



## Review article

## Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning



Shakhawath Hossain<sup>a,b,1</sup>, Aleksei Kabelev<sup>a,1</sup>, Albin Parrow<sup>a</sup>, Christel A.S. Bergström<sup>a,b</sup>,  
Per Larsson<sup>a,b,\*</sup>

<sup>a</sup> Department of Pharmacy, Uppsala Biomedical Center, Uppsala University, 751 23 Uppsala, Sweden

<sup>b</sup> Swedish Drug Delivery Forum (SDDF), Uppsala University, Sweden

## A B S T R A C T

In this review we will discuss how computational methods, and in particular classical molecular dynamics simulations, can be used to calculate solubility of pharmaceutically relevant molecules and systems. To the extent possible, we focus on the non-technical details of these calculations, and try to show also the added value of a more thorough and detailed understanding of the solubilization process obtained by using computational simulations. Although the main focus is on classical molecular dynamics simulations, we also provide the reader with some insights into other computational techniques, such as the COSMO-method, and also discuss Flory-Huggins theory and solubility parameters. We hope that this review will serve as a valuable starting point for any pharmaceutical researcher, who has not yet fully explored the possibilities offered by computational approaches to solubility calculations.

## 1. Introduction

Understanding drug solubility is key to the modern drug discovery process. Achieving a sufficiently high solubility for a particular compound is absolutely crucial for processes as widely different as purification, formulation, production, drug absorption and disposition [1]. Poor solubility has many adverse effects such as reduced drug efficiency and erratic or low absorption, and may cause unwanted side effects or lack of therapeutic effect [2]. Because of this, much effort has been invested in developing tools for solubility predictions [3]. Virtual screening of solubility and solubility-related processes (e.g. dissolution, solubilization, supersaturation, precipitation) allows these properties to be evaluated before compound synthesis and is a valuable tool for computational chemists during compound library design. Computational methods that are already successfully used to calculate solubility include classical molecular simulation for solvation free energies [4,5], while partition coefficient calculations (log P) are used to predict solubility ratios or to provide a molecular understanding of the compound compartmentalization after solubilization [6]. Computational methods, in theory, allow a detailed understanding not only of the solubility itself, but also of the different factors and atomic level interactions that are important for the solubilization process [7–9]. In a typical pharmaceutical laboratory, however, solubility continues to be something that is determined experimentally rather than calculated or predicted, despite the feasibility of using computational methods for at least

certain aspects of it, sometimes also as a more cost-efficient technique, and one which might also provide additional insights into solubilization and partitioning.

Thermodynamic equilibrium solubility is defined as the concentration at which a solid compound is in equilibrium with itself in solution [10]. Solubilization, the process leading to solvation of a solute in some solvent, is essentially a two-stage process. Solubilization requires bonds between solute molecules in their solid form to be broken, followed by formation of cavities in the solvent in which solute molecules insert themselves and interact with the surrounding solvent. Importantly, the properties of the solid solute form remain the same irrespective of the solvent in which the solubilization occurs—provided that no solvates are formed during the dissolution process. Several of the simulation methods described below make it possible to calculate relative solubility in different solvents [11], although more specialized techniques exist that also allow the solid properties to be studied [12–14]. It is worth emphasizing that even though the solid properties are often more difficult to study with simulation models, knowledge of the relative solubility of a compound in a solvent is highly warranted in itself. Knowledge of the relative solubility—combined with calculated solvation free energies and experimental knowledge of the solubility in one reference environment—pave the way to calculating absolute solubilities in any new solvent [15].

Several classes of techniques have been developed in an attempt to create a universal simulation protocol capable of solubility prediction.

\* Corresponding author at: Department of Pharmacy, Uppsala Biomedical Center, Uppsala University, 751 23 Uppsala, Sweden.

E-mail address: [per.larsson@farmaci.uu.se](mailto:per.larsson@farmaci.uu.se) (P. Larsson).

<sup>1</sup> Equal contribution.

<https://doi.org/10.1016/j.ejpb.2019.02.007>

Received 30 October 2018; Received in revised form 5 February 2019; Accepted 13 February 2019

Available online 14 February 2019

0939-6411/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Most of these, such as empirical methods exploiting experimental data to analyze the contributions of specific functional groups to solubility [16,17], and machine-learning based methods requiring training on solubility data [18,19], are beyond of the scope of this review. Rather, our purpose here is to provide the non-experienced user an understanding of when it is possible to use molecular dynamics simulations instead of experiments to calculate solubility.

## 2. Brief introduction to molecular dynamics simulation

Since its introduction more than 40 years ago [20], molecular dynamics (MD) simulation has been used in many different research fields, and in some cases, it has become yet another tool with which a particular problem is studied. In this section we briefly summarize the main ideas behind classical MD simulations, and then present the various techniques and algorithms which the non-experienced user in a pharmaceutical laboratory can use to understand different aspects of solubility. There are already many excellent reviews on both technical and application aspects of MD [21,22], but still, in our experience, the knowledge about what can and cannot be done with computational methods (and why that is) for studying solubility is not widespread in the pharmaceutical community. To narrow the focus further, we limit ourselves here to mainly classical, particle-based MD simulations, using any of the standard biomolecular forcefields. We also briefly compare and contrast MD with some other methods, such as Monte Carlo schemes within the grand canonical ensemble, and continuum-based models like COSMO [23,24].

In a classical description of a system composed of discrete molecules, individual atoms move because they are experiencing forces exerted on them by all of the other atoms and molecules in the system. A critical requirement for using molecular simulation to study solubility is that the inter- and intramolecular forces in a particular system of interest can be modeled by a so-called force field, which provides the fundamental physics of the system. Within a particular force field, atoms and molecules interact primarily with each other through van der Waals' and electrostatic forces. Additionally, within each molecule, the atoms that are covalently bound to one another are held together by strong bonded forces. In a typical biomolecular force field, all of these forces are calculated on a per-atom basis and summed up. Depending on the combination of molecule and force fields, deriving suitable interaction parameters (through the process of parameterization) for a novel molecule can be more or less straightforward, but in all cases is a process that requires careful validation. Different force field will differ in the philosophy behind the parameterization of the force field (including what sort of validation work becomes necessary for a novel molecule of interest), i.e. the determination of default values for bond lengths, van der Waals' parameters and partial charges, to give a few examples.

The force field is what defines the physical interactions between atoms and molecules in the simulation, or computational experiment. The other major component of the experiment makes use of the forces that act between the atoms to make them move. This is achieved using numerical methods to solve Newton's equations of motion. The forces are obtained in practice by taking the gradient of the interaction potentials involving each atom. With this procedure, it is possible to obtain the position of each of the atoms in the system at a small time-step later; from these new positions, new forces are again calculated and used to update positions a second time, and so on. The time-evolution of the entire system can then be tracked by connecting each state to the previous one. There are a number of software packages that can be used for these types of simulations, each with its merits and limitations. The more common ones include Chemistry at HARvard Molecular Mechanics (CHARMM) [25,26], Amber [27], NAMD [28,29], GROMINGEN Machine for Chemical Simulations (Gromacs) [30,31], ACEMD [32], and Desmond [33].

The connection between a discrete molecular description of a

particular system (such as the study of a particular single drug molecule solvated in a surrounding solvent), and macroscopic observations on the same system in a traditional experimental setting is provided by the branch of physics called statistical mechanics. While the details of this connection are beyond the scope of this review, we will discuss it briefly on a case-by-case base in the coming sections. We will describe a number of ways in which classical simulation methods have been used to calculate and predict solubility, and to what extent these results agree with experimental solubility measurements.

## 3. Solubility

In what follows, we are primarily concerned with solubility of a solute in a solvent in the so-called infinite-dilution or solubility limit, i.e. the solvent is in such excess that the properties of the solute do not change with a change in the amount of solvent. Working at this limit computationally means that the solubility process can be simulated without considering any solute-solute interactions. Typically what is studied is either the transfer of a solute between: (i) a gas and liquid phase (free energy of solvation); (ii) gas and water phases (free energy of hydration); or (iii) two liquid phases. In terms of free energy, solubility can be expressed as the change in e.g. Gibbs free energy of transfer of solute *i* between two phases,  $\alpha$  and  $\beta$ , using the equation:

$$\Delta G_i^{\alpha\beta}(T, p, X^\alpha, X^\beta) = -RT \ln K_i^{\alpha\beta} \quad (1)$$

where *X* is the composition at equilibrium, *R* the gas constant, *p* pressure, and *T* temperature. The partition coefficient *K* is defined as the ratio of the number densities of solute *i* and phases  $\alpha$  and  $\beta$ , respectively. This is also what is typically measured experimentally in solubility determinations. Naively, this would suggest that the easiest way to compare an experimentally determined solubility with a simulation is to compare the values of *K* obtained from the two methods. However, a direct calculation of *K* from simulations is prohibitive in most cases, for various reasons; for example, it may not be possible to properly sample the solute in the two different phases to acquire enough statistics for conclusions to be reliably drawn. Different techniques have therefore been developed to calculate the solubility in terms of the transfer free energy ( $\Delta G_i^{\alpha\beta}$ ) instead. Equilibrium solubility requires the chemical potential  $\mu$  of all species to be the same, in addition to the pressure and temperature [34,35]. Through Henry's law, the solubility is related to the excess, or residual, chemical potential  $\mu_i^E$  [36], and in the limit of infinite dilution equals the free energy calculated from computational simulations. The word excess refers to the difference in chemical potential between that of a real and ideal system at the same temperature and pressure [37].

