup>

A possible method that may enable the optimisation of such structures would be the inclusion of an energetic term. For example, this could involve the use of an energy optimisation to be performed alongside the structural optimisation to the scattering profile, regardless of the optimisation methodology. The use of energetic considerations and a Markov state model optimisation has been shown to perform well for peptide self-assembly.<sup>62</sup> Additionally, the Empirical Potential Structure Refinement used by Hargreaves *et al.*<sup>63</sup> and in a coarse-grained fashion by Soper and Edler<sup>64</sup> and the method used by Ivanović *et al.*<sup>65</sup> all involve performing an energetic optimisation that is biased on the agreement with experimental data. This use of an energetic consideration could be included in fitoog by performing an energy minimisation

<sup>60</sup> Which, of course, could lead to an infinite set of optimisations.

<sup>61</sup>  $500 \times 6 + 500 \times 3$ , for 500  $\text{C}_{10}\text{TA}^+$  and  $\text{NO}_3^-$  molecules.

<sup>62</sup> U. Sengupta *et al.* *J. Chem. Phys.* 150.11 (2019), p. 115101.

<sup>63</sup> Hargreaves *et al.*, see n. 35.

<sup>64</sup> A. K. Soper *et al.* *BBA - General Subjects* 1861.6 (2017), pp. 1652–1660.

<sup>65</sup> Ivanović *et al.*, see n. 36.

step following each step of the PSO. This would have the added benefit of creating a range of surfactant conformations allowing a more realistic structure to form.

#### 5.4 CONCLUSIONS

This chapter presented the development of a software entitled *fitoog*, the aim of which was to try and use PSO in order to generate a reasonable starting structure for an MD simulation of multiple micelles in solution. Previously work had shown it was possible to build a single micelle that agreed well with dilute experimental data. However, there had been no previous work investigating the generation of multiple micelles based on the scattering data alone.

*fitoog* is a highly parallelised software, capable of running efficiently on high-performance computing resources. It was determined that the serial component of running *fitoog* made up 0.2 % of the overall calculation. From ten repetitions, the software was able to resolve the expected structure for a model test case quickly running on a workstation class machine. However, when applied to a larger, realistic system it was not possible to obtain any realistic structures. Additionally, when compared with a simple random number generation, it performed no better in minimisation of the figure of merit. This indicates that it may be necessary to consider energetic information in addition to structural detail to accurately develop a feasible starting structure for an MD simulation. Alternatively, it may be possible to use a short MD simulation to generate the starting structure for the *fitoog* run rather than a randomly generated structure.



# 6

## DEVELOPING OPEN-SOURCE TEACHING MATERIALS FOR CLASSICAL SIMULATION AND SCATTERING

### ABSTRACT

Classical molecular dynamics<sup>1</sup> simulations are a common component of multi-modal analyses from scattering measurements, such as small-angle scattering and reflectometry. Users of these experimental techniques often have no formal training in the theory and practice of MD simulation, leading to the possibility of these simulations being treated as a “black box” analysis technique. In this chapter, two open educational resources<sup>2</sup>; `py1j` and “The Interaction Between Simulation and Scattering” are described and their utility for introducing users of scattering methods to the tools of MD discussed. These resources are a Python library designed to allow users to interact with MD simulations and a series of interactive web pages, to introduce classical simulation and how it may be applied to scattering.

<sup>1</sup> MD.

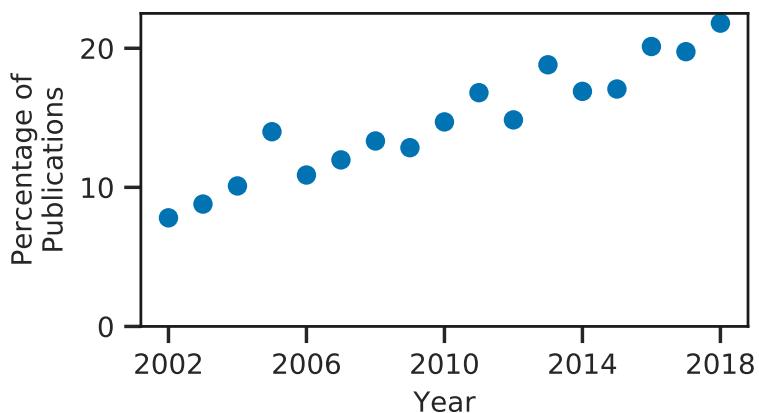
<sup>2</sup> OERs.

### CONTEXT

Often users of scattering methods are interested in the use of classical simulation to better understand the results that they have obtained. However, regularly those with an experimental background aim to apply molecular simulation methods without true consideration of the potential pitfalls of using such methods. These OERs were developed to introduce experimental colleagues<sup>3</sup> to classical simulation and MD. They help to facilitate learning by providing a simple, but mathematically rigorous interface and enabling the “worked example effect” throughout.

<sup>3</sup> Additionally, in the case of `py1j` undergraduate students.

Figure 6.1: The annual percentage of publications that mention “small angle scattering” that also mention “molecular dynamics”, determined from the numbers of matching Google Scholar results.



## 6.1 INTRODUCTION

The popularity of classical simulation, both all-atom and coarse-grained, as a technique for multi-modal analysis of scattering techniques, such as reflectometry and small angle scattering,<sup>4</sup> has grown linearly over the past two decades.<sup>5</sup> Figure 6.1 shows that as of 2019, ~20 % of all SAS publications also mention MD. Users of scattering techniques often have a background in experimental science and may have received little formal training in the theory or practice of computational modelling. This can lead to the use of MD simulations as a “black box” without necessarily understanding the underlying methodologies, or considering possible sources of error. To help support researchers use MD simulations in their analysis of scattering data while reducing the risk of modelling errors, a number of software tools, such as WAXSiS<sup>6</sup> and SASSIE<sup>7</sup> have been developed that present easy-to-use, graphical, web-based user interfaces.

A complementary approach is to organise educational activities, such as lectures and workshops, tailored to introduce molecular simulation techniques to audiences of scattering users. One example is the annual ISIS Neutron Training Course, which includes a module titled “An Introduction to Molecular Dynamics for Neutron Scattering”. This module covers the fundamentals of classical MD simulation, presents applications of these methods in neutron science, and gives students practical hands-on experience with the SASSIE software package.<sup>8</sup>

While lectures and workshops are an effective tool for education and training, participation can be limited due to difficulties attending in person<sup>9</sup> or physical limits on student numbers. An alternative educational strategy gaining popularity within scientific and engineering communities is the publication of OERs. These are courses, lectures, or learning resources published online that are freely available for use by anyone. In addition to their broad

<sup>4</sup> SAS.

<sup>5</sup> Pan *et al.*, see n. 58; L. Boldon *et al.* *Nano Rev.* 6.1 (2015), p. 25661; Hub, see n. 68; Koutsoubas, see n. 41; Darré *et al.*, see n. 9; Scoppola *et al.*, see n. 31.

<sup>6</sup> Chen *et al.*, see n. 10; C. J. Knight *et al.* *Nucleic Acids Res.* 43.W1 (2015), W225–W230.

<sup>7</sup> Perkins *et al.*, see n. 2.

<sup>8</sup> Perkins *et al.*, see n. 2.

<sup>9</sup> Due to location and cost.

accessibility, these resources have permissive “open” copyright licenses that allow their use in the “5R activities”: retain, reuse, revise, remix, and redistribute.<sup>10</sup> Publishing a resource as an OER increases the reach and impact, as others may use it in their own teaching not only in its original form, but are free to modify, and redistribute, the material to better suit their aims. The OERs developed as a part of this work both heavily leverage the Jupyter Notebook framework<sup>11</sup> to enable technology-enhanced OERs.

### 6.1.1 Using Jupyter Notebooks in Education

Project Jupyter<sup>12</sup> is a collection of standards, a community, and a set of software tools. The Jupyter Notebook is one of these software tools that is capable of creating, editing, and running a Jupyter Notebook file. This is a file that can contain executable code<sup>13</sup> and narrative text,<sup>14</sup> enabling the user to “tell an interactive, computational story”.<sup>15</sup> Furthermore, the interactive nature reduces the barrier of entry to computational methods that is often imposed on those learning, due often to the need to understand a command line interface.

The Jupyter Notebook framework has become a popular platform for OERs that teach computational skills, because it allows authors to include instructional text, images and other media, alongside the executable, editable code, in an example of “literate programming”.<sup>16</sup> This format encourages students to directly interact with code examples by running, editing, and rerunning these within the source document,<sup>17</sup> supporting exploratory experiential learning,<sup>18</sup> and enabling the “worked example effect”.<sup>19</sup> Furthermore, the modular nature of a Jupyter Notebook OER allows the resource designer to build computational tools to be used by those learning, such as Python libraries to aid understanding.

### 6.1.2 Teaching computational simulation

It was suggested by Aiello-Nicosia and Sperandeo-Mineo<sup>20</sup> that understanding the microscopic disordered motions of particles in gases is a difficult problem for many science students. Pallant and Tinker<sup>21</sup> commented that there is an educational challenge associated with helping students to rationalise the relationship between the mathematics underlying computational simulation and the behaviour of the system. Additionally, it has been noted that the visualisations often used in the traditional teaching of MD simulations may cause difficulties for students’ understanding, these can be categorised as follows:<sup>22</sup>

- visual subtlety: often simulations are presented as two-dimensional displays of three-dimensional objects, creating spatial relationships that may be difficult to interpret,
- complexity: high information depth in an image, perhaps of a

<sup>10</sup> D. Wiley. *Open Content - Definition*. URL: <http://opencontent.org/definition> (Accessed 2018-11-19).

<sup>11</sup> T. Kluyver et al. *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (2016), pp. 87–90.

<sup>12</sup> Kluyver et al., see n. 11.

<sup>13</sup> In this work this is exclusively in the Python programming language, however, the Notebook software is compatible with many languages.

<sup>14</sup> Either Markdown or formatted LATEX.

<sup>15</sup> L. A. Barba et al. *Teaching and Learning with Jupyter*. 2019. URL: <https://jupyter4edu.github.io/jupyter-edu-book/> (Accessed 2019-3-4).

<sup>16</sup> D. E. Knuth. *Comput. J.* 27.2 (1984), pp. 97–111.

<sup>17</sup> L. A. Barba et al. *CyberTraining: DSE-The Code Maker: Computational Thinking for Engineers with Interactive Contextual Learning*. 2017. URL: <https://doi.org/10.6084/m9.figshare.5662051.v1>.

<sup>18</sup> S. Papert. *Mindstorms*. 1993.

<sup>19</sup> R. A. Tarmizi et al. *J. Educ. Psycho.* 80.4 (1988), pp. 424–436.

<sup>20</sup> M. L. Aiello-Nicosia et al. *Eur. J. Phys.* 6.3 (1985), pp. 148–153.

<sup>21</sup> Amy Pallant et al. *J. Sci. Educ. Technol.* 13.1 (2004), pp. 51–66.

<sup>22</sup> L. L. Jones et al. *Chem. Educ. Res. Pract.* 6.3 (2005), pp. 136–149.

complex chemical model, will lead to increases, often unnecessarily in cognitive load,

- abstractness and conceptual depth: conventions are often used to represent phenomena that may be vague or unfamiliar to those learning about the methods.

Therefore, in addition to making use of software that will enable learners to interact with the computational methods being introduced, it is important that the discussion and visualisation of these methods are as straightforward as possible.

Currently, there are a handful of software packages that are designed to introduce, or educate about classical simulation, these include the Democritus flash application and the ArgonMD app developed by members of the Theory and Modeling in Chemical Science Centre for Doctoral Training.<sup>23</sup> Both of these tools provide an interface to interact with a two-dimensional MD simulation of a Lennard-Jonesium system.<sup>24</sup> Democritus was originally developed by Prof. Bill Smith from STFC,<sup>25</sup> however, development of this software has stagnated with no updates being made available since 2001. ArgonMD is a modern application built to work on mobile phone interfaces, in addition to computers. It offers some interesting features, such as the ability to create attractive pseudo-particles and to arbitrarily define a potential model function with touch-gestures. The ArgonMD app offers an exciting tool for introducing computational simulation, in particular, to those not familiar with computational interfaces or programming. However, the educational utility of this software is limited by the closed-source nature of the development<sup>26</sup> and that the MD algorithm being used is abstracted substantially from the user in an effort to create an easy to use interface.

In this chapter, I will introduce the Python-based software designed to introduce classical simulation to students and show its utility in introducing difficult problems such as that of microscopic disorder particle motion. Additionally, this software aims to educate the users about the underlying methods by making the code available and easily accessible. Then I will present how this software has been used in an OER aimed to introduce users of scattering techniques, such as those discussed previously in this thesis, to classical simulation. This OER aims to use the worked example effect to engage students in literate programming, this will reduce the barrier of entry to the use of classical simulation to aid in the analysis of experimental data.

## 6.2 PYLJ: AN OPEN-SOURCE TEACHING TOOL FOR CLASSICAL ATOMISTIC SIMULATION

<sup>23</sup> McCluskey *et al.*, “Pylj: A Teaching Tool for Classical Atomistic Simulation”, see n. ??; A. R. McCluskey *et al.* *Arm61/Pylj: Pylj-1.2.5*. 2019. URL: <http://doi.org/10.5281/zenodo.2587898>, PYthon Lennard-Jones.

<sup>24</sup> pylj<sup>27</sup> is an educational software and MD engine designed to introduce students to the details of classical simulation. Initially,

```

from pylj import md, sample

def md_simulation(
    number_of_particles, temperature, box_length, number_of_steps
):
    system = md.initialise(
        number_of_particles, temperature, box_length, "square"
    )
    sample_system = sample.Energy(system)
    system.time = 0
    for i in range(0, number_of_steps):
        system.integrate(md.velocity_verlet)
        system.md_sample()
        system.heat_bath(temperature)
        system.time += system.timestep_length
        system.step += 1
        if system.step % 10 == 0:
            sample_system.update(system)
    return system

```

`pylj` was only able to utilise a Lennard-Jones potential model,<sup>28</sup> however, recently<sup>29</sup> there is the ability to include any custom potential model, with a Buckingham potential packaged with the software.<sup>30</sup> In an effort to reduce the complexity and visual subtlety of typical MD simulations, `pylj` performs two-dimension MD simulation. From the beginning, this software was designed to operate in the Jupyter Notebook framework and therefore eliminate the need for the learner to interact with the command line interface, as is the case with common research-level MD packages like Gromacs, LAMMPS, or DL\_POLY.<sup>31</sup>

### 6.2.1 Software design

`pylj` was designed such that it may be operated at a series of different levels of abstraction. For example, an educator could write a simple function to allow the running of an MD simulation<sup>32</sup> or an interested student could manually interact with the source code. This abstraction is achieved through the modular design of `pylj`, where the `md.py` module implements all of the functionality related to MD simulation. The simulation is controlled by an overarching `System` class, which contains all of the information regarding the simulation that has been/is running.

The computationally intensive nature of the pairwise force and energy calculations necessary for MD simulation necessitated the use of Cython, a method of including compiled C language code within a Python package.<sup>33</sup> This enabled a ~10 times speed up in the determination of the pairwise interactions when compared with the use of pure Python implementation. The pure Python implementation is still present in the package, to allow students that are familiar with Python to understand the mechanism without the requirement to become familiar with C. From version 1.2.1,<sup>34</sup> the pure Python versions of the pairwise interactions have been built using numba just-in-time compilation,<sup>35</sup> which enabled a 5 times speed up. This means that a `pylj` simulation without the compiled C functions is now just twice as slow than the compiled version. Furthermore, this speed is now comparable to the length of time taken to render the visualisation, meaning that the pairwise

Code Block 6.1: An example of an NVT (constant number of particles, volume, and temperature) ensemble MD algorithm as implemented in `pylj`. The input variables are `number_of_particles` which is the number of particles in the simulation, `temperature` which is the temperature of the simulation in Kelvin, `box_length` which is the size of the simulation cell edge in Ångstrom, and `number_of_steps` which is the number of MD iterations to be performed. This will return a `pylj.System` class object containing a full description of the simulation.

<sup>28</sup> Lennard-Jones, see n. 73.

<sup>29</sup> McCluskey *et al.*, *Arm61/PyLj: PyLj-1.1.0*, see n. ??.

<sup>30</sup> Buckingham, see n. 75.

<sup>31</sup> Berendsen *et al.*, see n. 46; S. Plimpton. *J. Comput. Phys.* 117 (1995), pp. 1–19; W. Smith *et al.* *Mol. Simulat.* 28.5 (2002), pp. 385–471.

<sup>32</sup> An example of this can be seen in Code Block 6.1.

<sup>33</sup> Cython 3.0a0. URL: <https://cython.readthedocs.io/en/latest/> (Accessed 2019-3-4).

<sup>34</sup> A. R. McCluskey *et al.* *Arm61/PyLj: PyLj-1.2.1*. 2019. URL: <http://doi.org/10.5281/zenodo.2423866>.

<sup>35</sup> Numba: A High Performance Python Compiler. URL: <http://numba.pydata.org> (Accessed 2019-3-4).

Code Block 6.2: The Lennard-Jones potential model as implemented in `pylj`. The input variables are `dr` which is an array of floats describing the distances between the pairs of particles, `constants` which is an array of two floats giving the  $A$  and  $B$  parameters for the Lennard-Jones function, and the Boolean `force` which if true return the negative of the first derivative of the energy. This returns either the potential energy or the force depending on the `force` Boolean.

```
import numpy as np

def lennard_jones(dr, constants, force=False):
    if force:
        return 12 * constants[0] * np.power(dr, -13) - (
            6 * constants[1] * np.power(dr, -7)
        )
    else:
        return constants[0] * np.power(dr, -12) - (
            constants[1] * np.power(dr, -6)
        )
```

interactions are no longer the rate determining step in the use of `pylj`.

The `sample.py` module is integral to the utility of `pylj`, as this allows educators to easily create custom visualisation environments. Using this module, is it possible to create an MD simulation that enables the plotting of a huge variety of outputs. Plots such as instantaneous pressure against time, radial distribution function,<sup>36</sup> and instantaneous temperature histogram are included in `pylj` and it is easy for the users to create custom plots or build visualisations where multiple plots can be presented together.

Currently, the `forcefield.py`<sup>37</sup> module is the location where the potential model may be defined. This involves the definition of a single function that describes the potential model,<sup>38</sup> with a Boolean flag that defines where the force or energy should be returned. In future, for a planned `pylj` 2.0, the potential model definition will be adapted such that each model is an individual class containing functions for each of the different parameters to be calculated. This will more easily allow the growth of `pylj` to enable features such as particles that have different potential model parameterisations and the inclusion of mixing rules. This will also allow for the development of more complex potential models, such as those that allow for long-range electrostatic interaction that are not currently possible.

In order to give an idea of the capabilities and current use cases of the `pylj` software, three typical applications are discussed, a further application is evident in Section 6.3.

### 6.2.2 Applications: States of matter

<sup>36</sup> Abbreviated to RDF.

States of matter<sup>39</sup> is a common high school level science subject, where the different states of matter are rationalised in terms of the atomic density and interactions. Often this is introduced with pictorial examples showing a two-dimensional representation of a hexagonally close-packed crystal, a disordered liquid, and a low-density gas with the atoms represented as circular particles. `pylj` is capable of easily reproducing these diagram, as shown in Figure 6.2, while increasing student engagement by representing a “real” chemical system in thermal motion. `pylj` was recently used by Dr Benjamin Morgan of the University of Bath for such a demonstration in a seminar introducing chemical simulation to a

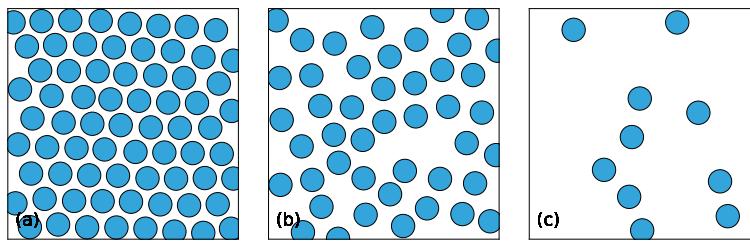


Figure 6.2: A snapshot of a `pylj` MD simulation for: (a) a solid, (b) a liquid, and (c) a gas.

cohort of mathematicians.