## 4. Free energy calculations

Four possible schemes are often presented for calculating free energy differences from simulations. These are: (i) methods based on histograms [38]; (ii) non-equilibrium simulations [39]; (iii) perturbation theory [40,41]; and (iv) using derivatives of the potential [42]. As free energy is a path-independent state variable, one can design simulations in a computer experiment that are a convenient, albeit somewhat artificial, way to compute the free energy. This computer approach simplifies calculations and allows one to circumvent the possibly difficult task of computationally simulating the corresponding real chemical process. The unphysical nature of such simulated transitions is why some of these free energy calculation methods are referred to as alchemical. For details we refer the reader to any of the more in-depth texts on the matter [43–46], but as a point of entry we note that the difference in free energy, here denoted as  $\Delta F$ , between two states *i* and *j*, can be expressed in terms of the so-called partition function *Z* as in Eq. (2):

$$\Delta F_{ij} = -k_B T \ln \left( \frac{Z_j}{Z_i} \right) \quad (2)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature, and  $Z_i$  and  $Z_j$  are the statistical mechanics partition functions of the two states of interest. While in principle these partition functions would provide a complete description of a system, often they are not known, which is the reason that absolute free energies of real systems are difficult to obtain from computer simulations. In terms of drug solubility these two states could, for example, correspond to an active pharmaceutical molecule being present or absent from a simulation box filled with solvent molecules.

#### 4.1. Free energy perturbation calculations

Originally proposed by Zwanzig in 1954 [40], the free energy perturbation (FEP) method can be used to describe the free energy difference between two different systems. These systems correspond to the initial system  $i$ , which is unperturbed, and  $j$ , which is called the perturbed system in relation to system  $i$ . The perturbed system may for example include an additional particle or a different energy function, as would be the case when using FEP to calculate solvation free energies. The free energy difference between these two states is then related according to Eq. (3):

$$\Delta F_{ij} = -k_B T \ln \left( e^{-\frac{E_j - E_i}{k_B T}} \right) \quad (3)$$

where  $k_B$  is Boltzmann's constant,  $T$  is temperature, and  $E_i$  and  $E_j$  are the energies associated with two different states of the system (Fig. 1).

The FEP method requires that the difference in energy between states  $i$  and  $j$  is sufficiently small, in order for the calculations to converge, and to provide a meaningful result. As a result, in practice it is normally necessary to divide a FEP calculation into a series of individual simulations in a piecewise manner through a coupling parameter  $\lambda$  (Fig. 1), and these simulations are then computed independently [47,48]. Luckily, communication between such simulations is not necessary, so the process can be trivially parallelized by running each part on a different processor/node, making calculations more efficient. Another situation in which FEP can be beneficial is when the final state's phase space is a subset of the reference state [49,50].

#### 4.2. Thermodynamic integration

An alternative to evaluating the difference in the free energy between subsequent states is to calculate the derivative of the total energy (the so-called Hamiltonian, including both potential and kinetic) in the system. Thermodynamic integration (TI) was originally proposed by Kirkwood in 1935 and is still one of the most widely used free energy calculation approaches. The TI method covers a transition between an initial and a final state, and similarly to FEP uses a coupling parameter

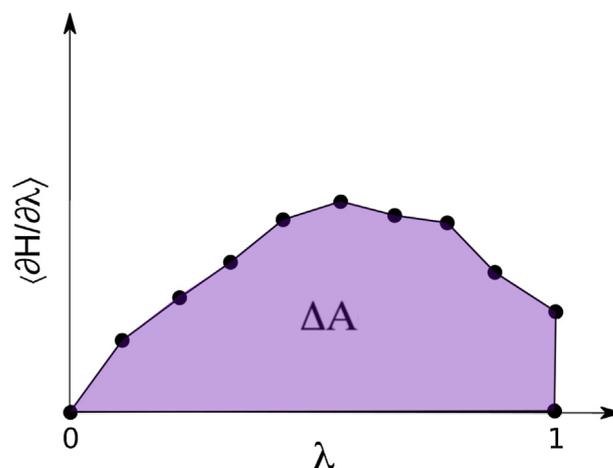


Fig. 2. Schematic representation of the thermodynamic integration (TI) method. Here  $H$  stands for the Hamiltonian of the system in a particular moment,  $\partial H / \partial \lambda$  is its derivative with respect to  $\lambda$  (the coupling parameter, describing the extent to which the system is in state 0 or 1, and  $\Delta A$ , the area under the curve, is the free energy difference between the initial and final states.

$\lambda$  to describe the transition between these two points. By introducing several intermediate  $\lambda$  values between 0 and 1, the energy function as a function of  $\lambda$  would be defined as in Eq. (4):

$$U(\lambda) = U_i + \lambda(U_f - U_i) \quad (4)$$

The potential energy  $U(\lambda)$  in each  $\lambda$ -state can be calculated as an ensemble average over configurations sampled from a MD simulation. The change in free energy between the initial and final states 0 and 1 can then be computed from the integral of the ensemble-averaged derivatives of potential energy over the coupling parameter  $\lambda$  in Eq. (5) [22]:

$$\Delta F(i \rightarrow f) = \int_0^1 \langle \frac{\delta U(\lambda)}{\delta \lambda} \rangle d\lambda \quad (5)$$

where  $\Delta F$  is the free energy and  $U(\lambda)$  the potential energy expressed as a function of the coupling parameter  $\lambda$  (Fig. 2).

#### 4.3. Applications

These free energy methods have been used extensively for calculation of solvation and hydration free energies as well as chemical potential calculations in different systems, and what will follow is not meant to be an exhaustive enumeration of all such work [51–55]. As a first pharmaceutically relevant example, Lüder et al. [56] used FEP to calculate the free energy of solvation of 46 different drugs in pure melt systems, and reached an average error of 1.8 kJ/mol over their entire dataset. Further, using TI, the hydration free energies for some polyphenol compounds (a class of bioactive natural compounds with pharmaceutically interesting properties) such as toluene and

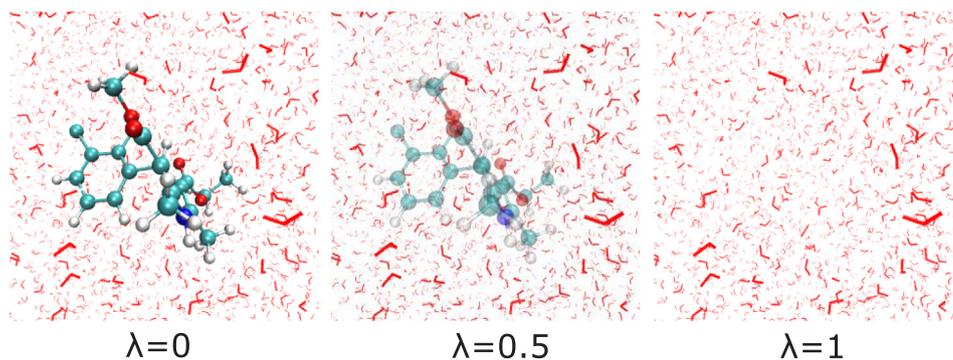


Fig. 1. Free energy perturbation calculation in which a drug molecule (felodipine) is being gradually decoupled from the surrounding solvent molecules. The coupling parameter  $\lambda$ , describes the gradual (but artificial) de-coupling of the system in a step-wise process going from  $\lambda = 0$  (felodipine is fully interacting with its surroundings) to  $\lambda = 1$ , in which all drug-solvent interactions are set to zero.

(+)-catechin, was calculated at different temperatures in a recent study by Gillet et al. [57]. Even though the actual hydration free energies differed ( $-47.78$  kJ/mol calculated for toluene vs  $-3.22$  kJ/mol as the measured value, a difference the authors attribute to differences in pressure and the water model they used), the obtained relative free energies at different temperatures showed consistent trends. The authors were also able to fit exponential functions to their results describing how solubility was moderated by temperature. These fitted functions had parameters that was found to be dependent on the properties of the solid solutes.

In a series of papers [56,58,59], Kjellander et al. studied the efficiency of FEP methods for the calculation of transfer free energies [58], first for drug molecules in TIP4P water using the OPLS-AA [60] force field, and then, as mentioned above, followed up by free energy of solvation in pure melts [56], and finally in amorphous matter [59]. In their simulations, the Coulomb and Lennard-Jones (LJ) interactions were scaled down independently. Interestingly, the authors also complemented the use of FEP with analysis based on using molecular energies and molecular surface areas to further understand the separate stages (cavity formation, solute insertion, dispersion interactions between solute and solvent) of the solubilization process.

Another study using FEP, work of Paluch et al. [61] illustrates how one can express the equilibrium mole fraction solubility via solvation free energy calculations and fugacity. As also suggested by Mobley and co-workers [11], the relative solubility of an active pharmaceutical ingredient in different solvents can be calculated from chemical potentials. These calculations had a root-mean-squared deviation (RMSD) of between 1.17 and 3.46 between calculated and experimental relative solubility values, depending on the method used. The advantage with such an approach is that the contribution to solubility from the fugacity of the solute in its solid form cancels out when the relative difference for the same solute in two different solvents is sought. The work from Mobley et al. shows that results from particle-based simulations agree better with experimental data than those from other methods like UNIFAC [16] (a semi-empirical method). The same approach was also used to calculate the excess solubility (solubility in an actual solution relative to what would be observed if the solvents mixed to form an ideal solution) [62] for acetanilide, acetaminophen, phenacetin, benzocaine and caffeine in mixtures of water and ethanol [37], with overall excellent quantitative agreement with experiment.

## 5. Widom insertion method

Widom [38] showed that the residual or excess chemical potential in the canonical ensemble, in which the number of particles in the system is kept constant, can be expressed as in Eq. (6):

$$\mu_{\text{ex}} = -k_B T \ln \langle e^{-\varphi/kT} \rangle_N \quad (6)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature of the system,  $\varphi$  is the interaction energy of a randomly placed test particle with the system of  $N$  particles, and  $\langle \dots \rangle_N$  denotes ensemble averaging over all configurations of the particles in the volume  $V$  [38,63].

Widom's test particle insertion scheme has been used to determine the solubility of small molecules in different systems with good agreement with experimental results in both cases [64,65]. However, due to the required random insertion of particles (Fig. 3), overlap often occurs between atoms of the inserted solutes and the rest of the system [66,67]. This issue becomes critical in a dense solvent or during the insertion of large molecules and can lead to errors in the calculated chemical potential (and therefore also the solubility) if adequate sampling is not performed. To alleviate this problem, Khawaja et al. [67] used a biased insertion technique, in which the molecules or particles are inserted into a pre-identified free space, and used this to determine the solubility of small molecules such as helium, carbon dioxide, and water in nitrile butadiene rubber. Their results were in good agreement

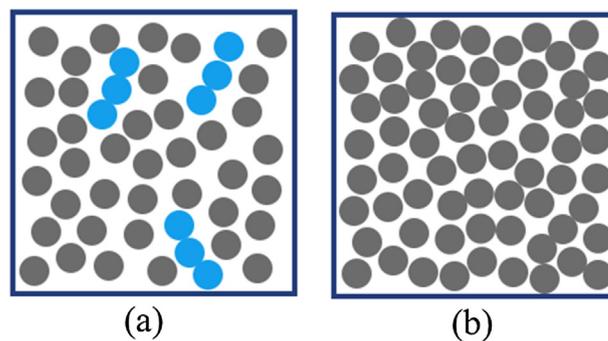


Fig. 3. Illustration of the Widom test particle insertion method. Solvent particles are gray and molecules are blue. (a) A loosely dense system in which the molecules can be inserted in a region with cavities, (b) A highly dense system in which the molecules are unable to insert by themselves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with experimental solubility values and this biased insertion technique is 40 times faster than unbiased, brute-force particle insertion. To solve the same issue for the insertion of long polymeric chains, an incremental insertion technique can be implemented in which smaller parts of the solute is added sequentially to the end of one of the chains [68]. Ferguson et al. [66] studied the solubility of alkane chains up to eicosane ( $C_{20}$ ) in water with this technique and found good agreement with the experimental results up to pentadecane ( $C_{15}$ ).

Further examples that are relevant in a pharmaceutical context is the application of the Widom insertion method to predict solubility of water in polylactide (PLA) polymers [69]. The authors obtain a value of the solubility parameter (more details below) of  $20.6 \text{ MPa}^{1/2}$ , which compares favorably to experimentally determined values of  $19\text{--}20.5 \text{ MPa}^{1/2}$  [70]. Even though the solute is water and not for example a drug molecule this showcases the usefulness of the Widom method. The solubility of water in PLA is important since it in turn can impact the stability of amorphous drug formulations, and again, advances of the insertion technique such as incremental insertion [66] makes it also potentially suitable for investigating the solubility of larger molecules in other dense media.

## 6. Solubility calculations in the grand canonical ensemble

The grand canonical ensemble (GCE) is the name of a thermodynamic system (or ensemble) in which the temperature and chemical potential of the species of interest are specified for a fixed volume, and the number of particles and energy are the variables [71]. The system reaches equilibrium when a number of successful insertion or deletion of particles balance each other. In this way, the GCE more closely mimics real systems in which the exchange of energy and particles occur in a regular manner. Note that in a typical canonical ensemble, the number of particles in a system is kept constant. Simulations in the grand canonical ensemble have been used to calculate chemical potential or to investigate fluid phase equilibria [72–74].

MD-simulations, but also Monte Carlo (MC)-methods, can be used to perform simulations in the GCE. Monte Carlo methods are arguably naturally better suited for this since the insertion or deletion of molecules in the GCE resembles the stochastic movements of molecules in MC-simulations [72]. The Monte Carlo method in the GCE has been used for calculating for example phase equilibria [73,75] and the solubility of small penetrants in some flexible chains [71]. The combination of MC and the GCE, the so-called grand canonical Monte Carlo (GCMC) scheme, is often combined with a histogram-reweighting method for phase equilibrium calculations [75,76]. Here, data on the frequency of energy and particle numbers are organized as histograms [75]. Thereafter, the free energy of a system over a range of

thermodynamic conditions is determined from a limited number of simulations. Potoff et al. [75] used GCMC to obtain the phase behavior of a variety of polar and non-polar binary mixtures and determined their phase diagrams with high accuracy. Liu et al. [77] used histogram-reweighted GCMC to obtain the phase behavior of CO<sub>2</sub>-H<sub>2</sub>O mixtures over a broad range of temperature and pressures. They found that the existing atomistic models for water or CO<sub>2</sub> cannot reproduce the experimental data over the entire temperature and pressure range which indicates the importance of improving the current models. Density-biased GCMC simulations was used by Rodgers et al. to investigate the solubility of alcohol molecules in a coarse-grained lipid bilayer model [78]. The model quantitatively determined the partitioning coefficient as a function of alcohol concentration in the system. However, the authors concluded that in order to capture the partitioning over a range of temperatures improved coarse-grained models are required.

Similarly to the Widom insertion method, application of GCMC to high-density systems is difficult mainly due to the substantially lower probability of finding a large enough cavity for particle insertion or sufficiently high energy configurations to delete molecules [72,74,79]. Mezei et al. [80] developed a cavity-based GCMC scheme which was later modified by Yau et al. [71] This works well for moderately dense systems, but the success rate of insertion or deletion of molecules is still somewhat low.

Instead of using MC, grand canonical ensemble molecular dynamics simulation (GCMD) techniques have also been developed based on an extended Hamiltonian system in which the physical system is linked to a particle bath to represent the particle number as a continuous dynamic extension variable [81–83]. This was then successfully used to determine the solubility of gases [79] and water [74] in different polymers. An approach that uses both the GCMC and MD techniques to study the solute behavior and solubility in explicit solvent aqueous systems and a solvated protein environment was presented by Lakkaraju et al. [84]. In this method,  $\mu_{ex}$  of both the solute and the water systematically oscillated over the GCMC-MD iterations to drive the solute and water energy and particle exchanges. The method is capable of approximating the hydration free energy of individual organic solutes, dilute aqueous mixtures of multiple solutes, and binding affinity of solutes to the protein.

Due to the conceptual appeal to mimic systems in a realistic way, the grand canonical ensemble can be a suitable technique to calculate solubility of relevant molecules in pharmaceutical research, in particular when combined with adaptive resolution simulation schemes [85], which allows one to focus on a specific region in space at high resolution, while the rest of the system can be treated at lower resolution [85,86]. This multi-scale approach can improve the computational efficiency when simulating large-scale systems.