### 6.2.3 Applications: Ideal gas law

When `pylj` was originally published,<sup>40</sup> the repository included an example of a possible laboratory exercise where the ideal gas law was modelled using MD simulation. This was achieved by varying the particle density and measuring the time-averaged pressure of the simulation. At low particle densities, where the interactions of the particles are unlikely, the `pylj` MD simulation agrees well with the ideal gas law. However, as the particle density increases such that the inter-particle interactions are more frequent, deviations are observed in agreement with the van der Waals equation, as can be seen from Figure 6.3. Using this exercise, it is possible to introduce a cohort of students to the insight available from chemical simulation, without a significant focus on the simulation methods increasing the accessibility to students in the first or second year of their undergraduate course.

<sup>40</sup> McCluskey *et al.*, “PyLj: A Teaching Tool for Classical Atomistic Simulation”, see n. ??.

### 6.2.4 Applications: Molecular dynamics

The final application of `pylj` is its use in teaching MD, with Code Block 6.1 showing a MD algorithm. The framework of `pylj` means that it is straightforward, and does not necessarily require substantial familiarity with the Python programming language. The MD algorithm presented above is simple and clear to implement, allowing the focus of the laboratory exercise to be on the students' understanding of the methodology. `pylj` has been applied in this

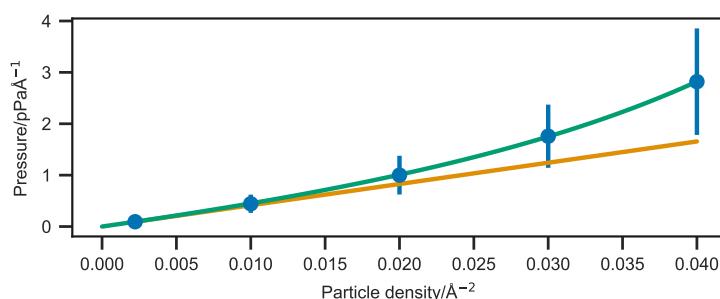


Figure 6.3: The deviation from the ideal gas law observed using 0.1 ns `pylj` simulations (blue circle), the ideal gas law is shown with a solid orange line, which the van der Waals equation of state is shown with a solid green line.

way in a third-year undergraduate laboratory exercise at the University of Bath which introduced students to MD simulations.

### 6.2.5 Comparison to other packages

The open-source nature of the `pylj` package allows the 5R activities to take place, meaning that anyone can use part or all of `pylj` to improve their own teaching. This is not available from other software packages that provide a visualisation environment for the simulation. In addition to allowing for the 5R activities, the fact that the source code is available allows users to read the underlying code and therefore better understand the computational methodology taking place.

The fact that `pylj` has been built on the popular Jupyter Notebooks framework means that there is a low barrier of entry to using the software.<sup>41</sup> This is currently the case for the use of `pylj` in the “The interaction between simulation and scattering” OER discussed below.

## 6.3 THE INTERACTION BETWEEN SIMULATION AND SCATTERING

The Jupyter Notebook framework and `pylj` software were then used to enable learning and understanding in the OER entitled “The interaction between simulation and scattering”.<sup>42</sup> This is an online, open-source, interactive learning resource written to introduce members of the scattering and diffraction community to MD simulations. The aim is to improve their understanding, and therefore reduce the treatment of MD as “black box” calculation by experimental colleagues. The OER comprises five lessons that introduce classical MD methods and show how these can be used to assist in the analysis of experimental scattering data by the calculation of a simulated scattering profile from the MD simulation. `pylj` is used to provide simple, but computationally authentic, examples of simulations, that demonstrate visually and programmatically the conceptual relationships between simulation and scattering techniques. Finally, a “real-world” example of calculation of the scattering profile from a simulation of a lysozyme protein in solution is shown.

### 6.3.1 Resource construction

The resource is available in two main formats. First, as a series of web pages, hosted at [pythoninchemistry.org/sim\\_and\\_scat](http://pythoninchemistry.org/sim_and_scat). Secondly, as the source-code repository used to build these web-pages.<sup>43</sup> The source content consists of a set of Jupyter Notebooks and Markdown files, which are automatically compiled using the `jupyter-book` tool<sup>44</sup> to generate the web version. This allows the resulting webpages to include text, equations, and figures, which describe key concepts and explain details of algorithms, as well

<sup>41</sup> For example, it is not necessary to install anything as it is possible to run a centrally supported Jupyter Notebook hub that provides access to `pylj` online.

<sup>42</sup> McCluskey *et al.*, “An Introduction to Classical Molecular Dynamics Simulation for Experimental Scattering Users”, see n. ??; A. R. McCluskey *et al.* *Pythoninchemistry/Sim\_and\_scat: Sim\_and\_scat-v0.3-Preprint*. 2019. URL: <http://doi.org/10.5281/zenodo.2556824>, available at [pythoninchemistry.org/sim\\_and\\_scat](http://pythoninchemistry.org/sim_and_scat).

<sup>43</sup> McCluskey *et al.*, *Pythoninchemistry/Sim\_and\_scat: Sim\_and\_scat-v0.3-Preprint*, see n. 42.

<sup>44</sup> S. Lau *et al.* *Jupyter/Jupyter-Book Beta-2*. 2019. URL: <https://github.com/jupyter/jupyter-book>.

as Python code blocks, which provide specific examples. The web pages have Thebelab and BinderHub integrations,<sup>45</sup> which allow students to launch interactive versions of these webpages that allow execution and modification of the included Python code. The ability to read the resource as an “interactive document” improves the ability for the students to engage in the “worked example effect”.<sup>46</sup> The resource is provided under a CC-BY license,<sup>47</sup> while the *jupyter-book* software is shared under an MIT license,<sup>48</sup> both of which are open and highly permissive. This allows readers to reuse or remix the material to enhance their own educational platform and for secondary authors to contribute to improving the source material.

### 6.3.2 Resource outline

The resource follows a simple outline that introduces key aspects of MD simulations.

**Home:** The welcome page introduces the resource, explains the purpose, and gives the user information about how the resource may be used, including details of the Thebelab and BinderHub integrations. This page also contained details of the permitted use/reuse, sharing of the content of the resource and licensing. Finally, a list of authors and contributors was given.

**Classical methods:** This section introduces concepts related to classical simulation methods. Including the use of interatomic potential functions, alongside some examples, such as the Lennard-Jones and Buckingham potential models.<sup>49</sup> The problem of parameterising a potential model is then suggested, showing the use of higher accuracy quantum mechanical calculations to do so. The presence of off-the-shelf, general potential models are discussed; with the caveat that they may still require system-specific optimisation. Finally, mixing rules are mentioned; again discussing the possible problems that a user may encounter if applying these blindly to specific systems.

**Molecular dynamics:** With the concept of a classical interatomic potential introduced, the resource then begins to focus on how these are used in MD simulations. It is shown how a one dimensional NVE (constant number, volume, and energy) MD simulation may be built, using the Velocity-Verlet algorithm and the Lennard-Jones potential model.<sup>50</sup> The Velocity-Verlet algorithm is introduced in terms of Newton’s laws of motion and the generalised equations of motion. Finally, a range of key factors that can affect MD simulations are discussed; simulation ensembles, the distance cut-off for an interatomic potential, and the use of periodic boundary conditions.

**py1j and interaction with scattering:** The later aspect of the resource covers using MD simulations to understand scattering profiles. This is presented as a practical example, using the *py1j*

<sup>45</sup> B. Ragan-Kelley *et al.* *Minrk/Thebelab v0.3.3*. 2019. URL: <https://github.com/minrk/thebelab>; B. Ragan-Kelley *et al.* *Jupyterhub/Binderhub FirstLight*. 2019. URL: <https://github.com/jupyterhub/binderhub>; Jupyter *et al.* In: *Proceedings of the 17th Python in Science Conference*. SciPy2018. 2018.

<sup>46</sup> Tarmizi *et al.*, see n. 19.

<sup>47</sup> Creative Commons. *Creative Commons – Attribution 4.0 International – CC-BY-4.0*. 2019. URL: <https://creativecommons.org/licenses/by/4.0/>.

<sup>48</sup> Open Source. *The MIT License*. 2019. URL: <https://opensource.org/licenses/MIT>.

<sup>49</sup> Lennard-Jones, see n. 73; Buckingham, see n. 75.

<sup>50</sup> Swope *et al.*, see n. 116; Lennard-Jones, see n. 73.

<sup>51</sup> McCluskey *et al.*, “Pylj: A Teaching Tool for Classical Atomistic Simulation”, see n. ??; McCluskey *et al.*, *Arm61/Pylj: Pylj-1.2.5*, see n. 27.

<sup>52</sup> Debye, see n. 62.

<sup>53</sup> Watson *et al.*, see n. 64; Svergun, see n. 64.

<sup>54</sup> Berendsen *et al.*, see n. 46.

<sup>55</sup> J. A. Lemkul. *GROMACS Tutorial: Lysozyme in Water*. URL: <http://www.mdtutorials.com/gmx/lysozyme/index.html> (Accessed 2019-3-8).

<sup>56</sup> Michaud-Agrawal *et al.*, see n. 64; Gowers *et al.*, see n. 64.

<sup>57</sup> D. Franke *et al.* *Nat. Methods* 12.5 (2015), pp. 419–422.

<sup>58</sup> Perkins *et al.*, see n. 2; D. Svergun *et al.* *J. Appl. Crystallogr.* 28.6 (1995), pp. 768–773.

package.<sup>51</sup> A two-dimensional MD simulation of argon atoms interacting through a Lennard-Jones potential is demonstrated. The users are first shown this working `pylj` simulation and invited to interact with the simulation and the custom plotting functionality of `pylj`. The concept of an RDF is then shown, and the user is given the opportunity to run some `pylj` simulations with the RDF being output alongside the simulation window. Next, the Debye equation<sup>52</sup> is presented and it is shown how it may be used to calculate scattering data from a simulation. The user is invited to observe the effect of simulation temperature on the resulting scattering profile. There is also a short discussion of alternative, faster, algorithms for calculating scattering profiles, such as the Fibonacci Sequence or Golden Vectors method.<sup>53</sup>

**“Real” simulation and scattering:** Having shown the development of a scattering profile from an idealised system, the user is directed to a popular resource for the GROMACS<sup>54</sup> MD software. This resource gives a quick introduction to using GROMACS to simulate a lysozyme molecule in buffer.<sup>55</sup> The user may then use their own simulated trajectory or one that can be downloaded from the OER. It is shown how the system may be visualised, introducing the MDAnalysis Python package for the analysis of MD trajectories,<sup>56</sup> and show how the scattering profile developed from the lysozyme simulation compared with experimental data.<sup>57</sup> The module finishes by pointing the student to resources to more easily resolve scattering data from the molecular simulation, such as SASSIE and CRYSTOL.<sup>58</sup> The focus of this resource is to introduce simulation methodologies to users of scattering to aid their understanding, not to derive the exact mechanics of the calculation of scattering from a simulation. Resources for this purpose already exist and have well-developed tutorials, so it is not necessary to recreate such software here.

## 6.4 CONCLUSIONS

In the chapter, two OERs focussed on classical simulation and MD were shown. The first of these resources was the `pylj` Python package, which designed for use at any education level to give an easy, visual example of classical simulation. This software is open-source and actively developed, with the growth of capability and applications in the future. Currently, the code is used in the third year computational chemistry laboratory at the University of Bath and there is an ongoing discussion for it to be used in future in the second year computational chemistry laboratory at the University of Bristol. Additionally, the webpage for `pylj` at [pythoninchemistry.org/pylj](http://pythoninchemistry.org/pylj) has been viewed over 400 times since launching in June 2018, indicating the popularity of the software.

The second OER is the online, interactive learning module for the introduction of users of experimental scattering methods to

classical simulation. This module is shared under an open, permissive license and in future, its use/reuse may beuptaken by educators of scattering worldwide. Furthermore, there is scope to introduce the module as a flipped learning component<sup>59</sup> within scattering courses at the ISIS Neutron and Muon Source and Diamond Light Source.

While the `pylj` software improves the availability of resources to introduce MD simulation, this interactive learning module represents a unique resource to enable learning and understanding. Previously, MD simulation has been taught in a chalk-and-talk fashion, or with examples of working simulations, however, this resource enables students to learn about the simulation methodology by interacting directly with working code. This allows for the worked example effect to be invoked during learning, improving the educational power of the materials.

<sup>59</sup> *Flipped Learning | Higher Education Academy*. URL: <https://www.heacademy.ac.uk/knowledge-hub/flipped-learning-0> (Accessed 2019-3-12).



# 7

## SUMMARY & FUTURE WORK

The work described in this thesis aimed to investigate the use of different coarse-graining methodologies to improve and aid in the analysis of scattering data from scattering experiments, in particular, reflectometry and small angle scattering.<sup>1</sup> The different coarse-graining methods varied both in what was being coarse-grained, from the potential model to descriptions for entire surfactants, and made use of a series of optimisation and sampling techniques to improve the inference from these analyses. Some of these methods showed greater success, for example, the use of classical simulation-driven analysis for reflectometry and the chemically-consistent monolayer models. Others were less successful, however, this work represents a significant step forward in the development of analysis methodologies for scattering experiments. Finally, teaching materials for introducing classical simulation approaches to users of scattering were developed that provide a new platform for engagement and understanding in simulation-driven analysis.

<sup>1</sup> SAS.

### 7.1 CHEMICALLY-CONSISTENT MODELLING OF X-RAY AND NEUTRON REFLECTOMETRY

The use of coarse-graining is commonplace in the analysis of reflectometry measurements, as the models that are typically used are comprised of layers which represent different chemical components of the underlying structure. In the work contained in Chapter 3, this approach was extended to use a chemically-consistent model for the analysis of a set of phospholipid reflectometry measurements at an air-deep eutectic solvent interface. This model was coarse-grained such that the system was described as consisting of two layers consisting of phospholipid heads and tails. The use of this coarse-grained method allowed for X-ray reflectometry<sup>2</sup> measurements conducted at different surface pressures<sup>3</sup> to be co-refined, by keeping chemical features constant<sup>4</sup> across the different SPs for a given phospholipid. This allowed for the constraints, that are typically applied in the modelling of phospholipid monolayers at an air-water interface, on the head and tail volume

<sup>2</sup> XRR.

<sup>3</sup> SPs.

<sup>4</sup> Such as the head and tail volumes.

to be removed to reflect of the effect of the non-aqueous solvent and SP on these. This method allowed for a unique insight into the structure of the phospholipid monolayer at the air-DES interface, showing a strong similarity to those formed at the air-water interface. However, it was possible to observe that the head group volume for the PG-containing phospholipid appeared to swell as a result of interactions with the ionic solvent.

This work was published alongside a fully-reproducible electronic supplementary information,<sup>5</sup> which gave access to the chemically-consistent model Python class. This will allow others to use this model in their data analysis, additionally there is scope to include this model<sup>6</sup> in an accessible repository for models that may be used with the `refnx` package.<sup>7</sup> The future of XRR and neutron reflectometry<sup>8</sup> analysis will build on the sharing of these models enabling science to be performed by science-domain experts, who have little reflectometry analysis experience. This is already the case in SAS where a large library of functional models exist and users can pick those which fit their needs.<sup>9</sup>

Another novel innovation of this work was the use of Markov chain Monte Carlo sampling to probe the inverse uncertainties of a given model, in addition to the inter-parameter correlations. However, it was noted that the use of MCMC can only probe the parameter space available within the given experimental uncertainties, which in particular for XRR measurements are believed to be significantly underestimated.<sup>10</sup> Therefore, in order to fully leverage the power of this sampling process for inverse uncertainty determination, it is first necessary to determine accurate uncertainties for the experimental measurements, which is a non-trivial task. As there is a growing interest in advanced modelling approaches, such as Bayesian inference and machine learning, a concerted effort from large scale facilities and instrument manufacturers to accurately define the uncertainties of a given measurement will be required.

## 7.2 APPLYING ATOMISTIC AND COARSE-GRAINED SIMULATION TO REFLECTOMETRY ANALYSIS

Chapter 3 focussed on the use of conventional off-the-shelf atomistic and coarse-grained potential models to simulate a phospholipid monolayer at the air-water interface. This is the first study to compare three-levels of simulation coarse-graining to a chemically-consistent modelling approach for the analysis of neutron reflectometry data. The reflectometry was calculated directly from the simulation trajectory and compared with the chemically-consistent analysis process developed in Chapter 3. While the chemically-consistent method produced a better fit to the experimental data, the atomistic Lipid and united-atom Berger models offered very good agreement, in particular when considering the substantial constraint implicit in the determination of a reflectometry profile

<sup>5</sup> McCluskey *et al.*, “Bayesian Determination of the Effect of a Deep Eutectic Solvent on the Structure of Lipid Monolayers”, see n. ??; McCluskey *et al.*, *Lipids\_at\_airdes* (Version 1.0), see n. 68.

<sup>6</sup> In addition to the `MDSimulation` Python class developed in Chapter 4.

<sup>7</sup> Nelson *et al.*, see n. 29; Nelson *et al.*, see n. 66; A. R. Nelson. *Refnx-Models*. URL: <https://github.com/refnx/refnx-models> (Accessed 2019-4-7).

<sup>8</sup> NR

<sup>9</sup> *SASfit*, see n. ??; *SASview* for Small Angle Scattering, see n. ??.

<sup>10</sup> Therefore, leading to an underestimation in the inverse uncertainties of the model.

directly from a simulation. The coarse-grained MARTINI potential model, however, did not fare well, with some severe difficulties, from water bead freezing to the inability to coarse-grain an 18-carbon chain with the MARTINI's 4-to-1 beading structure. The atomistic simulations were used to better understand the nature of a phospholipid monolayer at the air-water interface to improve the chemically-consistent model in the future. For example, it was observed that the uniform solvation that is commonly used for the phospholipid head group, and used in the chemically-consistent model, may not be accurate as the solvation varies throughout the monolayer depending on the chemistry of the head-solvent interactions.

This work offers two important opportunities for future work in the area. The first is the advancement of the use for simulation-driven analysis in NR measurements; where it was suggested that either the united-atom or atomistic simulations would be capable of reproducing experimental data. Therefore it would be pertinent to develop a workflow to produce starting structure for different, common, reflectometry systems, such as monolayers, bilayers, etc. This workflow could be implemented on computing resources at the large scale facilities and allow users to state the phospholipid type and the expected structure, leveraging existing software such as Packmol and GROMACS.<sup>11</sup> to build a starting structure and run a simulation. Such a workflow would allow users of NR and XRR instruments to easily set up and run simulations to match the experiments that they are conducting. Furthermore, the computational requirements to perform these simulations are becoming more accessible with improving compute resource at national facilities, so it would be possible for the user to perform these simulations concurrently with experiments.

The second area of future work would be the improvement of the chemically-consistent monolayer model based on the findings from this work. As mentioned above, the atomistic simulation showed the inadequacy of the use of a uniform solvent, therefore it may be useful to investigate the use of a three-layer model to describe a phospholipid monolayer system,<sup>12</sup> as it was shown that the solvation at the carbonyl group is greater than would be expected for a simple sigmoidal decay. Furthermore, it was noted that in disagreement with the work of Campbell *et al.*,<sup>13</sup> it appears the interfacial roughness between the layers is not carried in a conformal fashion. Instead, the roughness appeared to increase from the tail to the head of the phospholipid, this suggests that a more accurate monolayer model would not constrain the interface roughness to be conformal when only a single phospholipid type is present, as is the case in the chemically-consistent monolayer model.

<sup>11</sup> Martínez *et al.*, see n. 124; Lindahl *et al.*, see n. 46, in a similar fashion the building of a monolayer in this work.

<sup>12</sup> Where the head layer is split in two to allow for different solvation.

<sup>13</sup> Campbell *et al.*, see n. 26.

### 7.3 USING PARTICLE SWARM METHODS FOR SMALL ANGLE SCATTERING ANALYSIS

The simulation of micellar species typically involves either the simulation of a random solution of surfactant molecules until a micelle-like structure forms, or the artificial creation of a micelle of appropriate size based on *a priori* information. In an effort to improve the ability for the simulation-driven analysis of SAS data from micellar solutions, a particle swarm optimisation<sup>14</sup> algorithm was used to attempt to generate a starting structure. This PSO was implemented in the `fitoog` software and used a severely coarse-grained description of the molecules to try and optimise to the scattering profile alone. Despite some success on a small test system, this method was not able to be applied to a case with real experimental data.