## 7. Free energy methods for calculation of molecule partitioning

The partition coefficient,  $P$ , is the ratio of the concentration of a neutral or unionized solute molecule in a system of two immiscible solvents (Fig. 4). It is typically expressed as the logarithm of this concentration ratio, as in Eq. (7).

$$\log P = \log \left( \frac{[\text{solute}]_{\text{organic}}}{[\text{solute}]_{\text{water}}} \right) \quad (7)$$

In drug design strategies, a correct understanding of the interactions of a particular solute in both organic (lipophilic) and water (hydrophilic) phases is necessary [87], and can sometimes be just as critical as knowledge of the actual solubility. Therefore, the partition coefficient—which describes the difference in solubility of the solute in the organic and water phases—is important; it is for example used in drug design as a measure of the hydrophobicity of a solute, membrane permeability, as well as the bioavailability of drug compounds in different environments [6,88]. When computing the partition coefficient with

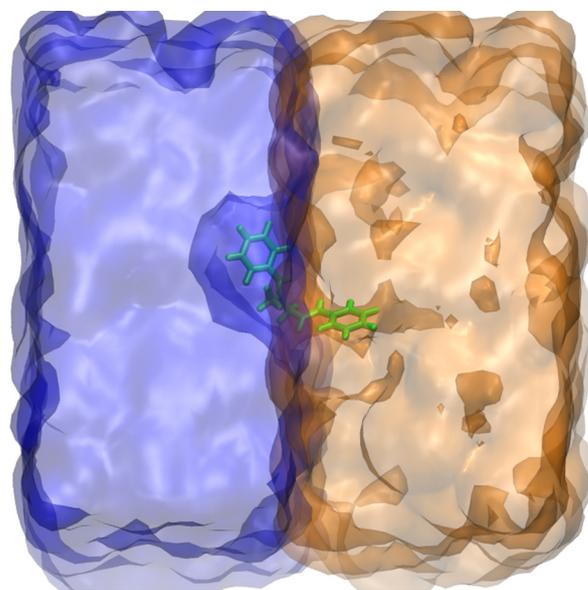


Fig. 4. Schematic representation of the partition coefficient as the equilibrium concentrations of a dissolved substance in a two-phase system with two largely immiscible solvents (in this case water, blue surface, and hexadecane, orange surface). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

simulations, oils, chloroform or alkanes can be used as the organic phase, but 1-octanol is commonly used due to its amphiphilic nature (a polar head group attached to a nonpolar tail). This renders the molecule similar to the main constituents of lipid membranes and gives it the capability of mimicking different properties of a real biological system [88].

Because of the practical importance of partition coefficients, many experimental techniques have been developed to measure them [89,90]. However, these experimental measurements use actual compounds which can be costly and the synthesis of the compounds also often takes a long time [88]. Therefore, to accelerate the drug design and discovery process, many theoretical and computational models have been developed to determine the partition coefficient. Quantitative structure property relationship (QSPR) models can relate the log  $P$  value with different descriptors or molecular properties such as molecular surface, volume, weight, functional groups, dipole moment, etc. [91]. Additive fragment-based techniques and machine learning algorithms are also used to calculate the partition coefficient [6]. However, these methods vary in accuracy and efficiency and are all required to be trained on experimental data sets [6,92]. As an alternative, alchemical free-energy calculation techniques (for example using FEP or TI) with molecular dynamics simulation, can be used to calculate the partition coefficient, since it is directly proportional to the transfer free energy between two solvents [6,88,91,92]: The relationship between the partition coefficient and transfer free energy can be expressed as in Eq. (8):

$$\log P = \frac{-\Delta G_{\text{transfer}}}{2.303RT} = \frac{\Delta G_o - \Delta G_w}{2.303RT} \quad (8)$$

where  $R$  is the molar Boltzmann constant,  $T$  is the temperature, and  $\Delta G_o$  and  $\Delta G_w$  are the free energies in an organic solvent e.g. 1-octanol and water, respectively. This approach provides reliable predictions of the partition coefficient and can be applied to many combinations or organic solvent, water and solute. In addition, MD simulation provides molecular level insights about the partition mechanism of the solutes into the solvents.

A large number of different studies have calculated partition coefficients from different free energy calculation techniques. Best et al. [93] used the FEP method for some small organic molecules in a

saturated octanol-water system. The  $\log P$  values were generally in good agreement with the experimental values except for acetamide-acetone. DeBolt and Kollman [94] successfully calculated the  $\log P$  of benzene and phenol in an octanol-water system using the FEP method with modified force fields. They also investigated the structural and thermodynamic behavior of liquid octanol and postulated that octanol's capability to serve as a biophase analog is mainly due to the formation of preferentially polar versus nonpolar regions in the media.

One of the issues of using free energy methods is that the calculated free energies are relative to a reference solute, which in turn produces a partition coefficient relative to that of the reference solute [91]. Therefore, to determine the absolute partition coefficient, an experimental partition coefficient is required for the reference solute. An accurate molecular model or force field is also required for both the reference solute and solute of interest. Garrido et al. [95] found excellent agreement (average deviations of 0.2, 1.1, 0.8 and 1.2 kJ mol<sup>-1</sup>) between experimental and simulated  $\log P$  values for propane, benzene, ethanol and acetone respectively through the calculation of the free energies of solvation using MD simulation and thermodynamic integration. The absolute average deviation between the experiments and simulation was 0.28  $\log P$  units. However, this deviation could be lowered to 0.14  $\log P$  units by optimizing partial atomic charges of acetone in the water phase. Huang et al. [96] reported predicted  $\log P$  values for different solute classes including alkanes, chlorides, bromides, fluorides, compounds with sulfur, phosphorus etc. They used an implicit solvation model in which 3D-RISM [97] theory was used to calculate the solvation free energy for different solute classes including alkanes, chlorides, compounds with sulphur, phosphorous etc. They obtained a good prediction of  $\log P$  compared to the experimental values. Ogata et al. [88] calculated the octanol-water  $\log P$  for 75 compounds, including 17 drug compounds, using free energy calculation with the Bennett's acceptance ratio (BAR) [41] method. For the set of 17 drug molecules in particular, the agreement obtained between experimental and computational  $\log P$  values had an R<sup>2</sup> value of 0.86 and a mean squared error (MAE) for the solvation free energies of 2.26 kJ/mol. Using the BAR method, Wolf and Groenhof [98] calculated hydration free energies and chloroform-water partition coefficients for different nucleic acids, molecules which are similar in size and complexity to many pharmaceutical drugs. They found, at least for their dataset, that the CHARMM [25] force fields reproduced the experimental data more accurately (absolute error of 0.65  $\log P$  units) than other force fields i.e. AMBER [27], GROMOS [99], and OPLS-AA [60]. These studies clearly indicate the importance of an accurate force field or molecular model all molecules in the system before using free energy and MD simulation techniques to calculate the  $\log P$  values of the molecules.

Adding further to the plethora of different kind of computational schemes, Bhatnagar et al. used an adaptive biasing force method with MD simulations (ABF-MD) to calculate the partition coefficients of n-alkanes. One of the major advantages of this method is that it does not need reference solutes. The so-called Gibbs ensemble Monte Carlo (GEMC) is another method which can calculate the transfer free energies and determine the partition coefficients directly by simulating two phases simultaneously [100,101]. In this scheme, molecules can transfer directly between the two phases during the simulation and the free energy of transfer can be calculated directly from the number density of particles in the two phases, according to Eq. (9) [102]:

$$\Delta G_{transfer} = 2.303RT \log \left( \frac{\rho_{\alpha}}{\rho_{\beta}} \right) \quad (9)$$

where  $\rho_{\alpha}$  and  $\rho_{\beta}$  represent the number density of particles in the two  $\alpha$  and  $\beta$ , respectively. This method works well for small molecules for which transfer of solute molecules occurs between phases at a reasonable rate.

In two other studies, the partition coefficient for 150 [103] and 650 [104] small compounds, respectively, was also predicted by free energy

calculation performed with coarse-grained MD (CGMD) simulations using the MARTINI [105] force-field. CGMD reduces the simulation time compared to all-atom MD and was found capable of reproducing the experimental octanol-water and hexane-water partition coefficients (in one of the studies [103] the deviations was between 0.67 and 0.90  $\log P$  units). Some discrepancies for the computed solvation free energies (mean absolute deviations larger than 10 kJ/mol [103]) were mainly attributed to the limited fluid range of the Lennard-Jones potential in the MARTINI coarse grained force field. Implementation of a hybrid all-atom/coarse-grained model unfortunately failed to improve the discrepancy for the estimation of solvation free energy [103].

It should be noted that to estimate the transfer free energy,  $\Delta G_{transfer}$  in Eq. (9), the solvents of the system are generally assumed to be completely immiscible and the solvation free energy of each solute is calculated separately. However, in theory, very polar solutes can carry water molecules to the organic phase [6], complicating the picture. Furthermore, as noted in the introduction, during free energy calculations only one solute molecule is present in the solvent, so that the calculation can be considered to take place at infinite dilution. During experiments, significant deviation of the solute concentration from infinite dilution can lead to differences between experimental and computational values. Other errors inherent to different experimental methods can be another reason for the discrepancies in some studies between the measurements and simulations. However, from the studies discussed above, it is evident that free energy methods with suitable force fields and enough sampling can predict partition coefficients accurately and efficiently; this is advantageous for rapid screening of large numbers of drug compounds.

Switching gears a bit, yet another method to calculate  $\log P$  is the Conductor-like Screening Model for Realistic Solvation (COSMO-RS) [23]; which extended to micelles and membranes is called COSMOmic [24]. This latter approach allows efficient prediction of partition coefficients of molecules in micelles and lipid bilayers. The details regarding COSMO-RS and COSMOmic can be found in [23,24,106]. Combinations of COSMOmic and MD methods [106] have been used to calculate the  $\log P$  of cytochrome P450 substrates and their metabolites [107], 4-ethylphenol, propanol, 5-phenylvaleric acid, and dibenz[a,h]anthracene [108], and other drug-like molecules [109] in various model membranes: dioleoylphosphatidylcholine (DOPC), palmitoyloleoyl-phosphatidylglycerol (POPG), 1-stearoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (SOPC), and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC). Paloncyova et al. [112] also compared the partition coefficients and thermodynamic properties obtained by COSMOmic calculation and MD simulation and there was good agreement between the two methods. The COSMOmic calculation is faster than the MD simulations. However, MD simulations can be more informative for estimating the membrane localization of drug-like molecules and for analyzing structural features of complex systems, e.g. lipids, drugs and proteins [112].

Finally, the umbrella sampling (US) method can also be used to calculate free energy profiles (potentials of mean force), and from them  $\log P$  [110,109,112]. Paloncyova et al. [111] compared the  $\log P$  obtained using DOPC and POPG bilayers with the octanol-water  $\log P$  and found that the octanol-water  $\log P$  correlates with only some types of membranes. This suggests that the typical organic-water system  $\log P$  cannot be straightforwardly used for all types of membranes and molecules. Another study by Holmboe et al. [110] compared the influence on the partition coefficient of danazo, felodipine and carbamazepine on membrane composition, in particular the influence of bile salt insertion in membranes.

## 8. Flory-Huggins theory and solubility parameters

Initially developed as a method to study the behavior of dilute polymer solutions [111], Flory-Huggins (FH) theory has more recently been used together with MD simulations to study not only qualitative

miscibility but also solubility in a number of pharmaceutically relevant systems [112–115]. FH-theory, originally based on the concept of entropy of mixing of two differently sized polymers, gives a way of estimating the change in free energy upon mixing of different molecules. Closely connected to FH-theory is also the so-called Hansen-Hildebrand (HH) solubility parameters [116]. Initially devised by Hildebrand, and then later extended and refined by Hansen, these parameters, which can be calculated from simulations using the FH-machinery, can be used to predict the solubility of a pharmaceutical compound in a solution. The HH-solubility parameters can be defined as the square root of the so-called cohesive energy density [117] of the (amorphous) system at room temperature. The cohesive energy density corresponds to, and can be measured experimentally from, the enthalpy change upon evaporation [118]. Such measurements however can be more or less complicated depending on the possibility to evaporate a particular material [119].

Computationally, the cohesive energy density is typically calculated as the difference between the non-bonded energy in the condensed and gas phases [120,121], and from this the solubility parameter  $\delta$  can then be calculated according to Eq. (10):

$$\delta = \sqrt{E_{\text{coh}}/V} = \sqrt{E_{\text{vac}} - E_{\text{bulk}}/V} = \sqrt{CED} \quad (10)$$

where  $E_{\text{coh}}$  is the cohesive energy density,  $E_{\text{vac}}$  and  $E_{\text{bulk}}$  the potential energies of molecules in vacuum and the condensed phases respectively, and  $V$  system volume. Depending on whether the Hansen or Hildebrand solubility parameters are sought, the non-bonded energies can be further split to reflect contributions from different types of interactions. As mentioned by Choi et al. [122], care has to be taken when calculating the gas phase energies, since polymer conformations in this phase most likely differs from those in the condensed phase. A protocol for this has been published by Belmares et al. [123] among others.

Having calculated the CED, it is then possible to calculate the also the FH interaction parameter  $\chi_{FH}$ , which describes the miscibility of the system components, using Eqs. (11) and (12) [120,121].

$$\Delta E_{\text{mix}} = \phi_1 \left( \frac{E_{\text{coh}}}{V} \right)_1 + \phi_2 \left( \frac{E_{\text{coh}}}{V} \right)_2 - \left( \frac{E_{\text{coh}}}{V} \right)_{1,2} \quad (11)$$

$$\chi_{FH} = \frac{V}{RT} \Delta E_{\text{mix}} \quad (12)$$

$E_{\text{coh}}$  again the cohesive energy density as in Eq. (10),  $\phi$  refers to the molar fractions of the respective molecules in the system,  $V$  is volume,  $R$  the gas constant and  $T$  temperature.

In this way, the miscibility has been studied for a number of different systems using classical MD by many researchers. Perhaps due to its origin within polymer theory, the vast majority of these systems involve polymers in some way, and either with or without the addition of small molecules such as drugs. Studies on polymer-polymer blends include those of Yang et al. [124], who calculated the solubility parameters and miscibility of poly-3-hydroxybutyrate (PHB) and polyethylene oxide (PEO) using the Dreiding force field [125]. Their results were in agreement with literature reference values. PEO was also used in a study by Chaudari et al. [126] in which they describe the free energy change of mixing PEO in water. Prathab et al. [8] investigated interfacial interactions between polymethyl methacrylate and a number of other polymers with the COMPASS [127] forcefield to establish favorable interactions in their mixed systems. Another polymer-polymer mixing study of PEO together with polyvinyl chloride, and concluded that 70/30 or 30/70 blends are more miscible than a 50/50 blend [128]. The COMPASS forcefield was again used by de Arenaza et al. [129] to study poly-L-lactide and poly-D-lactide with polystyrene and PVP; they found that the calculated values deviated 10–20% from the measured ones. The temperature dependence on solubility can also be addressed using the cohesive energy density concept, as was done for example by Chen et al. [130], who modeled the glass–rubber transition for amorphous polymers.

Studies on systems including additional, non-polymeric molecules, include that of Gupta et al. [131] who compared experimental and computational results for solubility of indomethacin in combination with either PEO, glucose or sucrose using the COMPASS [127] force field. Xiang and Anderson [9] similarly performed simulations of indomethacin in poly-vinylpyrrolidone (PVP). Force-field parameters for these compounds were derived to be compatible with the Amber family of forcefields, and the FH-interaction parameter calculated from their simulations predicted miscibility between PVP and indomethacin at molar fractions in agreement with experimental results. Hyunh et al. [120] again used the COMPASS force field for MD simulations of solubility calculations of docetaxel, an anti-cancer agent, in a number of excipients (tricaprylin, tricaproin, tributyrin, vitamin E and beta-carophyllene). As can be shown from FH-theory. using a miscibility cutoff value of less than 0.5 for the FH interaction parameter and varying the fractions of drug and excipients, they were able to determine solubility values for their systems in agreement to within 2–6% of experiments. The combination of small molecules (4-*n*-pentyl-4-cyanobiphenyl (5CB)) and polymer polymethyl methacrylate (PMMA) was also investigated by Patnaik and Pachter [132]. Using FH theory, they concluded that while 5CB would mix with monomeric methyl methacrylate, it is not miscible with PMMA. Xiang and Anderson [133] calculated FH interaction parameters for mixtures of felodipine, hydroxypropylmethylcellulose (HPMC) and water, and the value of the FH-interaction parameter ( $-0.20 + -0.07$ ) predicted miscibility of felodipine and HPMC at all HPMC compositions. The simulations were also able to predict a disrupting effect of water on the hydrogen bonds between felodipine and HPMC.