The failure of the PSO to be able to produce a realistic micellar structure indicates that it may be necessary to consider some energetic optimisation alongside the purely structural one. Previous work has involved the use of potential models that are biased on the agreement with experimental scattering.<sup>15</sup> However, these still require a significant computational resource to run and obtain useful information. A pragmatic approach to enable experimental users to use simulation-driven analysis would be to implement the methods applied previously in an easy-to-operate software that would allow for access to high-performance computing resources. Such an implementation is already available within the biological scattering community in the form of the SASSIE<sup>16</sup> and WAXSiS<sup>17</sup> resources. However, there is still the opportunity for a straightforward solution to the problem of easily generating realistic starting structures for the simulation of micellar solutions, using the optimisation of structural and energetic information. This may come in the form of an optimised version of the coarse-grained implementation of the Empirical Potential Structure Refinement,<sup>18</sup> or a generalisation of the workflow developed by Ivanovic *et al.*<sup>19</sup>

### 7.4 DEVELOPING OPEN-SOURCE TEACHING MATERIALS FOR CLASSICAL SIMULATION AND SCATTERING

An important challenge for ensuring that users of scattering techniques apply simulation-driven analysis methods is to provide suitable training material. Thus to accompany the development of these coarse-grained methodologies applied to the analysis of scattering measurements, teaching resources have been created to introduce users of scattering methods to classical simulation.

Driven by the growing interest in using simulation to analyse SAS experiments,<sup>20</sup> the open learning module entitled “The interaction between simulation and scattering” was developed.<sup>21</sup> This was designed to introduce users of scattering methods to the underlying

<sup>14</sup> PSO.

<sup>15</sup> Hargreaves *et al.*, see n. 35; Soper *et al.*, see n. 64; Ivanović *et al.*, see n. 36.

<sup>16</sup> Perkins *et al.*, see n. 2.

<sup>17</sup> Knight *et al.*, see n. 6.

<sup>18</sup> Soper *et al.*, see n. 64.

<sup>19</sup> Ivanović *et al.*, see n. 36.

<sup>20</sup> Hub, see n. 68; Perkins *et al.*, see n. 2.

<sup>21</sup> McCluskey *et al.*, “An Introduction to Classical Molecular Dynamics Simulation for Experimental Scattering Users”, see n. ??; McCluskey *et al.*, `Pythoninchemistry/Sim_and_scat:Sim_and_scat-v0.3-Preprint`, see n. 42.

mechanics of classical molecular dynamics<sup>22</sup> simulations through the use of Python-based examples, to stimulate the worked example effect. In addition to showing how classical MD worked, this also showed how the scattering profile may be determined directly from the simulation trajectory. Alongside this open learning module, the Python package `pylj` was developed.<sup>23</sup> This package was created to introduce classical MD and the results available to these simulations in a computational laboratory fashion. Open-source and written in the readable Python language this software aimed to engage students to learn more about simulation methods by interacting with the simulations through the easy visualisation environment, which aimed to provide a straightforward way to rationalise the system being simulated. This package was used in the open learning module, has already been applied in the undergraduate computational chemistry laboratory at the University of Bath, and will in the future be used at the University of Bristol.

The open source nature of both of these resources ensures that they are available to anyone to use, change, or build upon. This means that in the future the open learning module will be able to grow as others, and myself, improve it. For example, it would be relevant to include a more detailed description of the particulars of calculating a “realistic” scattering profile from a simulation, as currently aspects such as instrumental resolution and absorption effects are not covered. Additionally, as more and more methods for producing starting structures for classical simulations that are relevant to scattering are developed, for example by the CCP-SAS<sup>24</sup> consortium, it would be useful to users for detailed descriptions of these to be included. Finally, this example of an open, online, interactive learning module has the potential to encourage the creation of many more from the chemical sciences. Such topics as, examples of how to use Python for chemical data analysis or topics such as analytical methods (e.g. NMR spectroscopy) would be useful resources for students and academics alike.

Currently, the `pylj` package only allows the simulation of a single particle type, with no inclusion of aspects such as mixing rules. The planned `pylj` 2.0 version would adapt the underlying structures of the `pylj` software to enable these features. This would substantially improve the utility of `pylj` in the chemical sciences, as it would be more straightforward to simulate mixtures of chemical species and aspects such as chemical bonding. As mentioned, the `pylj` software is used in the undergraduate chemistry course at the University of Bath and currently, a course is being developed by the University of Bristol that will feature the use of `pylj`. It is hoped that as more universities make use of the software, the community behind it will grow.

<sup>22</sup> MD.

<sup>23</sup> McCluskey *et al.*, “Pylj: A Teaching Tool for Classical Atomistic Simulation”, see n. ??; McCluskey *et al.*, *Arm61/Pylj: Pylj-1.2.5*, see n. 27.

<sup>24</sup> The collaborative computational project for SAS.



# A

## ADDITIONAL CODE BLOCKS FOR DIFFERENTIAL EVOLUTION

This appendix includes additional Code Blocks relevant to the introduction of the DE optimisation method discussed in Section 2.4.1. Code Block A.1 describes the mutation step of the DE algorithm, as described by Björck.<sup>1</sup> Code Block A.2 describes the

<sup>1</sup> Björck, “Fitting with Differential Evolution: An Introduction and Evaluation”.

```
import numpy as np

def mutation(p, b, km):
    m = np.zeros_like(p)
    R = np.random.randint(p.shape[1], size=(2, p.shape[1]))
    for j in range(p.shape[1]):
        m[:, j] = b + km * (p[:, R[0, j]] - p[:, R[1, j]])
    return m
```

recombination step of the DE algorithm, as described by Björck.

```
import numpy as np

def recombination(p, m, kr):
    o = np.array(p)
    rand = np.random.rand(p.shape[0], p.shape[1])
    o[rand < kr] = m[rand < kr]
    return o
```

Code Block A.3 describes the classical step of the DE algorithm, as described by Björck.

```
import numpy as np

def selection(p, o, f):
    new_p = np.array(p)
    for j in range(p.shape[1]):
        p_fom = f(p[:, j])
        o_fom = f(o[:, j])
        if o_fom < p_fom:
            new_p[:, j] = o[:, j]
    return new_p
```

Code Block A.1: The mutation step method for a DE algorithm, as described in M. Björck. *J. Appl. Crystallogr.* 44.6 (2011), pp. 1198–1204. The input variables are  $p$  which is an array of floats representing the parent population,  $b$  which is an array of floats representing the best vector of the parent population, and  $km$  the mutant constant. This function returns an array of floats representing the mutant vector.

Code Block A.2: The recombination step method for a DE algorithm, as described in M. Björck. *J. Appl. Crystallogr.* 44.6 (2011), pp. 1198–1204. The input variables are  $p$  which is an array of floats representing the parent population,  $m$  which is an array of floats representing the mutant vector, and  $kr$  the recombination constant. This function returns an array of floats representing the offspring population.

Code Block A.3: The classical selection step used in a DE algorithm, as described in M. Björck. *J. Appl. Crystallogr.* 44.6 (2011), pp. 1198–1204. The input variables are  $p$  which is an array of floats representing the parent population,  $m$  which is an array of floats representing the offspring vector, and  $f$  is the function to be minimised. This function returns an array of floats representing the new parent population.



## B

# ADDITIONAL TABLES AND PLOTS FROM CHAPTER 3

This appendix contains additional tables and plots associated with the results presented in Chapter 3. Tables B.1–B.3 give the best fit values from the chemically consistent model of all four lipids at the lowest, second-lowest, and highest surface pressures<sup>1</sup> investigated. The equivalent data for the second-highest SP can be found in Table 3.3. The two dimensional probability distribution function at the lowest, second-lowest, and highest SPs measured for all of the lipids are shown in Figures B.1–B.3. The equivalent for the second-highest SPs can be found in Figure 3.6.

<sup>1</sup>SPs.

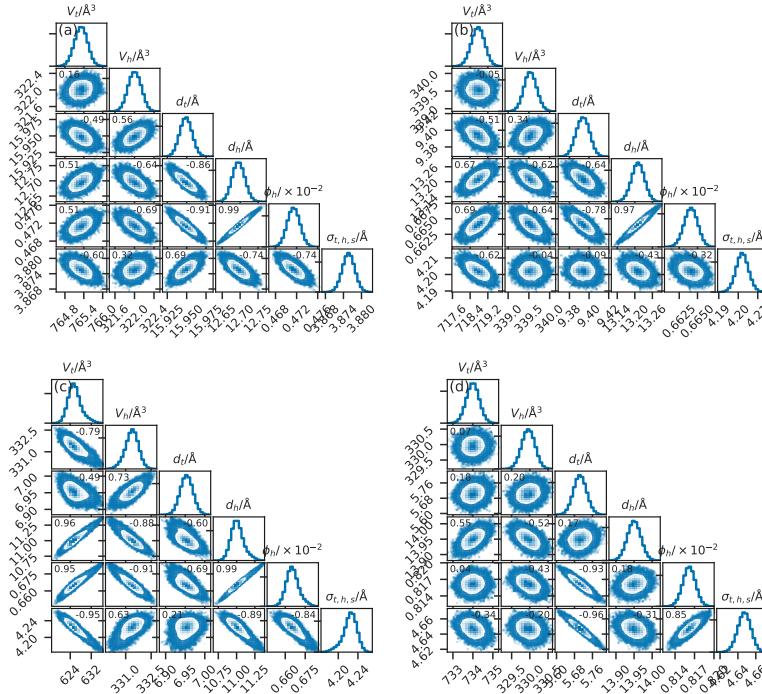


Figure B.1: The probability distribution functions from the chemically-consistent modelling of (a) DPPC at  $15 \text{ mN m}^{-1}$ , (b) DMPC at  $20 \text{ mN m}^{-1}$ , (c) DLPC  $20 \text{ mN m}^{-1}$ , and (d) DMPG at  $15 \text{ mN m}^{-1}$ . The Pearson correlation coefficient for each pair of parameters is given in the top corner of each two-dimensional PDF.

Table B.1: The best-fit values, and associated 95% confidence intervals for each of the varying parameters for each phospholipid at the lowest SP measured from XRR. The values of  $\phi_h$  were obtained from the appropriate use of Equation 3.2.

| Phospholipid<br>SP/mN m <sup>-1</sup> | DPPC                    | DMPC                    | D LPC             | D MFG             |
|---------------------------------------|-------------------------|-------------------------|-------------------|-------------------|
|                                       | 15                      | 20                      | 20                | 15                |
| $V_t/\text{\AA}^3$                    | $765.31 \pm 0.38$       | $718.76 \pm 0.54$       | $625.50 \pm 3.70$ | $734.00 \pm 0.58$ |
| $V_h/\text{\AA}^3$                    | $322.00 \pm 0.24$       | $339.53 \pm 0.27$       | $331.38 \pm 0.81$ | $329.96 \pm 0.34$ |
| $d_h/\text{\AA}$                      | $12.70_{-0.03}^{+0.03}$ | $13.21_{-0.04}^{+0.04}$ | $11.00 \pm 0.13$  | $13.95 \pm 0.03$  |
| $d_t\text{\AA}$                       | $15.95 \pm 0.01$        | $9.39 \pm 0.01$         | $6.95 \pm 0.03$   | $5.70 \pm 0.05$   |
| $\sigma_{t,h,s}/\text{\AA}$           | $3.88 \pm 0.00$         | $4.20 \pm 0.01$         | $4.22 \pm 0.03$   | $4.65 \pm 0.01$   |
| $\phi_h/\times 10^{-2}$               | $47.14 \pm 0.21$        | $66.41 \pm 0.17$        | $66.53 \pm 0.76$  | $81.62 \pm 0.18$  |

Table B.2: The best-fit values, and associated 95% confidence intervals for each of the varying parameters for each phospholipid at the second lowest SP measured from XRR. The values of  $\phi_h$  were obtained from the appropriate use of Equation 3.2.

| Phospholipid<br>SP/mN m <sup>-1</sup> | DPPC<br>20                              | DMPC<br>25                              | DLPC<br>25                             | DMPG<br>20    |
|---------------------------------------|---|---|--|---------------|
| $V_t/\text{\AA}^3$                    | 765.31 ± 0.38                           | 718.76 ± 0.54                           | 625.50 ± 3.70                          | 734.00 ± 0.58 |
| $V_h/\text{\AA}^3$                    | 322.00 ± 0.24                           | 339.53 ± 0.27                           | 331.38 ± 0.81                          | 329.96 ± 0.34 |
| $d_h/\text{\AA}$                      | 12.70 <sup>+0.03</sup> <sub>-0.03</sub> | 13.21 <sup>+0.04</sup> <sub>-0.04</sub> | 11.00 ± 0.13                           | 13.95 ± 0.03  |
| $d_t\text{\AA}$                       | 16.57 <sup>+0.01</sup> <sub>-0.01</sub> | 12.12 ± 0.01                            | 8.29 ± 0.05                            | 10.64 ± 0.01  |
| $\sigma_{t,h,s}/\text{\AA}$           | 4.09 ± 0.00                             | 3.92 ± 0.01                             | 4.36 <sup>+0.02</sup> <sub>-0.03</sub> | 3.94 ± 0.01   |
| $\phi_h/\times 10^{-2}$               | 45.07 ± 0.21                            | 56.66 ± 0.22                            | 60.07 ± 0.96                           | 65.71 ± 0.14  |

Table B.3: The best-fit values, and associated 95% confidence intervals for each of the varying parameters for each phospholipid at the highest SP measured from XRR. The values of  $\phi_i$  were obtained from the appropriate use of Equation 3.2.

| Phospholipid<br>SP/mN m <sup>-1</sup> | DPPC<br>30              | DMPC<br>40              | DPLC<br>35        | DMRG<br>30        |
|---------------------------------------|-------------------------|-------------------------|-------------------|-------------------|
| $V_t/\text{\AA}^3$                    | $765.31 \pm 0.38$       | $718.76 \pm 0.54$       | $625.50 \pm 3.70$ | $734.00 \pm 0.58$ |
| $V_h/\text{\AA}^3$                    | $322.00 \pm 0.24$       | $339.53 \pm 0.27$       | $331.38 \pm 0.81$ | $329.96 \pm 0.34$ |
| $d_h/\text{\AA}$                      | $12.70_{-0.03}^{+0.03}$ | $13.21_{-0.04}^{+0.04}$ | $11.00 \pm 0.13$  | $13.95 \pm 0.03$  |
| $d_t\text{\AA}$                       | $16.91 \pm 0.01$        | $13.82 \pm 0.01$        | $10.30 \pm 0.07$  | $13.99 \pm 0.01$  |
| $\sigma_{t,h,s}/\text{\AA}$           | $4.90 \pm 0.00$         | $4.53 \pm 0.01$         | $4.35 \pm 0.04$   | $4.44 \pm 0.00$   |
| $\phi_h/\times 10^{-2}$               | $43.95 \pm 0.22$        | $50.58 \pm 0.25$        | $50.40 \pm 1.19$  | $54.91 \pm 0.19$  |

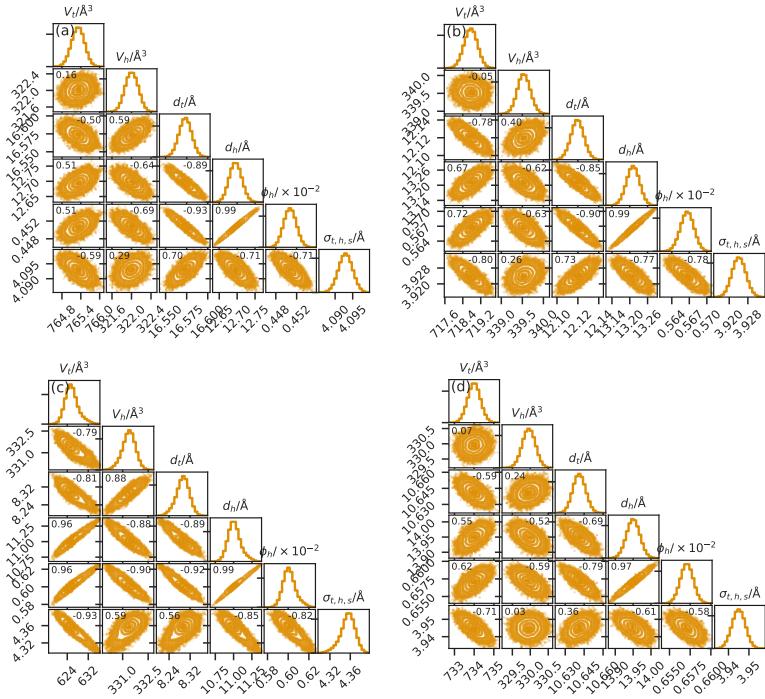


Figure B.2: The probability distribution functions from the chemically-consistent modelling of (a) DPPC at  $20 \text{ mN m}^{-1}$ , (b) DMPC at  $25 \text{ mN m}^{-1}$ , (c) DLPC  $25 \text{ mN m}^{-1}$ , and (d) DMPG at  $20 \text{ mN m}^{-1}$ . The Pearson correlation coefficient for each pair of parameters is given in the top corner of each two-dimensional PDF.

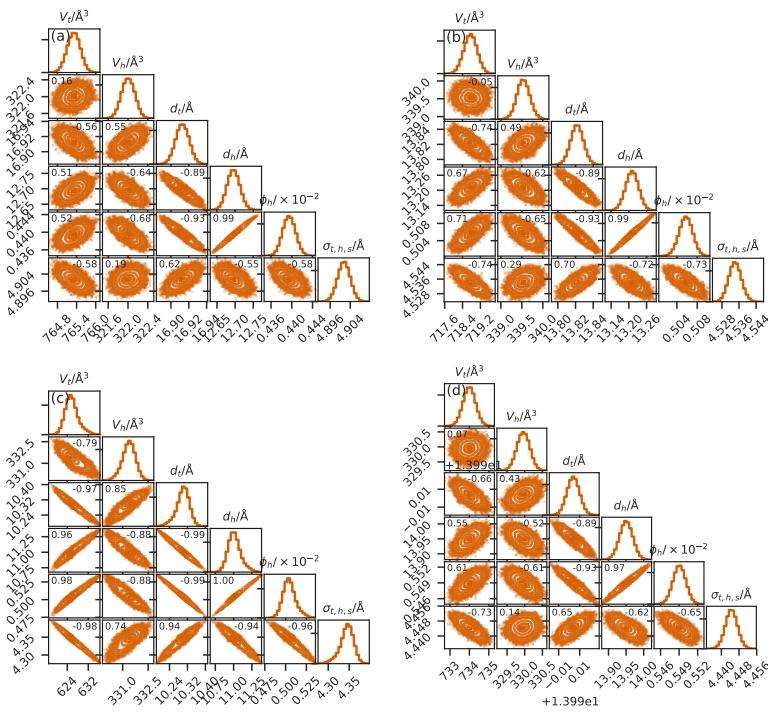
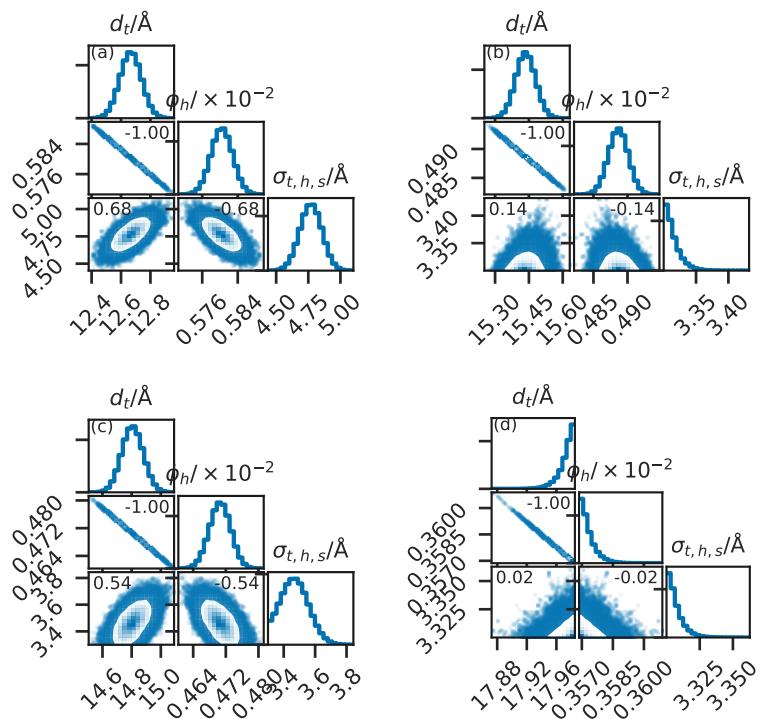


Figure B.3: The probability distribution functions from the chemically-consistent modelling of (a) DPPC at  $30 \text{ mN m}^{-1}$ , (b) DMPC at  $40 \text{ mN m}^{-1}$ , (c) DLPC  $35 \text{ mN m}^{-1}$ , and (d) DMPG at  $30 \text{ mN m}^{-1}$ . The Pearson correlation coefficient for each pair of parameters is given in the top corner of each two-dimensional PDF.