An interesting alternative to the use of cohesive energy density for calculating solubility parameters was recently published by Callaway et al. [134] They proposed instead an interaction energy based on the molecular volume enclosed by a Connolly surface. They note that this leads to a more realistic description of the interaction energy between two molecules, which might otherwise fail to take into account the combination of molecular size and affinity strength. Wu and co-authors devised an entirely new force field (TEAM) to validate predictions of solubility parameters for polymers of different chain lengths. They found that their computed solubility parameters were progressively underestimated with increasing polymer length [135].

## 9. Equations of state parameters from molecular simulation

Another route to predict drug solubility from simulations it to combine them with an equation of state (EOS). The most prominent example of an EOS is arguably the ideal gas law, which relates pressure and volume to the absolute temperature and the number of moles of the gas. A number of studies have used such a combined EOS-MD approach to calculate the solubility of different compounds. For example, Zhong and Masuoka [136] used the Peng-Robinson EOS to calculate gas solubilities in molten polymers. Using the Sanchez-Lacombe equation, Gauter et al. [137] predicted isotherms of ethylene and n-hexane in PE. Also using the Sanchez-Lacombe equation, Kiparissides et al. [138] determined the solubility of ethylene and found good agreement with measurements. Other efforts to this end include publications by Sato et al. [139–142], Kroon et al. [143], and Lisal et al. [144]

The common theme for these studies is that they require the determination of a number of thermodynamic state variables before any modeling of, or any conclusions about, the solubility. This is something that might not be straightforward, and in some cases even involve additional modeling or parameter fitting. Computational modeling can be helpful in such cases, as many of the relevant parameters can be estimated using simulations instead. This was illustrated in a study by Kanellopoulos et al. [145] who again employed the Sanchez-Lacombe equation, but this time using simulations with the COMPASS forcefield to determine pressure, temperature and density, and hence solubility of  $\alpha$ -olefins in polyolefins. Fermeglia and Prici [146] showed, for two

different EOS, how to use molecular simulation to extract characteristic parameters for four synthetic polymers.

## 10. Conclusions

Solubility is at the heart of pharmaceutical research. Understanding it is crucial for any research project, whether within academia or in industry. The purpose of this review has been to outline the various ways in which computational modeling, in particular molecular simulation, can be used to calculate and predict solubility. We have tried to showcase examples where simulations can guide or even preempt experiments. The increase in computational power allows more and more complex systems to be studied. In particular small-molecule simulations have matured from a tool for single solute investigations into something that allows a large number of compounds to be rapidly screened for their compatibility with a particular solvent. This can be done just as efficiently as the corresponding screen in the traditional laboratory, but also faster and arguably also at a lower cost. In addition, and perhaps just as importantly, a molecular level description of solubilization might provide further insights into the underlying mechanisms and processes.

## Acknowledgments

This work has received support from the Innovative Medicines Initiative Joint Undertaking (<http://www.imi.europa.eu>) under Grant agreement no. 115369, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies

This study is also part of the science program of the Swedish Drug Delivery Forum (SDDF). Financial support from Vinnova (Dnr 2017-02690) is gratefully acknowledged, as is support from the European Research Council Grant 638965

## References

- [1] D.J. Constable, C. Jimenez-Gonzalez, R.K. Henderson, Perspective on solvent use in the pharmaceutical industry, *Org. Process Res. Dev.* 11 (2007) 133–137.
- [2] T.M. Allen, P.R. Cullis, Drug delivery systems: entering the mainstream, *Science* 303 (2004) 1818–1822.
- [3] J.S. Delaney, Predicting aqueous solubility from structure, *Drug Discov. Today* 10 (2005) 289–295.
- [4] G. Duarte Ramos Matos, et al., Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database, *J. Chem. Eng. Data* 62 (2017) 1559–1569.
- [5] M. Lundborg, E. Lindahl, Automatic GROMACS topology generation and comparisons of force fields for solvation free energy calculations, *J. Phys. Chem. B* 119 (2014) 810–823.
- [6] C.C. Bannan, G. Calabró, D.Y. Kyu, D.L. Mobley, Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water, *J. Chem. Theory Comput.* 12 (2016) 4015–4024.
- [7] K. Kim, T. Kwon, B.J. Sung, C. Kim, Effect of methane–sugar interaction on the solubility of methane in an aqueous solution, *J. Colloid Interface Sci.* 500 (2017) 113–118.
- [8] B. Prathab, V. Subramanian, T.M. Aminabhavi, Molecular dynamics simulations to investigate polymer–polymer and polymer–metal oxide interactions, *Polymer* 48 (2007) 409–416.
- [9] T.X. Xiang, B.D. Anderson, Molecular dynamics simulation of amorphous indomethacin-poly(vinylpyrrolidone) glasses: solubility and hydrogen bonding interactions, *J. Pharm. Sci.* 102 (2013) 876–891.
- [10] C.A.S. Bergström, P. Larsson, Computational prediction of drug solubility in water-based systems: qualitative and quantitative approaches used in the current drug discovery and development setting, *Int. J. Pharm.* 540 (2018) 185–193.
- [11] S. Liu, et al., Using MD simulations to calculate how solvents modulate solubility, *J. Chem. Theory Comput.* 12 (2016) 1930–1941.
- [12] M.J. Schnieders, et al., The structure, thermodynamics and solubility of organic crystals from simulation with a polarizable force field, *J. Chem. Theory Comput.* 8 (2012) 1721–1736.
- [13] E.G. Noya, M.M. Conde, C. Vega, Computing the free energy of molecular solids by the Einstein molecule approach: Ices XIII and XIV, hard-dumbbells and a patchy model of proteins, *J. Chem. Phys.* 129 (2008) 104704.
- [14] L. Li, T. Totton, D. Frenkel, Computational methodology for solubility prediction: application to the sparingly soluble solutes, *J. Chem. Phys.* 146 (2017) 214110.
- [15] L. Chebil, et al., Solubilities inferred from the combination of experiment and simulation. Case study of Quercetin in a variety of solvents, *J. Phys. Chem. B* 114 (2010) 12308–12313.
- [16] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.* 21 (1975) 1086–1099.
- [17] R. Wittig, J. Lohmann, J. Gmehling, Vapor–liquid equilibria by UNIFAC group contribution. 6. Revision and extension, *Ind. Eng. Chem. Res.* 42 (2003) 183–188.
- [18] M. Hewitt, et al., In silico prediction of aqueous solubility: the solubility challenge, *J. Chem. Inf. Model.* 49 (2009) 2572–2587.
- [19] L.D. Hughes, D.S. Palmer, F. Nigsch, J.B. Mitchell, Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P, *J. Chem. Inf. Model.* 48 (2008) 220–232.
- [20] A. Rahman, Correlations in the motion of atoms in liquid argon, *Phys. Rev.* 136 (1964) A405–A411.
- [21] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, 2017.
- [22] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 1, 2001.
- [23] A. Klamt, Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.* 99 (1995) 2224–2235.
- [24] A. Klamt, U. Hünig, S. Spycher, J. Keldenich, COSMOmic: a mechanistic approach to the calculation of membrane–water partition coefficients and internal distributions within membranes and micelles, *J. Phys. Chem. B* 112 (2008) 12148–12157.
- [25] B.R. Brooks, et al., CHARMM: the biomolecular simulation program, *J. Comput. Chem.* 30 (2009) 1545–1614.
- [26] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Chem. Physiol.* (1983).
- [27] D.A. Pearlman, et al., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Comput. Phys. Commun.* 91 (1995) 1–41.
- [28] J.C. Phillips, et al., Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (2005) 1781–1802.
- [29] L. Kalé, et al., NAMD2: greater scalability for parallel molecular dynamics, *J. Comput. Phys.* 151 (1999) 283–312.
- [30] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56.
- [31] S. Páll, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling exascale software challenges in molecular dynamics simulations with GROMACS, *Lecture Notes in Computer Science: Solving Software Challenges for Exascale*, Springer International Publishing, 2015, pp. 3–27.
- [32] M.J. Harvey, G. Giupponi, G.D. Fabritiis, ACEMD: accelerating biomolecular dynamics in the microsecond time scale, *J. Chem. Theory Comput.* 5 (2009) 1632–1639.
- [33] K. Bowers, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters, in: 43–43 (IEEE, 2006).
- [34] H.T. Davis, *Statistical Mechanics of Phases, Interfaces, and Thin Films*, VCH, New York, 1996.
- [35] W.R. Fawcett, *Liquids, Solutions, and Interfaces: From Classical Macroscopic Descriptions to Modern Microscopic Details*, Oxford University Press, 2004.
- [36] K.S. Shing, K.E. Gubbins, K. Lucas, Henry constants in non-ideal fluid mixtures: computer simulation and theory, *Mol. Phys.* 65 (1988) 1235–1252.
- [37] A.S. Paluch, S. Parameswaran, S. Liu, A. Kolavennu, D.L. Mobley, Predicting the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol mixtures via molecular simulation, *J. Chem. Phys.* 142 (2015) 01B621\_1.
- [38] B. Widom, Some topics in the theory of fluids, *J. Chem. Phys.* 39 (1963) 2808–2812.
- [39] C. Jarzynski, Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.* 78 (1997) 2690.
- [40] R.W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.* 22 (1954) 1420–1426.
- [41] C.H. Bennett, Efficient estimation of free energy differences from Monte Carlo data, *J. Comput. Phys.* 22 (1976) 245–268.
- [42] J.G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* 3 (1935) 300–313.
- [43] C. Chipot, A. Pohorille, *Free Energy Calculations*, Springer, 2007.
- [44] C.D. Christ, A.E. Mark, W.F. Van Gunsteren, Basic ingredients of free energy calculations: a review, *J. Comput. Chem.* 31 (2010) 1569–1582.
- [45] N. Hansen, W.F. Van Gunsteren, Practical aspects of free-energy calculations: a review, *J. Chem. Theory Comput.* 10 (2014) 2632–2647.
- [46] M.R. Shirts, D.L. Mobley, J.D. Chodera, Alchemical free energy calculations: ready for prime time, *Annu. Rep. Comput. Chem.* 3 (2007) 41–59.
- [47] M.R. Shirts, V.S. Pande, Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration, *J. Chem. Phys.* 122 (2005) 144107.
- [48] N. Lu, J.K. Singh, D.A. Kofke, Appropriate methods to combine forward and reverse free-energy perturbation averages, *J. Chem. Phys.* 118 (2003) 2977–2984.
- [49] N. Lu, D.A. Kofke, Optimal intermediates in staged free energy calculations, *J. Chem. Phys.* 111 (1999) 4414–4423.
- [50] D. Wu, D.A. Kofke, Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations, *J. Chem. Phys.* 123 (2005) 084109.
- [51] E. Sanz, C. Vega, Solubility of KF and NaCl in water by molecular simulation, *J. Chem. Phys.* 126 (2007) 014507.