Figure B.4: The probability distribution functions from the chemically-consistent modelling of the NR data; (a) DPPC at  $15 \text{ mN m}^{-1}$ , (b) DPPC at  $20 \text{ mN m}^{-1}$ , (c) DMPC at  $20 \text{ mN m}^{-1}$ , and (d) DMPC at  $25 \text{ mN m}^{-1}$ . The Pearson correlation coefficient for each pair of parameters is given in the top corner of each two-dimensional PDF.



# C

## ADDITIONAL TABLES AND PLOT FROM CHAPTER 4

This appendix contains additional plots and tables associated with Chapter 4. Figures C.1-C.3 give the NR and SLD profiles for the chemically-consistent model alongside those obtained directly from MD simulations at an APM associated with surface pressures of 20, 40 and 50 mN m<sup>-1</sup>. The equivalent data at a APM associated with a surface pressure of 30 mN m<sup>-1</sup> can be found in Figure 4.3.

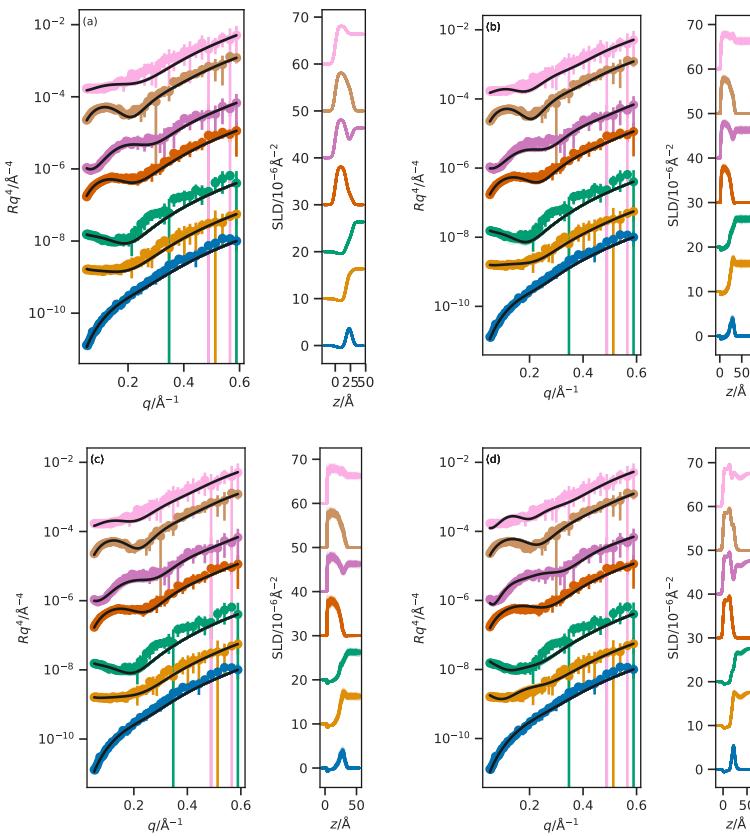
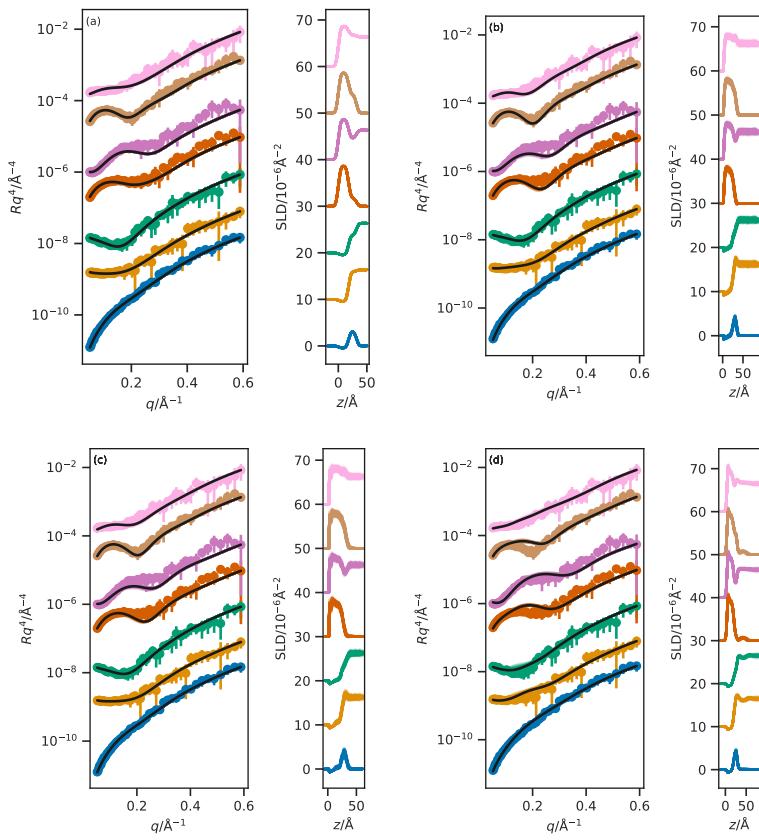


Figure C.1: The NR profiles (left) and SLD profiles (right) determined at an APM associated with a surface pressure of 20 mN m<sup>-1</sup> for; (a) the chemically-consistent model, (b) the Slipid all-atom potential model simulations, (c) the Berger united-atom potential model simulations, and (d) the MARTINI coarse-grained potential model simulations. From top-to-bottom the contrasts are as follows; d<sub>83</sub>-D<sub>2</sub>O, d<sub>70</sub>-D<sub>2</sub>O, d<sub>70</sub>-ACMW, h-D<sub>2</sub>O, d<sub>13</sub>-D<sub>2</sub>O, d<sub>13</sub>-ACMW. The different contrast NR profiles have been offset in the y-axis by an order of magnitude and the SLD profiles offset in the y-axis by  $10 \times 10^{-6} \text{ Å}^{-2}$ , for clarity.

Tables C.1-C.3 give the figure of merit for agreement between the model and the data from each of the analysis method at a APM

Figure C.2: The NR profiles (left) and SLD profiles (right) determined at an APM associated with a surface pressure of  $40 \text{ mN m}^{-1}$  for; (a) the chemically-consistent model, (b) the Slipid all-atom potential model simulations, (c) the Berger united-atom potential model simulations, and (d) the MARTINI coarse-grained potential model simulations. From top-to-bottom the contrasts are as follows;  $d_{83}-\text{D}_2\text{O}$ ,  $d_{83}\text{-ACMW}$ ,  $d_{70}-\text{D}_2\text{O}$ ,  $d_{70}\text{-ACMW}$ ,  $\text{h}-\text{D}_2\text{O}$ ,  $d_{13}-\text{D}_2\text{O}$ ,  $d_{13}\text{-ACMW}$ . The different contrast NR profiles have been offset in the y-axis by an order of magnitude and the SLD profiles offset in the y-axis by  $10 \times 10^{-6} \text{ Å}^{-2}$ , for clarity.



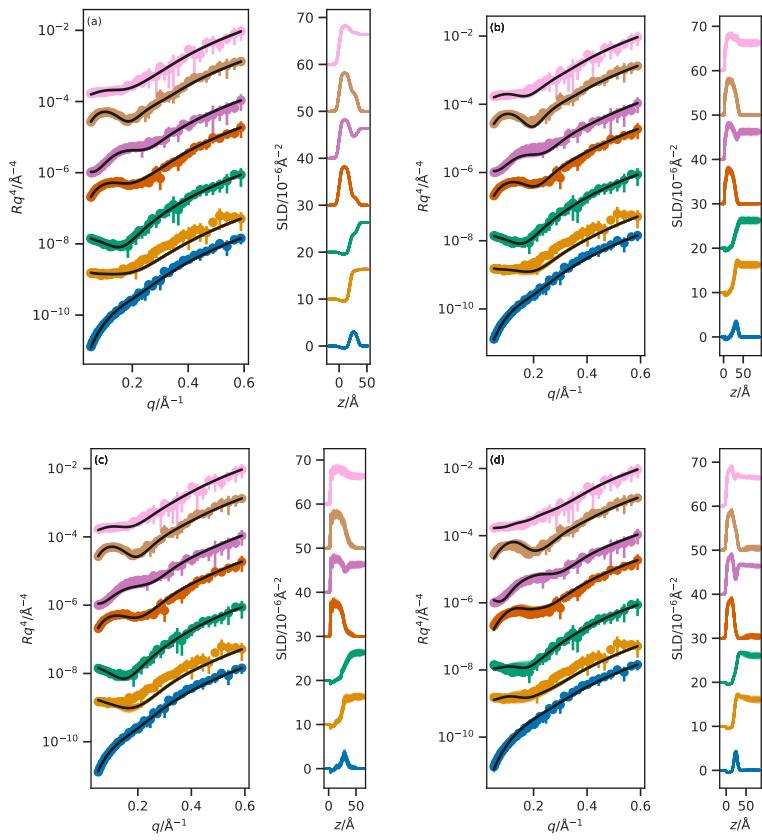


Figure C.3: The NR profiles (left) and SLD profiles (right) determined at an APM associated with a surface pressure of  $50 \text{ mN m}^{-1}$  for; (a) the chemically-consistent model, (b) the Slipid all-atom potential model simulations, (c) the Berger united-atom potential model simulations, and (d) the MARTINI coarse-grained potential model simulations. From top-to-bottom the contrasts are as follows;  $d_{83}\text{-D}_2\text{O}$ ,  $d_{83}\text{-ACMW}$ ,  $d_{70}\text{-D}_2\text{O}$ ,  $d_{70}\text{-ACMW}$ ,  $h\text{-D}_2\text{O}$ ,  $d_{13}\text{-D}_2\text{O}$ ,  $d_{13}\text{-ACMW}$ . The different contrast NR profiles have been offset in the y-axis by an order of magnitude and the SLD profiles offset in the y-axis by  $10 \times 10^{-6} \text{\AA}^{-2}$ , for clarity.

associated with a surface pressure of 20, 40 and 50 mN m<sup>-1</sup>. The equivalent data at a APM associated with a surface pressure of 30 mN m<sup>-1</sup> can be found in Table 4.4.

Table C.1: The  $\chi^2$  values for each of the reflectometry models at an APM associated with a surface pressure of 20 mN m<sup>-1</sup>.

| Contrast                          | Chemically-consistent | Slipids         | Berger          | MARTINI         |
|-----------------------------------|-----------------------|-----------------|-----------------|-----------------|
| h-D <sub>2</sub> O                | 107.98                | 178.23          | 136.85          | 219.52          |
| d <sub>13</sub> -D <sub>2</sub> O | 126.72                | 275.75          | 134.10          | 410.51          |
| d <sub>13</sub> -ACMW             | 122.13                | 144.86          | 135.12          | 163.44          |
| d <sub>70</sub> -D <sub>2</sub> O | 120.82                | 596.71          | 474.22          | 1154.94         |
| d <sub>70</sub> -ACMW             | 212.43                | 125.80          | 143.58          | 728.36          |
| d <sub>83</sub> -D <sub>2</sub> O | 414.13                | 435.28          | 566.50          | 2800.97         |
| d <sub>83</sub> -ACMW             | 111.04                | 123.85          | 340.70          | 1170.32         |
| Average                           | 173.61 ± 103.66       | 268.64 ± 247.12 | 275.87 ± 214.69 | 949.72 ± 857.48 |

Table C.2: The  $\chi^2$  values for each of the reflectometry models at an APM associated with a surface pressure of 40 mN m<sup>-1</sup>.

| Contrast                          | Chemically-consistent | Slipids         | Berger          | MARTINI         |
|-----------------------------------|-----------------------|-----------------|-----------------|-----------------|
| h-D <sub>2</sub> O                | 52.56                 | 204.90          | 119.12          | 343.36          |
| d <sub>13</sub> -D <sub>2</sub> O | 43.57                 | 253.44          | 87.77           | 389.37          |
| d <sub>13</sub> -ACMW             | 48.62                 | 80.34           | 44.94           | 143.96          |
| d <sub>70</sub> -D <sub>2</sub> O | 286.74                | 641.01          | 655.17          | 388.97          |
| d <sub>70</sub> -ACMW             | 156.97                | 193.67          | 188.39          | 1526.73         |
| d <sub>83</sub> -D <sub>2</sub> O | 383.58                | 256.57          | 396.31          | 433.68          |
| d <sub>83</sub> -ACMW             | 100.56                | 248.24          | 284.53          | 2436.08         |
| Average                           | 153.23 ± 123.50       | 268.31 ± 221.16 | 253.75 ± 259.59 | 808.88 ± 843.09 |

Table C.3: The  $\chi^2$  values for each of the reflectometry models at an APM associated with a surface pressure of 50 mN m<sup>-1</sup>.

| Contrast                          | Chemically-consistent | Slipids         | Berger          | MARTINI           |
|-----------------------------------|-----------------------|-----------------|-----------------|-------------------|
| h-D <sub>2</sub> O                | 31.90                 | 148.57          | 165.63          | 1353.62           |
| d <sub>13</sub> -D <sub>2</sub> O | 92.92                 | 300.25          | 622.46          | 866.44            |
| d <sub>13</sub> -ACMW             | 90.45                 | 50.38           | 51.73           | 327.25            |
| d <sub>70</sub> -D <sub>2</sub> O | 120.39                | 643.27          | 743.62          | 572.19            |
| d <sub>70</sub> -ACMW             | 88.06                 | 139.17          | 204.57          | 3063.36           |
| d <sub>83</sub> -D <sub>2</sub> O | 188.17                | 228.03          | 263.74          | 269.57            |
| d <sub>83</sub> -ACMW             | 75.90                 | 167.48          | 89.11           | 4315.89           |
| Average                           | 98.26 ± 44.19         | 239.59 ± 219.99 | 305.84 ± 307.31 | 1538.33 ± 1612.94 |

The solvent penetration at APMs associated with surface pressures of 20, 40 and 50 mN m<sup>-1</sup> is shown in Figures C.4-C.6 using the intrinsic density profile. The equivalent figure for an APM associated with a surface pressure of 30 mN m<sup>-1</sup> can be found in Figure 4.6.

The quantification of the interfacial roughness determined from the Slipids simulations at APMs associated with surface pressures of 20, 40 and 50 mN m<sup>-1</sup> are given in Tables

reftab:spread1-C.6. The equivalent data for an APM associated with a surface pressure of 30 mN m<sup>-1</sup> can be found in Table 4.6.

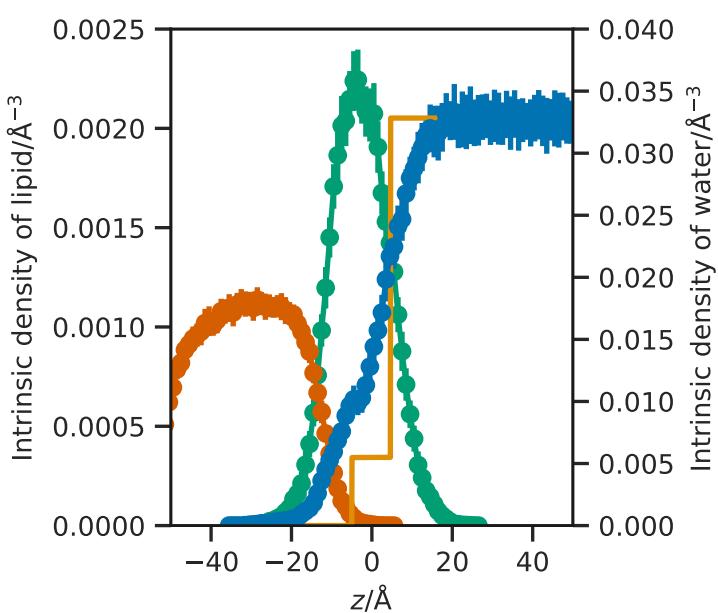


Figure C.4: The simulation time-averaged intrinsic density profile of the water molecules (blue dots), and phospholipid components (head groups: green dots, tail groups: red dots) at an APM associated with a surface pressure of  $20 \text{ mN m}^{-1}$ , where the phosphorus atoms of the phospholipid heads create the intrinsic surface at  $z = 0\text{\AA}$ , and the equivalent number density from the chemically-consistent model (orange line).

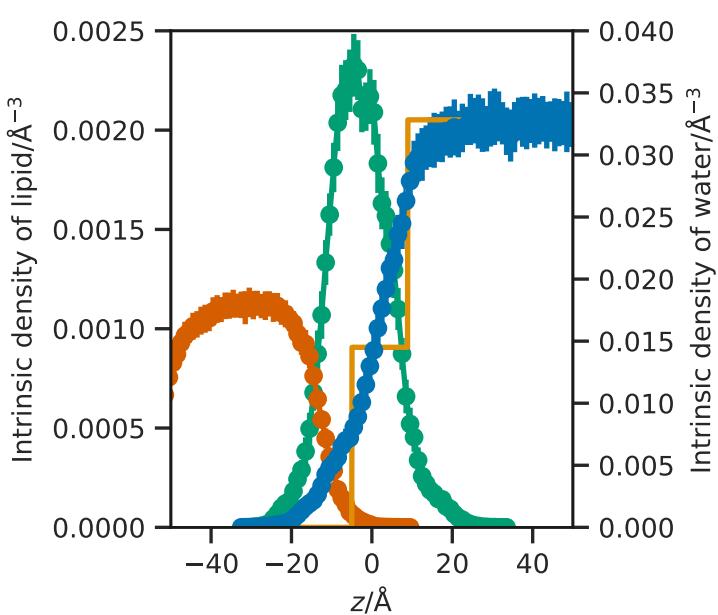


Figure C.5: The simulation time-averaged intrinsic density profile of the water molecules (blue dots), and phospholipid components (head groups: green dots, tail groups: red dots) at an APM associated with a surface pressure of  $40 \text{ mN m}^{-1}$ , where the phosphorus atoms of the phospholipid heads create the intrinsic surface at  $z = 0\text{\AA}$ , and the equivalent number density from the chemically-consistent model (orange line).

Figure C.6: The simulation time-averaged intrinsic density profile of the water molecules (blue dots), and phospholipid components (head groups: green dots, tail groups: red dots) at an APM associated with a surface pressure of  $50 \text{ mN m}^{-1}$ , where the phosphorus atoms of the phospholipid heads create the intrinsic surface at  $z = 0\text{\AA}$ , and the equivalent number density from the chemically-consistent model (orange line).

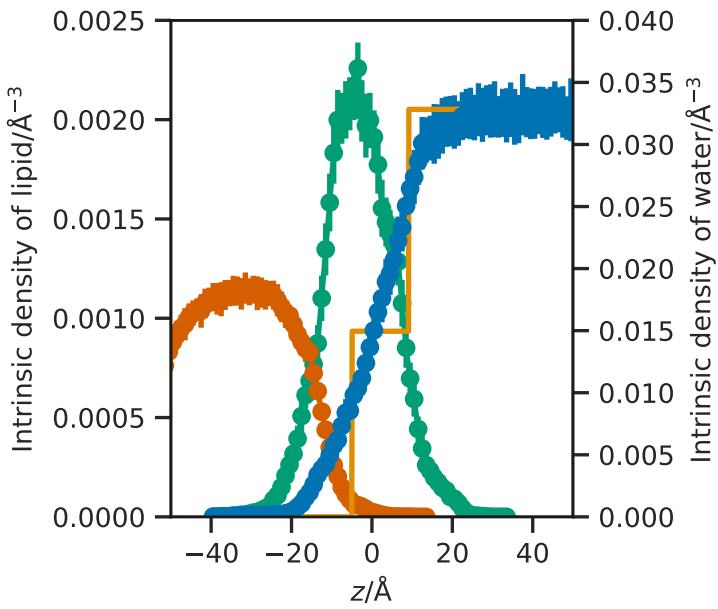


Table C.4: The mean, 95 % quantile, and their spread for the  $z$ -dimension position of atoms representative of difference parts of the phospholipid, at an APM associated with a surface pressure of  $20 \text{ mN m}^{-1}$ .

| Position     | Mean/\text{\AA} | 95 % quantile/\text{\AA} | Spread/\text{\AA} |
|--------------|-----------------|--------------------------|-------------------|
| Start-Head   | 64.2            | 73.7                     | 9.5               |
| Mid-Head     | 65.1            | 74.2                     | 9.1               |
| End-Head     | 68.2            | 77.2                     | 9.0               |
| Start-Tail 1 | 69.4            | 78.0                     | 8.5               |
| Start-Tail 2 | 70.4            | 79.3                     | 8.8               |
| Mid-Tail 1   | 78.3            | 84.0                     | 5.8               |
| Mid-Tail 2   | 79.6            | 85.6                     | 6.0               |
| End-Tail 1   | 88.5            | 90.5                     | 2.1               |
| End-Tail 2   | 89.6            | 90.7                     | 1.1               |

Table C.5: The mean, 95 % quantile, and their spread for the  $z$ -dimension position of atoms representative of difference parts of the phospholipid, at an APM associated with a surface pressure of  $40 \text{ mN m}^{-1}$ .

| Position     | Mean/\text{\AA} | 95 % quantile/\text{\AA} | Spread/\text{\AA} |
|--------------|-----------------|--------------------------|-------------------|
| Start-Head   | 69.4            | 78.6                     | 9.2               |
| Mid-Head     | 70.3            | 79.8                     | 9.5               |
| End-Head     | 73.5            | 82.8                     | 9.4               |
| Start-Tail 1 | 74.7            | 83.6                     | 8.9               |
| Start-Tail 2 | 75.9            | 85.0                     | 9.1               |
| Mid-Tail 1   | 83.7            | 89.6                     | 5.9               |
| Mid-Tail 2   | 85.3            | 91.9                     | 6.6               |
| End-Tail 1   | 94.2            | 96.9                     | 2.7               |
| End-Tail 2   | 95.6            | 97.1                     | 1.5               |

| Position     | Mean/ $\text{\AA}$ | 95 % quantile/ $\text{\AA}$ | Spread/ $\text{\AA}$ |
|--------------|--------------------|-----------------------------|----------------------|
| Start-Head   | 71.3               | 83.2                        | 11.9                 |
| Mid-Head     | 72.6               | 83.9                        | 11.2                 |
| End-Head     | 75.6               | 86.8                        | 11.2                 |
| Start-Tail 1 | 76.7               | 87.8                        | 11.1                 |
| Start-Tail 2 | 77.9               | 89.0                        | 11.1                 |
| Mid-Tail 1   | 85.4               | 94.8                        | 9.4                  |
| Mid-Tail 2   | 87.3               | 96.4                        | 9.1                  |
| End-Tail 1   | 95.8               | 100.3                       | 4.5                  |
| End-Tail 2   | 97.6               | 100.4                       | 2.9                  |

Table C.6: The mean, 95 % quantile, and their spread for the z-dimension position of atoms representative of difference parts of the phospholipid, at an APM associated with a surface pressure of  $50 \text{ mN m}^{-1}$ .



## REFERENCES

- Abelès, F. "Sur La Propagation Des Ondes Électromagnétiques Dans Les Milieux Sratifiés". *Ann. Phys.* 12.3 (1948), pp. 504–520. doi: 10.1051/anphys/194812030504.
- Ackley, D. H. "A Connectionist Machine for Genetic Hillclimbing". PhD. Michigan, US: University of Michigan, 1987. 260 pp.
- Aiello-Nicosia, M. L. and R. M. Sperandeo-Mineo. "Computer Simulation of a Two-Dimensional Ideal Gas: A Simple Molecular Dynamics Method for Teaching Purposes". *Eur. J. Phys.* 6.3 (1985), pp. 148–153. doi: 10.1088/0143-0807/6/3/005.
- Albrecht, O., H. Gruler, and E. Sackmann. "Polymorphism of Phospholipid Monolayers". *J. Phys. France* 39.3 (1978), pp. 301–313. doi: 10.1051/jphys:01978003903030100.
- Allen, D. T., Y. Saaka, L. C. Pardo, M. J. Lawrence, and C. D. Lorenz. "Specific Effects of Monovalent Counterions on the Structural and Interfacial Properties of Dodecyl Sulfate Monolayers". *Phys. Chem. Chem. Phys.* 18.44 (2016), pp. 30394–30406. doi: 10.1039/C6CP05714D.
- Amdahl, G. M. "Validity of a Single Processor Approach to Achieving Large-Scale Computing Capabilities". In: *AFIPS Conference Proceedings*. AFIPS Conference. 1967. doi: 10.1145/1465482.1465560.
- Andersen, H. C. "Molecular Dynamics Simulations at Constant Pressure and/or Temperature". *J. Chem. Phys.* 72.4 (1980), pp. 2384–2393. doi: 10.1063/1.439486.
- Anderson, P. M. and M. R. Wilson. "Molecular Dynamics Simulations of Amphiphilic Graft Copolymer Molecules at a Water/Air Interface". *J. Chem. Phys.* 121.17 (2004), p. 8503. doi: 10.1063/1.1796251.
- Anton, N., H. Mojzisova, E. Porcher, J.-P. Benoit, and P. Saulnier. "Reverse Micelle-Loaded Lipid Nano-Emulsions: New Technology for Nano-Encapsulation of Hydrophilic Materials". *Int. J. Pharm.* 398.1-2 (2010), pp. 204–209. doi: 10.1016/j.ijpharm.2010.07.039.
- Araujo, C. F., J. A. P. Coutinho, M. M. Nolasco, S. F. Parker, P. J. A. Ribeiro-Claro, S. Rudić, B. I. G. Soares, and P. D. Vaz. "Inelastic Neutron Scattering Study of Reline: Shedding Light on the Hydrogen Bonding Network of Deep Eutectic Solvents". *Phys.*

- Chem. Chem. Phys.* 19.27 (2017), pp. 17998–18009. doi: 10.1039/C7CP01286A.
- ArgonMD*. URL: <http://www.argonmd.co.uk> (Accessed 2019-4-4).
- Armen, R. S., O. D. Uitto, and S. E. Feller. “Phospholipid Component Volumes: Determination and Application to Bilayer Structure Calculations”. *Biophys. J.* 75.2 (1998), pp. 734–744. doi: 10.1016/S0006-3495(98)77563-0.
- Arnold, T., A. J. Jackson, A. Sanchez-Fernandez, D. Magnone, A. E. Terry, and K. J. Edler. “Surfactant Behavior of Sodium Dodecylsulfate in Deep Eutectic Solvent Choline Chloride/Urea”. *Langmuir* 31.47 (2015), pp. 12894–12902. doi: 10.1021/acs.langmuir.5b02596.
- Arnold, T., C. Nicklin, J. Rawle, J. Sutter, Trevor Bates, B. Nutter, G. McIntyre, and M. Burt. “Implementation of a Beam Deflection System for Studies of Liquid Interfaces on Beamline I07 at Diamond”. *J. Synchrotron Radiat.* 19.3 (2012), pp. 408–416. doi: 10.1107/S0909049512009272.
- Balgavý, P., N Kučerka, V. I. Gordeliy, and V. G. Cherezov. “Evaluation of Small-Angle Neutron Scattering Curves of Unilamellar Phosphatidylcholine Liposomes Using a Mutishell Model of Bilayer Neutron Scattering Length Density”. *Acta Phys. Slovaca* 51.1 (2001), pp. 53–68.
- Banjare, M. K., K. Behera, M. L. Satnami, S. Pandey, and K. K. Ghosh. “Self-Assembly of a Short-Chain Ionic Liquid within Deep Eutectic Solvents”. *RSC Adv.* 8.15 (2018), pp. 7969–7979. doi: 10.1039/C7RA13557B.
- Banks, J. L., H. S. Beard, Y. Cao, A. E. Cho, Wolfgang Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, and R. M. Levy. “Integrated Modeling Program, Applied Chemical Theory (IMPACT)”. *J. Comput. Chem.* 26.16 (2005), pp. 1752–1780. doi: 10.1002/jcc.20292.
- Barba, L. A., L. J. Barker, D. Blank, J. Brown, A. Downey, T. George, L. Heagy, K. Mandli, J.K. Moore, D. Lippert, K. E. Niemeyer, R. Watkins, R. H. West, E. Wickes, C. Willing, and M. Zingale. *Teaching and Learning with Jupyter*. 2019. URL: <https://jupyter4edu.github.io/jupyter-edu-book/> (Accessed 2019-3-4).
- Barba, L. A., A. Wickenheiser, and R. Watkins. *CyberTraining: DSE – The Code Maker: Computational Thinking for Engineers with Interactive Contextual Learning*. 2017. URL: <https://doi.org/10.6084/m9.figshare.5662051.v1>.
- Barker, R. D., L. E. McKinley, and S. Titmuss. “Neutron Reflectivity as a Tool for Physics-Based Studies of Model Bacterial Membranes”. In: *Biophysics of Infection*. Ed. by M. C. Leake. New York: Springer, 2016, pp. 261–282. ISBN: 978-3-319-32189-9.

- Bayerl, T. M., R. K. Thomas, J. Penfold, A. Rennie, and E. Sackmann. "Specular Reflection of Neutrons at Phospholipid Monolayers. Changes of Monolayer Structure and Headgroup Hydration at the Transition from the Expanded to the Condensed Phase State". *Biophys. J.* 57.5 (1990), pp. 1095–1098. doi: 10.1016/S0006-3495(90)82628-X.
- Bello, G., A. Bodin, M. J. Lawrence, D. Barlow, A. J. Mason, R. D. Barker, and R. D. Harvey. "The Influence of Rough Lipopolysaccharide Structure on Molecular Interactions with Mammalian Antimicrobial Peptides". *BBA - Biomembranes* 1858.2 (2016), pp. 197–209. doi: 10.1016/j.bbamem.2015.11.007.
- Berendsen, H. J. C., J. R. Grigera, and T. P. Straatsma. "The Missing Term in Effective Pair Potentials". *J. Phys. Chem.* 91.24 (1987), pp. 6269–6271. doi: 10.1021/j100308a038.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Di-Nola, and J. R. Haak. "Molecular Dynamics with Coupling to an External Bath". *J. Chem. Phys.* 81.8 (1984), pp. 3684–3690. doi: 10.1063/1.448118.
- Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. "GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation". *Comput. Phys. Commun.* 91.1-3 (1995), pp. 43–56. doi: 10.1016/0010-4655(95)00042-E.
- Berger, O., O. Edholm, and F. Jähnig. "Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature". *Biophys. J.* 72.5 (1997), pp. 2002–2013. doi: 10.1016/S0006-3495(97)78845-3.
- Björck, M. "Fitting with Differential Evolution: An Introduction and Evaluation". *J. Appl. Crystallogr.* 44.6 (2011), pp. 1198–1204. doi: 10.1107/S0021889811041446.
- Björck, M. and G. Andersson. "GenX: An Extensible X-Ray Reflectivity Refinement Program Utilizing Differential Evolution". *J. Appl. Crystallogr.* 40.6 (2007), pp. 1174–1178. doi: 10.1107/S0021889807045086.
- Bobone, S., Y. Gerelli, M. De Zotti, G. Bocchinfuso, A. Farrotti, B. Orioni, F. Sebastiani, E. Latter, J. Penfold, R. Senesi, F. Formaggio, A. Palleschi, C. Toniolo, G. Fragneto, and L. Stella. "Membrane Thickness and the Mechanism of Action of the Short Peptibol Trichogin GA IV". *(BBA) - Biomembranes* 1828.3 (2013), pp. 1013–1024. doi: 10.1016/j.bbamem.2012.11.033.
- Boldon, L., F. Laliberte, and L. Liu. "Review of the Fundamental Theories behind Small Angle X-Ray Scattering, Molecular Dynamics Simulations, and Relevant Integrated Application". *Nano Rev.* 6.1 (2015), p. 25661. doi: 10.3402/nano.v6.25661.
- Bowerman, S., A. S. J. B. Rana, A. Rice, G. H. Pham, E. R. Strieter, and J. Wereszczynski. "Determining Atomistic SAXS Models of Tri-Ubiquitin Chains from Bayesian Analysis of Accelerated

- Molecular Dynamics Simulations". *J. Chem. Theory Comput.* 13.6 (2017), pp. 2418–2429. doi: 10.1021/acs.jctc.7b00059.
- Braun, L., M. Uhlig, R. von Klitzing, and R. A. Campbell. "Polymers and Surfactants at Fluid Interfaces Studied with Specular Neutron Reflectometry". *Adv. Colloid Interface Sci.* 247 (2017), pp. 130–148. doi: 10.1016/j.cis.2017.07.005.
- Brini, E., E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt. "Systematic Coarse-Graining Methods for Soft Matter Simulations – a Review". *Soft Matter* 9.7 (2013), pp. 2108–2119. doi: 10.1039/C2SM27201F.
- Bryant, S. J., R. Atkin, and G. G. Warr. "Spontaneous Vesicle Formation in a Deep Eutectic Solvent". *Soft Matter* 12.6 (2016), pp. 1645–1648. doi: 10.1039/C5SM02660A.
- Bryant, S. J., R. Atkin, and G. G. Warr. "Effect of Deep Eutectic Solvent Nanostructure on Phospholipid Bilayer Phases". *Langmuir* 33.27 (2017), pp. 6878–6884. doi: 10.1021/acs.langmuir.7b01561.
- Buckingham, R. A. "The Classical Equation of State of Gaseous Helium, Neon and Argon". *Proc. Royal Soc. Lond. A.* 168.933 (1938), pp. 264–283. doi: 10.1098/rspa.1938.0173.
- Bucknall, D. G., S. A. Butler, H. E. Hermes, and J. S. Higgins. "Neutron Reflectivity Investigations of Semi-Crystalline Polymer Interfaces". *Physica B* 241–243 (1997), pp. 1071–1073. doi: 10.1016/S0921-4526(97)00793-X.
- Bussi, G., D. Donadio, and M. Parrinello. "Canonical Sampling through Velocity Rescaling". *J. Chem. Phys.* 126.1 (2007), p. 014101. doi: 10.1063/1.2408420.
- Campbell, R. A., Y. Saaka, Y. Shao, Y. Gerelli, R. Cubitt, E. Nazaruk, D. Matyszewska, and M. J. Lawrence. "Structure of Surfactant and Phospholipid Monolayers at the Air/Water Interface Modeled from Neutron Reflectivity Data". *J. Colloid Interface Sci.* 531 (2018), pp. 98–108. doi: 10.1016/j.jcis.2018.07.022.
- Campbell, R. A., H. P. Wacklin, I. Sutton, R. Cubitt, and G. Fragneto. "FIGARO: The New Horizontal Neutron Reflectometer at the ILL". *Eur. Phys. J. Plus* 126.11 (2011), p. 107. doi: 10.1140/epjp/i2011-11107-8.
- Carbone, P., H. A. K. Varzaneh, X. Chen, and F. Müller-Plathe. "Transferability of Coarse-Grained Force Fields: The Polymer Case". *J. Chem. Phys.* 128.6 (2008), p. 064904. doi: 10.1063/1.2829409.
- Chen, P.-C. and J. S. Hub. "Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-Ray Scattering Data". *Biophys. J.* 107.2 (2014), pp. 435–447. doi: 10.1016/j.bpj.2014.06.006.
- Clifton, L. A., M. Sanders, C. Kinane, T. Arnold, K. J. Edler, C. Neylon, R. J. Green, and R. F. Frazier. "The Role of Protein Hydrophobicity in Thionin–Phospholipid Interactions: A Comparison of A1 and A2-Purothionin Adsorbed Anionic Phospholipid

- Monolayers". *Phys. Chem. Chem. Phys.* 14.39 (2012), p. 13569. doi: 10.1039/c2cp42029e.
- Commons, Creative. *Creative Commons – Attribution 4.0 International – CC-BY-4.0*. 2019. URL: <https://creativecommons.org/licenses/by/4.0/>.
- Cordobés, F., J. Muñoz, and C. Gallegos. "Linear Viscoelasticity of the Direct Hexagonal Liquid Crystalline Phase for a Heptane/Nonionic Surfactant/Water System". *J. Colloid Interface Sci.* 187.2 (1997), pp. 401–417. doi: 10.1006/jcis.1996.4707.
- Cordomí, A., G. Caltabiano, and L. Pardo. "Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters". *J. Chem. Theory Comput.* 8.3 (2012), pp. 948–958. doi: 10.1021/ct200491c.
- Coulomb, C. A. "Premier Mémoire Sur l'électricité et Le Magnétisme". *Histoire de l'Académie Royale des Sciences*. Imprimerie Royale (1788), pp. 569–577.
- Coulomb, C. A. "Second Mémoire Sur l'électricité et Le Magnétisme". *Histoire de l'Académie Royale des Sciences*. Imprimerie Royale (1788), pp. 578–611.
- Cython 3.0a0*. URL: <https://cython.readthedocs.io/en/latest/> (Accessed 2019-3-4).
- Dabkowska, A. P., L. E. Collins, D. J. Barlow, R. Barker, S. E. McLain, M. J. Lawrence, and C. D. Lorenz. "Modulation of Dipalmitoylphosphatidylcholine Monolayers by Dimethyl Sulfoxide". *Langmuir* 30.29 (2014), pp. 8803–8811. doi: 10.1021/la501275h.
- Dai, Y., J. van Spronsen, G.-J. Witkamp, R. Verpoorte, and Y. H. Choi. "Natural Deep Eutectic Solvents as New Potential Media for Green Technology". *Anal. Chim. Acta* 766 (2013), pp. 61–68. doi: 10.1016/j.aca.2012.12.019.
- Daillant, J., L. Bosio, and J. J. Benattar. "X-Ray Reflectivity Study of the Liquid-Expanded Liquid-Condensed Phase Transition". *EPL* 12.8 (1990), pp. 715–720. doi: 10.1209/0295-5075/12/8/008.
- Dane, A. D., A. Veldhuis, D. K. G. de Boer, A. J. G. Leenaers, and L. M. C. Buydens. "Application of Genetic Algorithms for Characterization of Thin Layered Materials by Glancing Incidence X-Ray Reflectometry". *Physica B* 253.3-4 (1998), pp. 254–268. doi: 10.1016/S0921-4526(98)00398-6.
- Darré, L., J. Iglesias-Fernandez, A. Kohlmeyer, H. Wacklin, and C. Domene. "Molecular Dynamics Simulations and Neutron Reflectivity as an Effective Approach To Characterize Biological Membranes and Related Macromolecular Assemblies". *J. Chem. Theory Comput.* 11.10 (2015), pp. 4875–4884. doi: 10.1021/acs.jctc.5b00635.
- De Broglie, L. "Recherches Sur La Théorie Des Quanta". *Ann. Phys. (Paris)* 10.3 (1925), pp. 22–125.