- [52] A.S. Paluch, S. Jayaraman, J.K. Shah, E.J. Maginn, A method for computing the solubility limit of solids: application to sodium chloride in water and alcohols, *J. Chem. Phys.* 133 (2010) 124504.
- [53] J.L. Aragones, E. Sanz, C. Vega, Solubility of NaCl in water by molecular simulation revisited, *J. Chem. Phys.* 136 (2012) 244508.
- [54] J.T. Wescott, L.R. Fisher, S. Hanna, Use of thermodynamic integration to calculate the hydration free energies of n-alkanes, *J. Chem. Phys.* 116 (2002) 2361–2369.
- [55] T.P. Straatsma, H.J.C. Berendsen, Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations, *J. Chem. Phys.* (1988).
- [56] K. Lüder, L. Lindfors, J. Westergren, S. Nordholm, R. Kjellander, In silico prediction of drug solubility: 2. Free energy of solvation in pure melts, *J. Phys. Chem. B* 111 (2007) 1883–1892.
- [57] R. Gillet, A. Fierro, L.M. Valenzuela, J.R. Pérez-Correa, Using molecular dynamics simulations to predict the effect of temperature on aqueous solubility for aromatic compounds, *Fluid Phase Equilibria* 472 (2018) 85–93.
- [58] J. Westergren, et al., In silico prediction of drug solubility: 1. Free energy of hydration, *J. Phys. Chem. B* 111 (2007) 1872–1882.
- [59] K. Lüder, L. Lindfors, J. Westergren, S. Nordholm, R. Kjellander, In silico prediction of drug solubility. 3. Free energy of solvation in pure amorphous matter, *J. Phys. Chem. B* 111 (2007) 7303–7311.
- [60] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* 118 (1996) 11225–11236.
- [61] A.S. Paluch, E.J. Maginn, Predicting the solubility of solid phenanthrene: a combined molecular simulation and group contribution approach, *AIChE J.* 59 (2013) 2647–2661.
- [62] J.P. O'Connell, J.M. Prausnitz, Thermodynamics of gas solubility in mixed solvents, *Ind. Eng. Chem. Fundam.* 3 (1964) 347–351.
- [63] D.S. Corti, Alternative derivation of Widom's test particle insertion method using the small system grand canonical ensemble, *Mol. Phys.* 93 (1998) 417–420.
- [64] A.M.A. Dias, R.P. Bonifácio, I.M. Marrucho, A.A. Pádua, M.C. Gomes, Solubility of oxygen in n-hexane and in n-perfluorohexane. Experimental determination and prediction by molecular simulation, *Phys. Chem. Chem. Phys.* 5 (2003) 543–549.
- [65] M.F. Costa Gomes, A.A. Pádua, Interactions of carbon dioxide with liquid fluorocarbons, *J. Phys. Chem. B* 107 (2003) 14020–14024.
- [66] A.L. Ferguson, P.G. Debenedetti, A.Z. Panagiotopoulos, Solubility and molecular conformations of n-alkane chains in water, *J. Phys. Chem. B* 113 (2009) 6405–6414.
- [67] M. Khawaja, A.P. Sutton, A.A. Mostofi, Molecular simulation of gas solubility in nitrile butadiene rubber, *J. Phys. Chem. B* 121 (2016) 287–297.
- [68] S.K. Kumar, I. Szleifer, A.Z. Panagiotopoulos, Determination of the chemical potentials of polymeric systems from Monte Carlo simulations, *Phys. Rev. Lett.* 66 (1991) 2935.
- [69] T.-X. Xiang, B.D. Anderson, Water uptake, distribution, and mobility in amorphous poly (d, l-lactide) by molecular dynamics simulation, *J. Pharm. Sci.* 103 (2014) 2759–2771.
- [70] H. Tsuji, K. Sumida, Poly(L-lactide): v. effects of storage in swelling solvents on physical properties and structure of poly(L-lactide), *J. Appl. Polym. Sci.* 79 (2001) 1582–1589.
- [71] D.H. Yau, S.Y. Liem, K.-Y. Chan, A contact cavity-biased method for grand canonical Monte Carlo simulations, *J. Chem. Phys.* 101 (1994) 7918–7924.
- [72] H. Eslami, F. Müller-Plathe, Molecular dynamics simulation in the grand canonical ensemble, *J. Comput. Chem.* 28 (2007) 1763–1773.
- [73] J.R. Errington, A.Z. Panagiotopoulos, Phase equilibria of the modified Buckingham exponential-6 potential from Hamiltonian scaling grand canonical Monte Carlo, *J. Chem. Phys.* 109 (1998) 1093–1100.
- [74] H. Eslami, N. Mehdipour, Grand canonical ensemble molecular dynamics simulation of water solubility in polyamide-6, 6, *Phys. Chem. Chem. Phys.* 13 (2011) 669–673.
- [75] J.J. Potoff, J.R. Errington, A.Z. Panagiotopoulos, Molecular simulation of phase equilibria for mixtures of polar and non-polar components, *Mol. Phys.* 97 (1999) 1073–1083.
- [76] A. Hemmen, A.Z. Panagiotopoulos, J. Gross, Grand canonical Monte Carlo simulations guided by an analytic equation of state transferable anisotropic mie potentials for ethers, *J. Phys. Chem. B* 119 (2015) 7087–7099.
- [77] Y. Liu, A.Z. Panagiotopoulos, P.G. Debenedetti, Monte Carlo simulations of high-pressure phase equilibria of CO<sub>2</sub>-H<sub>2</sub>O mixtures, *J. Phys. Chem. B* 115 (2011) 6629–6635.
- [78] J.M. Rodgers, M. Webb, B. Smit, Alcohol solubility in a lipid bilayer: efficient grand-canonical simulation of an interfacial active molecule, *J. Chem. Phys.* 132 (2010) 064107.
- [79] H. Eslami, F. Müller-Plathe, Molecular dynamics simulation of sorption of gases in polystyrene, *Macromolecules* 40 (2007) 6413–6421.
- [80] M. Mezei, Grand-canonical ensemble Monte Carlo study of dense liquid: Lennard-Jones, soft spheres and water, *Mol. Phys.* 61 (1987) 565–582.
- [81] T. Çağın, B.M. Pettitt, Molecular dynamics with a variable number of molecules, *Mol. Phys.* 72 (1991) 169–175.
- [82] T. Cagin, B.M. Pettitt, Grand molecular dynamics: a method for open systems, *Mol. Simul.* 6 (1991) 5–26.
- [83] G.C. Lynch, B.M. Pettitt, Grand canonical ensemble molecular dynamics simulations: reformulation of extended system dynamics approaches, *J. Chem. Phys.* 107 (1997) 8594–8610.
- [84] S.K. Lakkaraju, E.P. Raman, W. Yu, A.D. MacKerell Jr, Sampling of organic solutes in aqueous and heterogeneous environments using oscillating excess chemical potentials in grand canonical-like Monte Carlo-molecular dynamics simulations, *J. Chem. Theory Comput.* 10 (2014) 2281–2290.
- [85] A. Agarwal, J. Zhu, C. Hartmann, H. Wang, L. Delle Site, Molecular dynamics in a grand ensemble: Bergmann-Lebowitz model and adaptive resolution simulation, *New J. Phys.* 17 (2015) 083042.
- [86] B. Shadrack Jabes, C. Krekeler, R. Klein, L. Delle Site, Probing spatial locality in ionic liquids with the grand canonical adaptive resolution molecular dynamics technique, *J. Chem. Phys.* 148 (2018) 193804.
- [87] N.M. Garrido, A.J. Queimada, M. Jorge, E.A. Macedo, I.G. Economou, 1-octanol/water partition coefficients of n-Alkanes from molecular simulations of absolute solvation free energies, *J. Chem. Theory Comput.* 5 (2009) 2436–2446.
- [88] K. Ogata, M. Hatakeyama, S. Nakamura, Effect of atomic charges on octanol-water partition coefficient using alchemical free energy calculation, *Molecules* 23 (2018) 425.
- [89] H. DeVoe, M.M. Miller, S.P. Wasik, Generator columns and high-pressure liquid-chromatography for determining aqueous solubilities and octanol-water partition-coefficients of hydrophobic substances, *J. Res. Natl. Bur. Stand.* 86 (1981) 361–366.
- [90] J. Sangster, Octanol-water partition coefficients of simple organic compounds, *J. Phys. Chem. Ref. Data* 18 (1989) 1111–1229.
- [91] N. Bhatnagar, G. Kamath, I. Chelst, J.J. Potoff, Direct calculation of 1-octanol-water partition coefficients from adaptive biasing force molecular dynamics simulations, *J. Chem. Phys.* 137 (2012) 014502.
- [92] C. Hansch, J.E. Quinlan, G.L. Lawrence, Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids, *J. Org. Chem.* 33 (1968) 347–350.
- [93] S.A. Best, K.M. Merz, C.H. Reynolds, Free energy perturbation study of octanol/water partition coefficients: comparison with continuum GB/SA calculations, *J. Phys. Chem. B* 103 (1999) 714–726.
- [94] S.E. DeBolt, P.A. Kollman, Investigation of structure, dynamics, and solvation in 1-octanol and its water-saturated solution: molecular dynamics and free-energy perturbation studies, *J. Am. Chem. Soc.* 117 (1995) 5316–5340.
- [95] N.M. Garrido, M. Jorge, A.J. Queimada, E.A. Macedo, I.G. Economou, Using molecular simulation to predict solute solvation and partition coefficients in solvents of different polarity, *Phys. Chem. Chem. Phys.* 13 (2011) 9155–9164.
- [96] W. Huang, N. Blinov, A. Kovalenko, Octanol-water partition coefficient from 3D-RISM-KH molecular theory of solvation with partial molar volume correction, *J. Phys. Chem. B* 119 (2015) 5588–5597.
- [97] J. Johnson, et al., Small molecule hydration energy and entropy from 3D-RISM, *J. Phys. Condens. Matter* 28 (2016) 344002.
- [98] M.G. Wolf, G. Groenhof, Evaluating nonpolarizable nucleic acid force fields: a systematic comparison of the nucleobases hydration free energies and chloroform-to-water partition coefficients, *J. Comput. Chem.* 33 (2012) 2225–2232.
- [99] M. Christen, et al., The GROMOS software for biomolecular simulation: GROMOS05, *J. Comput. Chem.* 26 (2005) 1719–1751.
- [100] B. Chen, J.I. Siepmann, Partitioning of alkane and alcohol solutes between water and (dry or wet) 1-octanol, *J. Am. Chem. Soc.* 122 (2000) 6464–6467.
- [101] B. Chen, J.I. Siepmann, Microscopic structure and solvation in dry and wet octanol, *J. Phys. Chem. B* 110 (2006) 3555–3563.
- [102] M.G. Martin, J.I. Siepmann, Calculating Gibbs free energies of transfer from Gibbs ensemble Monte Carlo simulations, *Theor. Chem. Acc.* 99 (1998) 347–350.
- [103] S. Genheden, Solvation free energies and partition coefficients with the coarse-grained and hybrid all-atom/coarse-grained MARTINI models, *J. Comput. Aided Mol. Des.* 31 (2017) 867–876.
- [104] T. Bereau, K. Kremer, Automated parametrization of the coarse-grained Martini force field for small organic molecules, *J. Chem. Theory Comput.* 11 (2015) 2783–2791.
- [105] S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. de Vries, The MARTINI force field: coarse grained model for biomolecular simulations, *J. Phys. Chem. B* 111 (2007) 7812–7824.
- [106] S. Jakobtorweihen, T. Ingram, I. Smirnova, Combination of COSMOmic and molecular dynamics simulations for the calculation of membrane-water partition coefficients, *J. Comput. Chem.* 34 (2013) 1332–1340.
- [107] M. Paloncýová, K. Berka, M. Otyepka, Convergence of free energy profile of coumarin in lipid bilayer, *J. Chem. Theory Comput.* 8 (2012) 1200–1211.
- [108] S. Jakobtorweihen, et al., Predicting solute partitioning in lipid bilayers: free energies and partition coefficients from molecular dynamics simulations and COSMOmic, *J. Chem. Phys.* 141 (2014) 07B622.1.
- [109] M. Paloncýová, R. DeVane, B. Murch, K. Berka, M. Otyepka, Amphiphilic drug-like molecules accumulate in a membrane below the head group region, *J. Phys. Chem. B* 118 (2014) 1030–1039.
- [110] M. Holmboe, P. Larsson, J. Anwar, C.A.S. Bergström, Partitioning into colloidal structures of fasted state intestinal fluid studied by molecular dynamics simulations, *Langmuir* 32 (48) (2016) 12732–12740, <https://doi.org/10.1021/acs.langmuir.6b03008>.
- [111] E.O. Kraemer, Molecular weights of celluloses and cellulose derivatives, *Ind. Eng. Chem.* 30 (1938) 1200–1203.
- [112] F.S. Moolman, M. Meunier, P.W. Labuschagne, P.-A. Truter, Journal of chemical theory and computation, *Polymer* 46 (2005) 6192–6200.
- [113] T.M. Madkour, A combined statistical mechanics and molecular dynamics approach for the evaluation of the miscibility of polymers in good, poor and non-solvents, *Chem. Phys.* 274 (2001) 187–198.
- [114] M. Maus, K.G. Wagner, A. Kornherr, G. Zifferer, Molecular dynamics simulations for drug dosage form development: thermal and solubility characteristics for hot-melt extrusion, *Mol. Simul.* 34 (2008) 1197–1207.
- [115] M. Valavi, M. Ukrainczyk, M.R. Dehghani, Prediction of solubility of active pharmaceutical ingredients by semi-predictive Flory Huggins/Hansen model, *J.*