- De Haan, V.-O. and G. G. Drijkoningen. "Genetic Algorithms Used in Model Finding and Fitting for Neutron Reflection Experiments". *Physica B* 198.1-3 (1994), pp. 24–26. doi: 10.1016/0921-4526(94)90118-X.
- Debye, P. "Zerstreuung von Röntgenstrahlen". *Ann. Phys.* 351.6 (1915), pp. 809–823. doi: 10.1002/andp.19153510606.
- Democritus: A Molecular Dynamics Tutorial*. URL: <http://www.cse.scitech.ac.uk/ccg/software/Democritus/Democritus.html> (Accessed 2019-4-4).
- Edler, K. J. and D. T. Bowron. "Combining Wide-Angle and Small-Angle Scattering to Study Colloids and Self-Assembly". *Curr. Opin. Colloid Interface Sci.* 20.4 (2015), pp. 227–234. doi: 10.1016/j.cocis.2015.07.002.
- Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. "A Smooth Particle Mesh Ewald Method". *J. Chem. Phys.* 103.19 (1995), pp. 8577–8593. doi: 10.1063/1.470117.
- Evans, P. R. "Rotations and Rotation Matrices". *Acta Crystallogr. D* 57.10 (2001), pp. 1355–1359. doi: 10.1107/S0907444901012410.
- Ewald, P. P. "Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale". *Ann. Phys.* 369.3 (1921), pp. 253–287. doi: 10.1002/andp.19213690304.
- Farago, B. "Recent Developments and Applications of NSE in Soft Matter". *Curr. Opin. Colloid Interface Sci.* 14.6 (2009), pp. 391–395. doi: 10.1016/j.cocis.2009.07.004.
- Flipped Learning | Higher Education Academy*. URL: <https://www.heacademy.ac.uk/knowledge-hub/flipped-learning-0> (Accessed 2019-3-12).
- Foglia, F., G. Fragneto, L. A. Clifton, M. J. Lawrence, and D. J. Barlow. "Interaction of Amphotericin B with Lipid Monolayers". *Langmuir* 30.30 (2014), pp. 9147–9156. doi: 10.1021/la501835p.
- Foglia, F., M. J. Lawrence, and D. J. Barlow. "Studies of Model Biological and Bio-Mimetic Membrane Structure: Reflectivity vs Diffraction, a Critical Comparison". *Curr. Opin. Colloid Interface Sci.* 20.4 (2015), pp. 235–243. doi: 10.1016/j.cocis.2015.08.001.
- Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman. "Emcee : The MCMC Hammer". *Publ. Astron. Soc. PAc.* 125.925 (2013), pp. 306–312. doi: 10.1086/670067.
- Franke, D., C. M. Jeffries, and D. I. Svergun. "Correlation Map, a Goodness-of-Fit Test for One-Dimensional X-Ray Scattering Spectra". *Nat. Methods* 12.5 (2015), pp. 419–422. doi: 10.1038/nmeth.3358.
- Frenkel, D. and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. San Diego, USA: Academic Press, 1996. ISBN: 978-0-12-267351-1.

- Garcia Sakai, V. and A. Arbe. "Quasielastic Neutron Scattering in Soft Matter". *Curr. Opin. Colloid Interface Sci.* 14.6 (2009), pp. 381–390. doi: 10.1016/j.cocis.2009.04.002.
- Garcia-Gutierrez, M. C. and D. R. Rueda. "Bases of Synchrotron Radiation, Light Sources, and Features of X-Ray Scattering Beamlines". In: *Applications of Synchrotron Light to Scattering and Diffraction in Materials and Life Sciences*. Ed. by M. Gomez, A. Nogales, M. C. Garcia-Gutierrez, and T. A. Ezquerro. Berlin, DE: Springer-Verlag Berlin Heidelberg, 2009, pp. 1–22. ISBN: 978-3-540-95968-7.
- Geoghegan, M., R. A. L. Jones, D. S. Sivia, J. Penfold, and A. S. Clough. "Experimental Study of Surface Segregation and Wetting in Films of a Partially Miscible Polymer Blend". *Phys. Rev. E* 53.1 (1996), pp. 825–837. doi: 10.1103/PhysRevE.53.825.
- Gerelli, Y. "Aurore: New Software for Neutron Reflectivity Data Analysis". *J. Appl. Crystallogr.* 49.1 (2016), pp. 330–339. doi: 10.1107/S1600576716000108.
- Gerelli, Y. "Aurore: New Software for Neutron Reflectivity Data Analysis. Corrigendum". *J. Appl. Crystallogr.* 49.2 (2016), pp. 712–712. doi: 10.1107/S1600576716002466.
- Gil, D. L. and D. Windover. "Limitations of X-Ray Reflectometry in the Presence of Surface Contamination". *J. Phys. D* 45.23 (2012), p. 235301. doi: 10.1088/0022-3727/45/23/235301.
- Gilks, W., S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Boca Raton, US: CRC Press, 1995. ISBN: 978-0-412-05551-5.
- Goodman, J. and J. Weare. "Ensemble Samplers with Affine Invariance". *Comm. App. Math. Comp. Sci.* 5 (2010), pp. 65–80. doi: 10.2140/camcos.2010.5.65.
- Goodwin, J. *Colloids and Interfaces with Surfactants and Polymers*. 2nd ed. Chichester, UK: John Wiley & Sons, 2009. ISBN: 978-0-470-51880-9.
- Gowers, R., M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney, and O. Beckstein. "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations". In: Python in Science Conference. Austin, Texas, 2016, pp. 98–105. doi: 10.25080/Majora-629e541a-00e.
- Graner, F., S. Perez-Oyarzun, A. Saint-Jalmes, C. Flament, and F. Gallet. "Phospholipidic Monolayers on Formamide". *J. Phys. II France* 5.2 (1995), pp. 313–322. doi: 10.1051/jp2:1995131.
- Grigoriev, D., R. Krustev, R. Miller, and U. Pison. "Effect of Monovalent Ions on the Monolayers Phase Behavior of the Charged Lipid DPPG". *J. Phys. Chem. B* 103.6 (1999), pp. 1013–1018. doi: 10.1021/jp983704s.
- Grillo, I. "Small-Angle Neutron Scattering and Applications in Soft Condensed Matter". In: *Soft-Matter Characterization*. Ed. by R.

- Borsali and R. Pecora. New York, USA: Springer, 2008, pp. 723–782. ISBN: 978-1-4020-4465-6.
- Gumbart, J., L. G. Trabuco, E. Schreiner, E. Villa, and K. Schulten. “Regulation of the Protein-Conducting Channel by a Bound Ribosome”. *Structure* 17.11 (2009), pp. 1453–1464. doi: 10.1016/j.str.2009.09.010.
- Gutiérrez, M. G., M. L. Ferrer, C. R. Mateo, and F. del Monte. “Freeze-Drying of Aqueous Solutions of Deep Eutectic Solvents: A Suitable Approach to Deep Eutectic Suspensions of Self-Assembled Structures”. *Langmuir* 25.10 (2009), pp. 5509–5515. doi: 10.1021/la900552b.
- Hammond, O. S., D. T. Bowron, and K. J. Edler. “Liquid Structure of the Choline Chloride-Urea Deep Eutectic Solvent (Re-line) from Neutron Diffraction and Atomistic Modelling”. *Green Chem.* 18.9 (2016), pp. 2736–2744. doi: 10.1039/C5GC02914G.
- Hammond, O. S., D. T. Bowron, A. J. Jackson, T. Arnold, A. Sanchez-Fernandez, N. Tsapatsaris, V. Garcia Sakai, and K. J. Edler. “Resilience of Malic Acid Natural Deep Eutectic Solvent Nanostructure to Solidification and Hydration”. *J. Phys. Chem. B* 121.31 (2017), pp. 7473–7483. doi: 10.1021/acs.jpcb.7b05454.
- Hargreaves, R., D. T. Bowron, and K. Edler. “Atomistic Structure of a Micelle in Solution Determined by Wide Q -Range Neutron Diffraction”. *J. Am. Chem. Soc.* 133.41 (2011), pp. 16524–16536. doi: 10.1021/ja205804k.
- Harvey, J. *Computational Chemistry*. Oxford: Oxford University Press, 2018. ISBN: 978-0-19-875550-0.
- Hassan, P. A., G. Fritz, and E. W. Kaler. “Small Angle Neutron Scattering Study of Sodium Dodecyl Sulfate Micellar Growth Driven by Addition of a Hydroscopic Salt”. *J. Colloid Interface Sci.* 257.1 (2003), pp. 154–162. doi: 10.1016/S0021-9797(02)00020-6.
- Hastings, W. K. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. *Biometrika* 57.1 (1970), pp. 97–109. doi: 10.1093/biomet/57.1.97.
- Hayashi, S. and S. Ikeda. “Micelle Size and Shape of Sodium Dodecyl Sulfate in Concentrated Sodium Chloride Solutions”. *J. Phys. Chem.* 84.7 (1980), pp. 744–751. doi: 10.1021/j100444a011.
- Hayter, J. B. and J. Penfold. “An Analytic Structure Factor for Macroion Solutions”. *Mol. Phys.* 42.1 (1981), pp. 109–118. doi: 10.1080/00268978100100091.
- Hayward, D. W., J. B. Gilroy, P. A. Rupar, L. Chabanne, C. Pizzey, M. A. Winnik, G. R. Whittell, I. Manners, and R. M. Richardson. “Liquid Crystalline Phase Behavior of Well-Defined Cylindrical Block Copolymer Micelles Using Synchrotron Small-Angle X-Ray Scattering”. *Macromolecules* 48.5 (2015), pp. 1579–1591. doi: 10.1021/ma502222f.
- Hazell, G., A. P. Gee, T. Arnold, K. J. Edler, and S. E. Lewis. “Langmuir Monolayers Composed of Single and Double Tail Sulfobet-

- taine Lipids". *J. Colloid Interface Sci.* 474 (2016), pp. 190–198. doi: 10.1016/j.jcis.2016.04.020.
- Heinrich, F., H. Nanda, H. Z. Goh, C. Bachert, M. Lösche, and A. D. Linstedt. "Myristoylation Restricts Orientation of the GRASP Domain on Membranes and Promotes Membrane Tethering". *J. Biol. Chem.* 289.14 (2014), pp. 9683–9691. doi: 10.1074/jbc.M113.543561.
- Helm, C. A., H. Möhwald, K. Kjær, and J. Als-Nielsen. "Phospholipid Monolayer Density Distribution Perpendicular to the Water Surface. A Synchrotron X-Ray Reflectivity Study". *EPL* 4.6 (1987), pp. 697–703. doi: 10.1209/0295-5075/4/6/010.
- Hess, B. "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation". *J. Chem. Theory Comput.* 4.1 (2008), pp. 116–122. doi: 10.1021/ct700200b.
- Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation". *J. Chem. Theory Comput.* 4.3 (2008), pp. 435–447. doi: 10.1021/ct700301q.
- Holland, J. H. *Adaptation in Natural and Artificial Systems*. 2nd ed. Boston, USA: MIT Press, 1992. ISBN: 978-0-262-58111-0.
- Hollinshead, C. M., R. D. Harvey, D. J. Barlow, J. R. P. Webster, A. V. Hughes, A. Weston, and M. J. Lawrence. "Effects of Surface Pressure on the Structure of Distearoylphosphatidylcholine Monolayers Formed at the Air/Water Interface". *Langmuir* 25.7 (2009), pp. 4070–4077. doi: 10.1021/la8028319.
- Hoogerheide, D. P., S. Y. Noskov, A. J. Kuszak, S. K. Buchanan, T. K. Rostovtseva, and H. Nanda. "Structure of Voltage-Dependent Anion Channel-Tethered Bilayer Lipid Membranes Determined Using Neutron Reflectivity". *Acta Crystallogr. D* 74.12 (2018), pp. 1219–1232. doi: 10.1107/S2059798318011749.
- Hoover, W. G. "Canonical Dynamics: Equilibrium Phase-Space Distributions". *Phys. Rev. A* 31.3 (1985), pp. 1695–1697. doi: 10.1103/PhysRevA.31.1695.
- Hsieh, Y.-T. and Y.-R. Liu. "Micelle Structure in a Deep Eutectic Solvent for the Electrochemical Preparation of Nanomaterials". *Langmuir* 34.35 (2018), pp. 10270–10275. doi: 10.1021/acs.langmuir.8b01896.
- Hub, J. S. "Interpreting Solution X-Ray Scattering Data Using Molecular Simulations". *Curr. Opin. Struct. Biol.* 49 (2018), pp. 18–26. doi: 10.1016/j.sbi.2017.11.002.
- Hughes, A. V. *RasCAL*. URL: <https://sourceforge.net/projects/rscl/> (Accessed 2016-8-8).
- Hughes, A. V., F. Ciesielski, A. C. Kalli, L. A. Clifton, T. R. Charlton, M. S. P. Sansom, and J. R. P. Webster. "On the Interpretation of Reflectivity Data from Lipid Bilayers in Terms of Molecular-Dynamics Models". *Acta Crystallogr. D* 72.12 (2016), pp. 1227–1240. doi: 10.1107/S2059798316016235.

- Humphrey, W., A. Dalke, and K. Schulten. "VMD: Visual Molecular Dynamics". *J. Mol. Graph.* 14.1 (1996), pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.
- ILL: Neutron for Science: Technical Characteristics.* URL: <https://www.ill.eu/reactor-environment-safety/high-flux-reactor/technical-characteristics/> (Accessed 2016-8-8).
- ISIS – How ISIS Works.* URL: <https://www.isis.stfc.ac.uk/Pages/What-does-ISIS-Neutron-Muon-Source-do.aspx> (Accessed 2018-9-25).
- ISIS – SANS2D.* URL: <https://www.isis.stfc.ac.uk/Pages/Sans2d.aspx> (Accessed 2018-9-25).
- Israelachvili, J. *Intermolecular and Surface Forces*. 3rd ed. Cambridge, USA: Academic Press, 2011. ISBN: 978-0-12-391933-5.
- Ivanović, M. T., L. K. Bruetzel, J. Lipfert, and J. S. Hub. "Temperature-Dependent Atomic Models of Detergent Micelles Refined against Small-Angle X-Ray Scattering Data". *Angew. Chemie Int. Ed.* 57.20 (2018), pp. 5635–5639. doi: 10.1002/anie.201713303.
- Jämbeck, J. P. M. and A. P. Lyubartsev. "An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes". *J. Chem. Theory Comput.* 8.8 (2012), pp. 2938–2948. doi: 10.1021/ct300342n.
- Jämbeck, J. P. M. and A. P. Lyubartsev. "Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids". *J. Phys. Chem. B* 116.10 (2012), pp. 3164–3179. doi: 10.1021/jp212503e.
- Jämbeck, J. P. M. and A. P. Lyubartsev. "Another Piece of the Membrane Puzzle: Extending Lipids Further". *J. Chem. Theory Comput.* 9.1 (2013), pp. 774–784. doi: 10.1021/ct300777p.
- Jensen, G. V., R. Lund, J. Gummel, T. Narayanan, and J. S. Pedersen. "Monitoring the Transition from Spherical to Polymer-like Surfactant Micelles Using Small-Angle X-Ray Scattering". *Angew. Chemie Int. Ed.* 53.43 (2014), pp. 11524–11528. doi: 10.1002/anie.201406489.
- Jmol: An Open-Source Java Viewer for Chemical Structures in 3D.* URL: <http://www.jmol.org/> (Accessed 2018-1-28).
- Johnson, S. J., T. M. Bayerl, D. C. McDermott, G. W. Adam, A. R. Rennie, R. K. Thomas, and E. Sackmann. "Structure of an Adsorbed Dimyristoylphosphatidylcholine Bilayer Measured with Specular Reflection of Neutrons". *Biophys. J.* 59.2 (1991), pp. 289–294. doi: 10.1016/S0006-3495(91)82222-6.
- Jones, E., T. Oliphant, P. Peterson, et al. *SciPy: Open Source Scientific Tools for Python.* URL: <http://www.scipy.org> (Accessed 2019-3-4).
- Jones, L. L., K. D. Jordan, and N. A. Stillings. "Molecular Visualization in Chemistry Education: The Role of Multidisciplinary Collaboration". *Chem. Educ. Res. Pract.* 6.3 (2005), pp. 136–149. doi: 10.1039/B5RP90005K.

- Jones, R. A. L. *Soft Condensed Matter*. Oxford, UK: Oxford University Press, 2002. ISBN: 978-0-19-850589-1.
- Jorge, M. "Molecular Dynamics Simulation of Self-Assembly of *n*-Decyltrimethylammonium Bromide Micelles". *Langmuir* 24.11 (2008), pp. 5714–5725. doi: 10.1021/la800291p.
- Jorgensen, W. L. and J. Tirado-Rives. "The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin". *J. Am. Chem. Soc.* 110.6 (1988), pp. 1657–1666. doi: 10.1021/ja00214a001.
- Jupyter *et al.* "Binder 2.0 - Reproducible, Interactive, Sharable Environments for Science at Scale." In: *Proceedings of the 17th Python in Science Conference*. SciPy2018. 2018. doi: 10.25080/Majora-4af1f417-011.
- Jurašin, D., M. Vinceković, A. Pustak, I. Šmit, M. Bujan, and N. Filipović-Vinceković. "Lamellar to Hexagonal Columnar Liquid Crystalline Phase Transition in a Catanionic Surfactant Mixture: Dodecylammonium Chloride–Sodium Bis(2-Ethylhexyl) Sulfosuccinate". *Soft Matter* 9.12 (2013), p. 3349. doi: 10.1039/c3sm27665a.
- Kamenka, N., Y. Chevalier, and R. Zana. "Aqueous Solutions of Zwitterionic Surfactants with Varying Carbon Number of the Intercharge Group. 1. Micelle Aggregation Numbers". *Langmuir* 11.9 (1995), pp. 3351–3355. doi: 10.1021/la00009a015.
- Kataoka, K., A. Harada, and Y. Nagasaki. "Block Copolymer Micelles for Drug Delivery: Design, Characterization and Biological Significance". *Adv. Drug Deliv. Rev.* 64 (2012), pp. 37–48. doi: 10.1016/j.addr.2012.09.013.
- Kennedy, J. and R. Eberhart. "Particle Swarm Optimization". In: *Proceedings of ICNN'95. International Conference on Neural Networks*. Perth, AU: IEEE, 1995, pp. 1942–1948. doi: 10.1109/ICNN.1995.488968.
- Kewalramani, S., H. Hlaing, B. M. Ocko, I. Kuzmenko, and M. Fukuto. "Effects of Divalent Cations on Phase Behavior and Structure of a Zwitterionic Phospholipid (DMPC) Monolayer at the Air–Water Interface". *J. Phys. Chem. Lett.* 1.2 (2010), pp. 489–495. doi: 10.1021/jz9002873.
- Khodakarimi, S., M. H. Hekmatshoar, and F. Abbasi. "X-Ray Reflectivity and Topography of the Solvent-Treated P3HT:PCBM Thin Films". *J. Mater. Sci. Mater. Electron.* 27.1 (2016), pp. 182–190. doi: 10.1007/s10854-015-3735-3.
- Kienzle, P. A., M. Douchet, D. J. McGilivray, K. V. O'Donovan, N. K. Berk, and C. F. Majkrzak. *NCNR Reflectometry Software*. URL: <http://www.ncnr.nist.gov/reflpak> (Accessed 2018-3-4).
- Klein, R. "Interacting Colloidal Suspensions". In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*. Ed. by Th. Zemb and P. Lindner. Amsterdam, NL: Elsevier Science B.V., 2002, pp. 351–380. ISBN: 978-0-08-093013-8.

- Kluyver, T., B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing. "Jupyter Notebooks-a Publishing Format for Reproducible Computational Workflows". *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (2016), pp. 87–90. doi: 10.3233/978-1-61499-649-1-87.
- Kmiecik, S., D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski. "Coarse-Grained Protein Models and Their Applications". *Chem. Rev.* 116.14 (2016), pp. 7898–7936. doi: 10.1021/acs.chemrev.6b00163.
- Knight, C. J. and J. S. Hub. "WAXSiS: A Web Server for the Calculation of SAXS/WAXS Curves Based on Explicit-Solvent Molecular Dynamics". *Nucleic Acids Res.* 43.W1 (2015), W225–W230. doi: 10.1093/nar/gkv309.
- Knuth, D. E. "Literate Programming". *Comput. J.* 27.2 (1984), pp. 97–111. doi: 10.1093/comjnl/27.2.97.
- Koutsoubas, A. "Combined Coarse-Grained Molecular Dynamics and Neutron Reflectivity Characterization of Supported Lipid Membranes". *J. Phys. Chem. B* 120.44 (2016), pp. 11474–11483. doi: 10.1021/acs.jpcb.6b05433.
- Kozhevnikov, I. V. "General Laws of X-Ray Reflection from Rough Surfaces: II. Conformal Roughness". *Crystallogr. Rep.* 57.4 (2012), pp. 490–498. doi: 10.1134/S1063774512030133.
- Kraack, H., B. M. Ocko, P. S. Pershan, E. Slutskin, and M. Deutsch. "Structure of a Langmuir Film on a Liquid Metal Surface". *Science* 298.5597 (2002), pp. 1404–1407. doi: 10.1126/science.1078372.
- Kraft, D. J., J. Hilhorst, M. A. P. Heinen, M. J. Hoogenraad, B. Luigjes, and W. K. Kegel. "Patchy Polymer Colloids with Tunable Anisotropy Dimensions". *J. Phys. Chem. B* 115.22 (2011), pp. 7175–7181. doi: 10.1021/jp108760g.
- Kratky, O and G Porod. "Diffuse Small-Angle Scattering of x-Rays in Colloid Systems". *J. Colloid Sci.* 4.1 (1949), pp. 35–70. doi: 10.1016/0095-8522(49)90032-X.
- Krauth, W. *Statistical Mechanics: Algorithms and Computations*. Oxford Master Series in Statistical, Computational, and Theoretical Physics 13. Oxford, UK: Oxford University Press, 2006. ISBN: 978-0-19-851536-4.
- Kubo, I., S. Adachi, H. Maeda, and A. Seki. "Phosphatidylcholine Monolayers Observed with Brewster Angle Microscopy and -A Isotherms". *Thin Solid Films* 393.1-2 (2001), pp. 80–85. doi: 10.1016/S0040-6090(01)01101-4.
- Kučerka, N., B. W. Holland, C. G. Gray, B. Tomberli, and J. Katsaras. "Scattering Density Profile Model of POPG Bilayers As Determined by Molecular Dynamics Simulations and Small-Angle Neutron and X-Ray Scattering Experiments". *J. Phys. Chem. B* 116.1 (2012), pp. 232–239. doi: 10.1021/jp208920h.

- Kučerka, N., M. A. Kiselev, and P. Balgavý. "Determination of Bi-layer Thickness and Lipid Surface Area in Unilamellar Dimyristoylphosphatidylcholine Vesicles from Small-Angle Neutron Scattering Curves: A Comparison of Evaluation Methods". *Eur. Biophys. J.* 33.4 (2004), pp. 328–334. doi: 10.1007/s00249-003-0349-0.
- Kunz, K., J. Reiter, A. Goetzelmann, and M. Stamm. "Model-Free Analysis of Neutron Reflectivity Data from Polymer Thin Films with the Simulated Annealing Technique". *Macromolecules* 26.16 (1993), pp. 4316–4323. doi: 10.1021/ma00068a038.
- Larsen, A. H., L. Arleth, and S. Hansen. "Analysis of small-angle scattering data using model fitting and Bayesian regularization". *J. Appl. Crystallogr.* 51.4 (2018), pp. 1151–1161. doi: 10.1107/S1600576718008956.
- Lau, S. and C. Holdgraf. *Jupyter/Jupyter-Book Beta-2*. 2019. URL: <https://github.com/jupyter/jupyter-book>.
- Law, C., D. J. Ashton, N. B. Wilding, and R. L. Jack. "Coarse-Grained Depletion Potentials for Anisotropic Colloids: Application to Lock-and-Key Systems". *J. Chem. Phys.* 145.8 (2016), p. 084907. doi: 10.1063/1.4961541.
- Leach, A. R. *Molecular Modelling: Principles and Applications*. Harlow, UK: Addison Wesley London Ltd, 1996. ISBN: 978-0-582-38210-7.
- Lemkul, J. A. *GROMACS Tutorial: Lysozyme in Water*. URL: <http://www.mdtutorials.com/gmx/lysozyme/index.html> (Accessed 2019-3-8).
- Lennard-Jones, J. E. "On the Determination of Molecular Fields. II. From the Equation of State of a Gas". *Proc. Royal Soc. Lond. A.* 106.738 (1924), pp. 463–477. doi: 10.1098/rspa.1924.0082.
- Levenberg, K. "A Method for the Solution of Certain Non-Linear Problems in Least Squares". *Quart. Appl. Math.* 2.2 (1944), pp. 164–168. doi: 10.1090/qam/10666.
- Lindahl, E., B. Hess, and D. van der Spoel. "GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis". *J. Mol. Model.* 7.8 (2001), pp. 306–317. doi: 10.1007/s008940100045.
- Lipfert, J., L. Columbus, V. B. Chu, S. A. Lesley, and S. Doniach. "Size and Shape of Detergent Micelles Determined by Small-Angle X-Ray Scattering". *J. Phys. Chem. B* 111.43 (2007), pp. 12427–12438. doi: 10.1021/jp0730161.
- Lovell, M. R. and R. M. Richardson. "Analysis Methods in Neutron and X-Ray Reflectometry". *Curr. Opin. Colloid Interface Sci.* 4.3 (1999), pp. 197–204. doi: 10.1016/S1359-0294(99)00039-4.
- Lu, J. R., E. M. Lee, and R. K. Thomas. "The Analysis and Interpretation of Neutron and X-Ray Specular Reflection". *Acta Crystallogr. A* 52.1 (1996), pp. 11–41. doi: 10.1107/S0108767395011202.
- Magnussen, O. M., B. M. Ocko, M. Deutsch, M. J. Regan, P. S. Pershan, D. Abernathy, G. Grübel, and J.-F. Legrand. "Self-

- Assembly of Organic Films on a Liquid Metal". *Nature* 384 (1996), pp. 250–252. doi: 10.1038/384250a0.
- Maillet, J.-B., V. Lachet, and P. V. Coveney. "Large Scale Molecular Dynamics Simulation of Self-Assembly Processes in Short and Long Chain Cationic Surfactants". *Phys. Chem. Chem. Phys.* 1.23 (1999), pp. 5277–5290. doi: 10.1039/a905216j.
- Malliaris, A., J. Le Moigne, J. Sturm, and R. Zana. "Temperature Dependence of the Micelle Aggregation Number and Rate of Intramicellar Excimer Formation in Aqueous Surfactant Solutions". *J. Phys. Chem.* 89.12 (1985), pp. 2709–2713. doi: 10.1021/j100258a054.
- Marquardt, D. W. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". *J. Soc. Indust. Appl. Math.* 11.2 (1963), pp. 431–441. doi: 10.1137/0111030.
- Marrink, S. J., V. Corradi, P. C. T. Souza, H. I. Ingólfsson, D. P. Tieleman, and M. S. P. Sansom. "Computational Modeling of Realistic Cell Membranes". *Chem. Rev.* 119.9 (2019), pp. 6184–6226. doi: 10.1021/acs.chemrev.8b00460.
- Marrink, S. J., X. Periole, D. P. Tieleman, and A. H. de Vries. "Comment on "On Using a Too Large Integration Time Step in Molecular Dynamics Simulations of Coarse-Grained Molecular Models" by M. Winger, D. Trzesniak, R. Baron and W. F. van Gunsteren, *Phys. Chem. Chem. Phys.*, 2009, 11, 1934". *Phys. Chem. Chem. Phys.* 12.9 (2010), p. 2254. doi: 10.1039/b915293h.
- Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations". *J. Phys. Chem. B* 111.27 (2007), pp. 7812–7824. doi: 10.1021/jp071097f.
- Marsh, D. "Molecular Volumes of Phospholipids and Glycolipids in Membranes". *Chem. Phy. Lipids* 163.7 (2010), pp. 667–677. doi: 10.1016/j.chmplyslip.2010.06.005.
- Martínez, L., R. Andrade, E. G. Birgin, and J. M. Martínez. "PACK-MOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations". *J. Comput. Chem.* 30.13 (2009), pp. 2157–2164. doi: 10.1002/jcc.21224.
- Mayans, M. O., W. J. Coadwell, D. Beale, D. B. A. Symons, and S. J. Perkins. "Demonstration by Pulsed Neutron Scattering That the Arrangement of the Fab and Fc Fragments in the Overall Structures of Bovine IgGi and IgG2 in Solution Is Similar". *Biochem. J.* 331 (1995), pp. 283–291. doi: 10.1042/bj3110283.
- McCluskey, A. R. *Arm61/Fitoog: Fitoog-0.1*. 2019. URL: <http://doi.org/10.5281/zenodo.2605181>.
- McCluskey, A. R. and K. J. Edler. "Model-Dependent Small-Angle Scattering for the Study of Complex Organic Materials". *Curr. Org. Chem.* 22.8 (2018), pp. 750–757. doi: 10.2174/1875692115666170612104439.
- McCluskey, A. R., J. Grant, A. J. Smith, J. L. Rawle, D. J. Barlow, M. J. Lawrence, S. C. Parker, and K. J. Edler. "Assessing Molec-

- ular Simulation for the Analysis of Lipid Monolayer Reflectometry". *J. Phys. Comm.* Accepted (2019). doi: 10.1088/2399-6528/ab12a9.
- McCluskey, A. R., J. Grant, A. R. Symington, T. Snow, J. Doutch, B. J. Morgan, S. C. Parker, and K. J. Edler. "An Introduction to Classical Molecular Dynamics Simulation for Experimental Scattering Users". *J. Appl. Crystallogr.* 52.3 (2019), pp. 665–668.
- McCluskey, A. R., J. Grant, A. R. Symington, T. Snow, J. Doutch, B. J. Morgan, S. C. Parker, and K. J. Edler. *Pythoninchemistry/Sim\_and\_scat: Sim\_and\_scat-v0.3-Preprint*. 2019. url: <http://doi.org/10.5281/zenodo.1403828>.
- McCluskey, A. R., B. J. Morgan, K. J. Edler, and S. C. Parker. *Arm61/Pylj: Pylj-1.1.0*. 2018. url: <http://doi.org/10.5281/zenodo.1403828>.
- McCluskey, A. R., B. J. Morgan, K. J. Edler, and S. C. Parker. "Pylj: A Teaching Tool for Classical Atomistic Simulation". *J. Open Source Educ.* 1.2 (2018), pp. 19–21. doi: 10.21105/jose.00019.
- McCluskey, A. R., A. Sanchez-Fernandez, K. J. Edler, S. C. Parker, A. J. Jackson, R. A. Campbell, and T. Arnold. "Bayesian Determination of the Effect of a Deep Eutectic Solvent on the Structure of Lipid Monolayers". *Phys. Chem. Chem. Phys.* 21.11 (2019), pp. 6133–6141. doi: 10.1039/C9CP00203K.
- McCluskey, A. R., A. Sanchez-Fernandez, K. J. Edler, S. C. Parker, A. J. Jackson, R. A. Campbell, and T. Arnold. *Lipids\_at\_airdes (Version 1.0)*. 2019. url: <http://doi.org/10.5281/zenodo.2577796>.
- McCluskey, A. R., A. R. Symington, B. J. Morgan, K. J. Edler, and S. C. Parker. *Arm61/Pylj: Pylj-1.2.1*. 2019. url: <http://doi.org/10.5281/zenodo.2423866>.
- McCluskey, A. R., A. R. Symington, B. J. Morgan, K. J. Edler, and S. C. Parker. *Arm61/Pylj: Pylj-1.2.5*. 2019. url: <http://doi.org/10.5281/zenodo.2587898>.
- McConlogue, C. W. and T. K. Vanderlick. "A Close Look at Domain Formation in DPPC Monolayers". *Langmuir* 13.26 (1997), pp. 7158–7164. doi: 10.1021/la970898e.
- MetalJet X-Ray Source Technology*. url: <https://www.excillum.com/technology/> (Accessed 2018-12-6).
- Metropolis, N., A.W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equation of State Calculations by Fast Computing Machines". *J. Chem. Phys.* 21.6 (1953), pp. 1087–1092. doi: 10.1063/1.1699114.
- Michaud-Agrawal, N., E. J. Denning, T. B. Woolf, and O. Beckstein. "MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations". *J. Comput. Chem.* 32.10 (2011), pp. 2319–2327. doi: 10.1002/jcc.21787.
- Miller, A. F., M. R. Wilson, M. J. Cook, and R. W. Richards. "Monte Carlo Simulations of an Amphiphilic Polymer at a Hydropho-

- bic/Hydrophilic Interface". *Mol. Phys.* 101.8 (2003), pp. 1131–1138. doi: 10.1080/0026897031000068569.
- Miyamoto, S. and P. A. Kollman. "Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models". *J. Comput. Chem.* 13.8 (1992), pp. 952–962. doi: 10.1002/jcc.540130805.
- Mohwald, H. "Phospholipid and Phospholipid-Protein Monolayers at the Air-Water Interface". *Annu. Rev. Phys. Chem.* 41 (1990), pp. 441–476. doi: 10.1146/annurev.pc.41.100190.002301.
- Morse, P. M. "Diatom Molecules According to the Wave Mechanics. II. Vibrational Levels". *Phys. Rev.* 34.1 (1929), pp. 57–64. doi: 10.1103/PhysRev.34.57.
- Nawaz, S., M. Redhead, G. Mantovani, C. Alexander, C. Bosquillon, and P. Carbone. "Interactions of PEO-PPO-PEO Block Copolymers with Lipid Membranes: A Computational and Experimental Study Linking Membrane Lysis with Polymer Structure". *Soft Matter* 8.25 (2012), p. 6744. doi: 10.1039/c2sm25327e.
- Negro, E., R. Latsuzbaia, A. H. de Vries, and G. J. M. Koper. "Experimental and Molecular Dynamics Characterization of Dense Microemulsion Systems: Morphology, Conductivity and SAXS". *Soft Matter* 10.43 (2014), pp. 8685–8697. doi: 10.1039/C4SM01763C.
- Nelson, A. R. *Refnx-Models*. URL: <https://github.com/refnx/refnx-models> (Accessed 2019-4-7).
- Nelson, A. R. J. "Co-Refinement of Multiple-Contrast Neutron/X-Ray Reflectivity Data Using MOTOFIT". *J. Appl. Crystallogr.* 39.2 (2006), pp. 273–276. doi: 10.1107/S0021889806005073.
- Nelson, A. R. J. "Towards a Detailed Resolution Smearing Kernel for Time-of-Flight Neutron Reflectometers. Corrigendum". *J. Appl. Crystallogr.* 47.3 (2014), p. 1162. doi: 10.1107/S1600576714009595.
- Nelson, A. R. J. and C. D. Dewhurst. "Towards a Detailed Resolution Smearing Kernel for Time-of-Flight Neutron Reflectometers". *J. Appl. Crystallogr.* 46.5 (2013), pp. 1338–1346. doi: 10.1107/S0021889813021936.
- Nelson, A. R. J. and S. W. Prescott. "Refnx: Neutron and X-Ray Reflectometry Analysis in Python". *J. Appl. Crystallogr.* 52.1 (2019), pp. 193–200. doi: 10.1107/S1600576718017296.
- Nelson, A. R. J., S. W. Prescott, I. Gresham, and A. R. McCluskey. *Refnx v0.1.2*. 2019. URL: <http://doi.org/10.5281/zenodo.2552023>.
- Nérot, L. and P. Croce. "Caractérisation Des Surfaces Par Réflexion Rasante de Rayons X. Application à l'étude Du Polissage de Quelques Verres Silicates". *Rev. Phys. Appl. (Paris)* 15.3 (1980), pp. 761–779. doi: 10.1051/rphysap:01980001503076100.
- Nosé, S. "A Unified Formulation of the Constant Temperature Molecular Dynamics Methods". *J. Chem. Phys.* 81.1 (1984), pp. 511–519. doi: 10.1063/1.447334.

- Numba: A High Performance Python Compiler.* URL: <http://numba.pydata.org> (Accessed 2019-3-4).
- Ott, F. *SimulReflec.* URL: <http://www-l1b.cea.fr/prism/programs/simulreflec/simulreflec.html> (Accessed 2019-3-4).
- Owejan, J. E., J. P. Owejan, S. C. DeCaluwe, and J. A. Dura. "Solid Electrolyte Interphase in Li-Ion Batteries: Evolving Structures Measured In Situ by Neutron Reflectometry". *Chem. Mater.* 24.11 (2012), pp. 2133–2140. doi: 10.1021/cm3006887.
- Pallant, Amy and Robert F. Tinker. "Reasoning with Atomic-Scale Molecular Dynamic Models". *J. Sci. Educ. Technol.* 13.1 (2004), pp. 51–66. doi: 10.1023/B:JOST.0000019638.01800.d0.
- Pambou, E., J. Crewe, M. Yaseen, F. N. Padia, S. Rogers, D. Wang, H. Xu, and J. R. Lu. "Structural Features of Micelles of Zwitterionic Dodecyl-Phosphocholine (C<sub>12</sub> PC) Surfactants Studied by Small-Angle Neutron Scattering". *Langmuir* 31.36 (2015), pp. 9781–9789. doi: 10.1021/acs.langmuir.5b02077.
- Pan, J., F. A. Heberle, S. Tristram-Nagle, M. Szymanski, M. Koepfinger, J. Katsaras, and N Kučerka. "Molecular Structures of Fluid Phase Phosphatidylglycerol Bilayers as Determined by Small Angle Neutron and X-Ray Scattering". *BBA - Biomembranes* 1818.9 (2012), pp. 2135–2148. doi: 10.1016/j.bbamem.2012.05.007.
- Pandey, A., R. Rai, M. Pal, and S. Pandey. "How Polar Are Choline Chloride-Based Deep Eutectic Solvents?" *Phys. Chem. Chem. Phys.* 16.4 (2014), pp. 1559–1568. doi: 10.1039/C3CP53456A.
- Pandit, S. A., D. Bostick, and M. L. Berkowitz. "An Algorithm to Describe Molecular Scale Rugged Surfaces and Its Application to the Study of a Water/Lipid Bilayer Interface". *J. Chem. Phys.* 119.4 (2003), pp. 2199–2205. doi: 10.1063/1.1582833.
- Papert, S. *Mindstorms*. New York City, US: Basic Books, 1993. ISBN: 978-0-465-04674-4.
- Parratt, L. G. "Surface Studies of Solids by Total Reflection of X-Rays". *Phys. Rev.* 95.2 (1954), pp. 359–369. doi: 10.1103/PhysRev.95.359.
- Parrinello, M. and A. Rahman. "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method". *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190. doi: 10.1063/1.328693.
- Pearson, K. "Notes on Regression and Inheritance in the Case of Two Parents". *Proc. Royal Soc. Lond.* 58 (1895), pp. 240–242.
- Pedersen, J. S. "Monte Carlo Simulation Techniques Applied in the Analysis of Small-Angle Scattering Data from Colloids and Polymer System". In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*. Ed. by Th. Zemb and P. Lindner. Amsterdam, NL: Elsevier Science B.V., 2002, pp. 381–390. ISBN: 978-0-08-093013-8.
- Perkins, S. J., A. S. Nealis, B. J. Sutton, and A. Feinstein. "Solution Structure of Human and Mouse Immunoglobulin M by Syn-

- chrotron X-Ray Scattering and Molecular Graphics Modelling". *J. Mol. Biol.* 221 (1991), pp. 1345–1366. doi: 10.1016/0022-2836(91)90937-2.
- Perkins, S. J., D. W. Wright, H. Zhang, E. H. Brookes, J. Chen, T. C. Irving, S. Krueger, D. J. Barlow, K. J. Edler, D. J. Scott, N. J. Terrill, S. M. King, P. D. Butler, and J. E. Curtis. "Atomistic Modelling of Scattering Data in the Collaborative Computational Project for Small Angle Scattering (CCP-SAS)". *J. Appl. Crystallogr.* 49.6 (2016), pp. 1861–1875. doi: 10.1107/S160057671601517X.
- Plimpton, S. "Fast Parallel Algorithms for Short-Range Molecular Dynamics". *J. Comput. Phys.* 117 (1995), pp. 1–19. doi: 10.1006/jcph.1995.1039.
- Pluhackova, K. and R. A. Böckmann. "Biomembranes in Atomistic and Coarse-Grained Simulations". *J. Phys. Condens. Matter* 27.32 (2015), p. 323103. doi: 10.1088/0953-8984/27/32/323103.
- Poli, R. "Analysis of the Publications on the Applications of Particle Swarm Optimisation". *J. Artif. Evol. Appl.* 2008 (2008), pp. 1–10. doi: 10.1155/2008/685175.
- Politsch, E. and G. Cevc. "Unbiased Analysis of Neutron and X-Ray Reflectivity Data by an Evolution Strategy". *J. Appl. Crystallogr.* 35.3 (2002), pp. 347–355. doi: 10.1107/S0021889802004351.
- Ragan-Kelley, B., Y. Panda, C. Willing, C. Holdgraf, K. Erdogan, T. Head, M. Bussonnier, H. Zarea, M. van Niekerk, C. Yick, and E. Sundell. *Jupyterhub/Binderhub FirstLight*. 2019. URL: <https://github.com/jupyterhub/binderhub>.
- Ragan-Kelley, B., N. M Thiéry, S. Corlay, S. Gutsche, C. Holdgraf, and T. Head. *Minrk/Thebelab v0.3.3*. 2019. URL: <https://github.com/minrk/thebelab>.
- Rahman, A. "Correlations in the Motion of Atoms in Liquid Argon". *Phys. Rev.* 136 (2A 1964), A405–A411. doi: 10.1103/PhysRev.136.A405.
- RCSB PDB: Protein Data Bank. URL: <http://www.rcsb.org> (Accessed 2018-1-28).
- Ribeiro, B. D., C. Florindo, L. C. Iff, M. A. Z. Coelho, and I. M. Marrucho. "Menthol-Based Eutectic Mixtures: Hydrophobic Low Viscosity Solvents". *ACS Sustain. Chem. Eng.* 3.10 (2015), pp. 2469–2477. doi: 10.1021/acssuschemeng.5b00532.
- Rodriguez-Loureiro, I., E. Scoppola, L. Bertinetti, A. Barbetta, G. Fragneto, and E. Schneck. "Neutron Reflectometry Yields Distance-Dependent Structures of Nanometric Polymer Brushes Interacting across Water". *Soft Matter* 13.34 (2017), pp. 5767–5777. doi: 10.1039/C7SM01066D.
- Rodríguez, J. L., M. B. Sierra, P. V. Messina, M. A. Morini, P.C. Schulz, P. del Burgo, E. Junquera, A. Rodríguez, and E. Aicart. "Surface and Bulk Properties of Aqueous Decyltrimethylammonium Bromide–Hexadecyltrimethylammonium Bromide Mixed System". *J. Colloid Interface Sci.* 314.2 (2007), pp. 699–706. doi: 10.1016/j.jcis.2007.06.010.

- Rondelli, V., G. Fragneto, S S., E. Del Favero, and L. Cantù. "Reflectivity from Floating Bilayers: Can We Keep the Structural Asymmetry?" *J. Phys. Conf. Ser.* 340 (2012), p. 012083. doi: 10.1088/1742-6596/340/1/012083.
- Rose, D. R., J. Phipps, J. Michniewicz, G. I. Birnbaum, F. R. Ahmed, A. Muir, W. F. Anderson, and S. Narang. "Crystal Structure of T4-Lysozyme Generated from Synthetic Coding DNA Expressed in Escherichia Coli". *Protein Eng. Des. Sel.* 2.4 (1988), pp. 277–282. doi: 10.1093/protein/2.4.277.
- Rosen, M. J. and J. T. Kunjappu. *Surfactants and Interfacial Phenomena*. 4th ed. Hoboken, USA: John Wiley & Sons, 2012. ISBN: 978-0-470-54194-4.
- Rudd, R. E and J. Q. Broughton. "Coarse-Grained Molecular Dynamics and the Atomic Limit of Finite Elements". *Phys. Rev. B* 58.10 (1998), R5893–R5896. doi: 10.1103/PhysRevB.58.R5893.
- Sammalkorpi, M., M. Karttunen, and M. Haataja. "Ionic Surfactant Aggregates in Saline Solutions: Sodium Dodecyl Sulfate (SDS) in the Presence of Excess Sodium Chloride (NaCl) or Calcium Chloride (CaCl<sub>2</sub>)". *J. Phys. Chem. B* 113.17 (2009), pp. 5863–5870. doi: 10.1021/jp901228v.
- Sanchez-Fernandez, A., T. Arnold, A. J. Jackson, S. L. Fussell, R. K. Heenan, R. A. Campbell, and K. J. Edler. "Micellization of Alkyltrimethylammonium Bromide Surfactants in Choline Chloride:Glycerol Deep Eutectic Solvent". *Phys. Chem. Chem. Phys.* 18.48 (2016), pp. 33240–33249. doi: 10.1039/C6CP06053F.
- Sanchez-Fernandez, A., K. J. Edler, T. Arnold, D. Alba Venero, and A. J. Jackson. "Protein Conformation in Pure and Hydrated Deep Eutectic Solvents". *Phys. Chem. Chem. Phys.* 19.13 (2017), pp. 8667–8670. doi: 10.1039/C7CP00459A.
- Sanchez-Fernandez, A., O. S. Hammond, K. J. Edler, T. Arnold, J. Doutch, R. M. Dalgliesh, P. Li, K. Ma, and A. J. Jackson. "Counterion Binding Alters Surfactant Self-Assembly in Deep Eutectic Solvents". *Phys. Chem. Chem. Phys.* 20.20 (2018), pp. 13952–13961. doi: 10.1039/C8CP01008K.
- Sanchez-Fernandez, A., G. L. Moody, L. C. Murfin, T. Arnold, A. J. Jackson, S. M. King, S. E. Lewis, and K. J. Edler. "Self-Assembly and Surface Behaviour of Pure and Mixed Zwitterionic Amphiphiles in a Deep Eutectic Solvent". *Soft Matter* 14.26 (2018), pp. 5525–5536. doi: 10.1039/C8SM00755A.
- Sanders, S. A. and A. Z. Panagiotopoulos. "Micellization Behavior of Coarse Grained Surfactant Models". *J. Chem. Phys.* 132.11 (2010), p. 114902. doi: 10.1063/1.3358354.
- Sapir, L., C. B. Stanley, and D. Harries. "Properties of Polyvinylpyrrolidone in a Deep Eutectic Solvent". *J. Phys. Chem. A* 120.19 (2016), pp. 3253–3259. doi: 10.1021/acs.jpca.5b11927.
- SASfit*. URL: <https://kur.web.psi.ch/sans1/SANSsoft/sasfit.html> (Accessed 2018-11-11).

- SASview for Small Angle Scattering.* URL: <http://www.sasview.org> (Accessed 2016-10-26).
- Schmaljohann, D. "Thermo- and pH-Responsive Polymers in Drug Delivery". *Adv. Drug Deliv. Rev.* 58.15 (2006), pp. 1655–1670. doi: 10.1016/j.addr.2006.09.020.
- Schnablegger, H. and Y. Singh. *The SAXS Guide: Getting Acquainted with the Principles*. 4th ed. Graz, AT: Anton Paar GmbH, 2017.
- Schramm, L. L., E. N. Stasiuk, and D. G. Marangoni. "Surfactants and Their Applications". *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* 99 (2003), pp. 3–48. doi: 10.1039/B208499F.
- Scherstenberger, P. "Contrast and Contrast Variation in Neutron, x-Ray and Light Scattering". In: *Neutron, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*. Ed. by Th. Zemb and P. Lindner. Amsterdam, NL: Elsevier Science B.V., 2002, pp. 145–170. ISBN: 978-0-08-093013-8.
- Scoppola, E. and E. Schneck. "Combining Scattering and Computer Simulation for the Study of Biomolecular Soft Interfaces". *Curr. Opin. Colloid Interface Sci.* 37 (2018), pp. 88–100. doi: 10.1016/j.cocis.2018.06.008.
- Segala, E., D. Guo, R. K. Y. Cheng, A. Bortolato, F. Deflorian, A. S. Doré, J. C. Errey, L. H. Heitman, A. P. IJzerman, F. H. Marshall, and R. M. Cooke. "Controlling the Dissociation of Ligands from the Adenosine A<sub>2A</sub> Receptor through Modulation of Salt Bridge Strength". *J. Med. Chem.* 59.13 (2016), pp. 6470–6479. doi: 10.1021/acs.jmedchem.6b00653.
- Sengupta, U., M. Carballo-Pacheco, and B. Strodel. "Automated Markov State Models for Molecular Dynamics Simulations of Aggregation and Self-Assembly". *J. Chem. Phys.* 150.11 (2019), p. 115101. doi: 10.1063/1.5083915.
- Shapiro, S. S. and M. B. Wilk. "An Analysis of Variance Test for Normality (Complete Samples)". *Biometrika* 52.3-4 (1965), pp. 591–611. doi: 10.1093/biomet/52.3-4.591.
- Shelley, J. C. and M. Y. Shelley. "Computer Simulation of Surfactant Solutions". *Curr. Opin. Colloid Interface Sci.* 5.1-2 (2000), pp. 101–110. doi: 10.1016/S1359-0294(00)00042-X.
- Shi, Y. and R. Eberhart. "A Modified Particle Swarm Optimizer". In: *1998 IEEE International Conference on Evolutionary Computation Proceedings*. IEEE World Congress on Computational Intelligence. Anchorage, US: IEEE, 1998, pp. 69–73. doi: 10.1109/ICEC.1998.699146.
- Simons, K. and D. Toomre. "Lipid Rafts and Signal Transduction". *Nat. Rev. Mol. Cell Biol.* 1 (2000), pp. 31–39. doi: 10.1038/35036052.
- Sivia, D. S. *Elementary Scattering Theory: For X-Ray and Neutron Users*. Oxford, UK: Oxford University Press, 2011. ISBN: 978-0-19-922868-3.

- Sivia, D. S., W. I. F. David, K. S. Knight, and S. F. Gull. "An Introduction to Bayesian Model Selection". *Physica D* 66.1-2 (1993), pp. 234–242. doi: 10.1016/0167-2789(93)90241-R.
- Sivia, D. S., W. A. Hamilton, and G. S. Smith. "Analysis of Neutron Reflectivity Data: Maximum Entropy, Bayesian Spectral Analysis and Speckle Holography". *Physica B* 173.1-2 (1991), pp. 121–138. doi: 10.1016/0921-4526(91)90042-D.
- Sivia, D. S. and J. Skilling. *Data Analysis: A Bayesian Tutorial*. 2nd ed. Oxford: Oxford University Press, 2006. ISBN: 978-0-19-856832-2.
- Sivia, D. S. and J. R. P. Webster. "The Bayesian Approach to Reflectivity Data". *Physica B* 248.1-4 (1998), pp. 327–337. doi: 10.1016/S0921-4526(98)00259-2.
- Skou, S., R. E. Gillilan, and N. Ando. "Synchrotron-Based Small-Angle X-Ray Scattering of Proteins in Solution". *Nat. Protoc.* 9.7 (2014), pp. 1727–1739. doi: 10.1038/nprot.2014.116.
- Small, D. "Lateral Chain Packing in Lipids and Membranes". *J. Lipids Res.* 25 (1984), pp. 1490–1500.
- Smith, E. L., A. P. Abbott, and K. S. Ryder. "Deep Eutectic Solvents (DESs) and Their Applications". *Chem. Rev.* 114.21 (2014), pp. 11060–11082. doi: 10.1021/cr300162p.
- Smith, W., C. W. Yong, and P. M. Rodger. "DL\_POLY: Application to Molecular Simulation". *Mol. Simulat.* 28.5 (2002), pp. 385–471. doi: 10.1080/08927020290018769.
- Soper, A. K. and K. J. Edler. "Coarse-Grained Empirical Potential Structure Refinement: Application to a Reverse Aqueous Micelle". *BBA - General Subjects* 1861.6 (2017), pp. 1652–1660. doi: 10.1016/j.bbagen.2017.02.028.
- Source, Open. *The MIT License*. 2019. url: <https://opensource.org/licenses/MIT>.
- Storn, R. and K. Price. "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces". *J. Global Optim.* 11 (1997), pp. 341–359. doi: 10.1023/A:1008202821328.
- Sun, P., H. Sun, W. Feng, Q. Zhao, and H. Zhao. "A Study of Acceleration Coefficients in Particle Swarm Optimization Algorithm Based on CPSO". In: *2010 2nd International Conference on Information Engineering and Computer Science*. 2010 2nd International Conference on Information Engineering and Computer Science (ICIECS). Wuhan, China, 2010, pp. 1–4. doi: 10.1109/ICIECS.2010.5677729.
- Sun, W.-J., R. M. Suter, M. A. Knewton, C. R. Worthington, S. Tristram-Nagle, R. Zhang, and J. F. Nagle. "Order and Disorder in Fully Hydrated Unoriented Bilayers of Gel-Phase Dipalmitoylphosphatidylcholine". *Phys. Rev. E* 49.5 (1994), pp. 4665–4676. doi: 10.1103/PhysRevE.49.4665.
- Svergun, D. I. "Solution Scattering from Biopolymers: Advanced Contrast-Variation Data Analysis". *Acta Crystallogr. A* 50.3 (1994), pp. 391–402. doi: 10.1107/S0108767393013492.

- Svergun, D., C. Barberato, and M. H. J. Koch. "CRYSTAL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates". *J. Appl. Crystallogr.* 28.6 (1995), pp. 768–773. doi: 10.1107/S0021889895007047.
- Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson. "A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters". *J. Chem. Phys.* 76.1 (1982), pp. 637–649. doi: 10.1063/1.442716.
- Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. 2nd ed. New York: John Wiley & Sons, 1980. ISBN: 978-0-471-04893-0.
- Tarek, M., S. Bandyopadhyay, and M. L. Klein. "Molecular Dynamics Studies of Aqueous Surfactants Systems". *J. Mol. Liq.* 78.1-2 (1998), pp. 1–6. doi: 10.1016/S0167-7322(98)00078-6.
- Tarmizi, R. A. and J. Sweller. "Guidance During Mathematical Problem Solving". *J. Educ. Psycho.* 80.4 (1988), pp. 424–436. doi: 10.1037/0022-0663.80.4.424.
- Tieleman, D. P., J. L. MacCallum, W. L. Ash, C. Kandt, Z. Xu, and L. Monticelli. "Membrane Protein Simulations with a United-Atom Lipid and All-Atom Protein Model: Lipid–Protein Interactions, Side Chain Transfer Free Energies and Model Proteins". *J. Phys. Condens. Matter* 18.28 (2006), S1221–S1234. doi: 10.1088/0953-8984/18/28/S07.
- Ulyanenkov, A., K. Omote, and J. Harada. "The Genetic Algorithm: Refinement of X-Ray Reflectivity Data from Multilayers and Thin Films". *Physica B* 283 (2000), pp. 237–241. doi: 10.1016/S0921-4526(99)01972-9.
- Ulyanenkov, A. and S. Sobolewski. "Extended Genetic Algorithm: Application to x-Ray Analysis". *J. Phys. D* 38 (10A 2005), A235–A238. doi: 10.1088/0022-3727/38/10A/046.
- Uusitalo, J. J., H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, and S.J. Marrink. "Martini Coarse-Grained Force Field: Extension to DNA". *J. Chem. Theory Comput.* 11.8 (2015), pp. 3932–3945. doi: 10.1021/acs.jctc.5b00286.
- Vaknin, D., K. Kjaer, J. Als-Nielsen, and M. Lösche. "Structural Properties of Phosphatidylcholine in a Monolayer at the Air/Water Interface". *Biophys. J.* 59.6 (1991), pp. 1325–1332. doi: 10.1016/S0006-3495(91)82347-5.
- Van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. "GROMACS: Fast, Flexible, and Free". *J. Comput. Chem.* 26.16 (2005), pp. 1701–1718. doi: 10.1002/jcc.20291.
- Van Osch, D. J. G. P., L. F. Zubeir, A. van den Bruinhorst, M. A. A. Rocha, and M. C. Kroon. "Hydrophobic Deep Eutectic Solvents as Water-Immiscible Extractants". *Green Chem.* 17.9 (2015), pp. 4518–4521. doi: 10.1039/C5GC01451D.

- Vishnyakov, A., M.-T. Lee, and A. V. Neimark. "Prediction of the Critical Micelle Concentration of Nonionic Surfactants by Dissipative Particle Dynamics Simulations". *J. Phys. Chem. Lett.* 4.5 (2013), pp. 797–802. doi: 10.1021/jz400066k.
- Von Hansen, Y., S. Gekle, and R. R. Netz. "Anomalous Anisotropic Diffusion Dynamics of Hydration Water at Lipid Membranes". *Phys. Rev. Lett.* 111.11 (2013), pp. 118103–118105. doi: 10.1103/PhysRevLett.111.118103.
- Wang, S. and R. G. Larson. "Coarse-Grained Molecular Dynamics Simulation of Self-Assembly and Surface Adsorption of Ionic Surfactants Using an Implicit Water Model". *Langmuir* 31.4 (2015), pp. 1262–1271. doi: 10.1021/la503700c.
- Watkins, E. B., C. E. Miller, D. J. Mulder, T. L. Kuhl, and J. Majewski. "Structure and Orientational Texture of Self-Organizing Lipid Bilayers". *Phys. Rev. Lett.* 102.23 (2009), pp. 238101–238104. doi: 10.1103/PhysRevLett.102.238101.
- Watson, M. C. and J. E. Curtis. "Rapid and Accurate Calculation of Small-Angle Scattering Profiles Using the Golden Ratio". *J. Appl. Crystallogr.* 46.4 (2013), pp. 1171–1177. doi: 10.1107/S002188981301666X.
- Weinbach, S. P., K. Kjaer, J. Als-Nielsen, M. Lahav, and L. Leiserowitz. "Self-Assembled Langmuir Monolayers and Trilayers at the Air-Formamide Interface". *J. Phys. Chem.* 97.20 (1993), pp. 5200–5203. doi: 10.1021/j100122a003.
- Wiley, D. *Open Content - Definition*. URL: <http://opencontent.org/definition> (Accessed 2018-11-19).
- Willis, B. T. M. and C. J. Carlile. *Experimental Neutron Scattering*. Oxford, UK: Oxford University Press, 2009. ISBN: 978-0-19-967377-3.
- Wojciechowski, K., M. Orczyk, T. Gutberlet, G. Brezesinski, T. Geue, and P. Fontaine. "On the Interaction between Digitonin and Cholesterol in Langmuir Monolayers". *Langmuir* 32.35 (2016), pp. 9064–9073. doi: 10.1021/acs.langmuir.6b01737.
- Wojciechowski, K., M. Orczyk, T. Gutberlet, and T. Geue. "Complexation of Phospholipids and Cholesterol by Triterpenic Saponins in Bulk and in Monolayers". *BBA - Biomembranes* 1858.2 (2016), pp. 363–373. doi: 10.1016/j.bbamem.2015.12.001.
- Wormington, M., C. Panaccione, K. M. Matney, and D. K. Bowen. "Characterization of Structures from X-Ray Scattering Data Using Genetic Algorithms". *Philos. Trans. R. Soc. London Ser. A* 357.1761 (1999), pp. 2827–2848. doi: 10.1098/rsta.1999.0469.
- Wu, G., R. Mallipeddi, and P. N. Suganthan. "Ensemble Strategies for Population-Based Optimization Algorithms – A Survey". *Swarm Evol. Comput.* 44 (2019), pp. 695–711. doi: 10.1016/j.swevo.2018.08.015.
- Xue, M., L. Cheng, I. Faustino, W. Guo, and S. J. Marrink. "Molecular Mechanism of Lipid Nanodisk Formation by Styrene-Maleic

- Acid Copolymers". *Biophys. J.* 115.3 (2018), pp. 494–502. doi: 10.1016/j.bpj.2018.06.018.
- Yeh, I.-C. and M. L. Berkowitz. "Ewald Summation for Systems with Slab Geometry". *J. Chem. Phys.* 111.7 (1999), pp. 3155–3162. doi: 10.1063/1.479595.
- Yesylevskyy, S. O., L. V. Schäfer, D. Sengupta, and S. J. Marrink. "Polarizable Water Model for the Coarse-Grained MARTINI Force Field". *PLoS Comput. Biol.* 6.6 (2010). Ed. by M. Levitt, e1000810. doi: 10.1371/journal.pcbi.1000810.
- Zagnoni, M. "Miniaturised Technologies for the Development of Artificial Lipid Bilayer Systems". *Lab on a Chip* 12.6 (2012), p. 1026. doi: 10.1039/c2lc20991h.
- Zahn, S., B. Kirchner, and D. Mollenhauer. "Charge Spreading in Deep Eutectic Solvents". *ChemPhysChem* 17.21 (2016), pp. 3354–3358. doi: 10.1002/cphc.201600348.
- Zhou, X.-L. and S.-H. Chen. "Theoretical Fundation of X-Ray and Neutron Reflectometry". *Phys. Rep.* 257 (1995), pp. 223–348. doi: 10.1016/0370-1573(94)00110-O.
- Zhou, Xiao-Lin and Sow-Hsin Chen. "Model-Independent Method for Reconstruction of Scattering-Length-Density Profiles Using Neutron or x-Ray Reflectivity Data". *Physical Review E* 47.5 (1993), pp. 3174–3190. doi: 10.1103/PhysRevE.47.3174.