- Mol. Liq. 246 (2017) 166–172.
- [116] J.H. Hildebrand, Factors determining solubility among non-electrolytes, Proc. Natl. Acad. Sci. USA 36 (1950) 7–15.
- [117] G. Scatchard, Equilibria in non-electrolyte solutions in relation to the vapor pressures and densities of the components, Chem. Rev. 8 (1931) 321–333.
- [118] A.F. Barton, *CRC Handbook of Solubility Parameters and Other Cohesion Parameters*, 2017.
- [119] J.E. Mark, *Polymer data handbook*, 2009.
- [120] L. Huynh, J. Grant, J.C. Leroux, P. Delmas, C. Allen, Predicting the solubility of the anti-cancer agent docetaxel in small molecule excipients using computational methods, Pharm. Res. 25 (2008) 147–157.
- [121] T. Spyriouni, C. Vergelati, A molecular modeling study of binary blend compatibility of polyamide 6 and poly (vinyl acetate) with different degrees of hydrolysis: an atomistic and mesoscopic approach, Macromolecules 34 (2001) 5306–5316.
- [122] P. Choi, T.A. Kavassalis, A. Rudin, Estimation of the three-dimensional solubility parameters of alkyl phenol ethoxylates using molecular dynamics, J. Colloid Interface Sci. 150 (1992) 386–393.
- [123] M. Belmares, et al., Hildebrand and Hansen solubility parameters from molecular dynamics with applications to electronic nose polymer sensors, J. Comput. Chem. 25 (2004) 1814–1826.
- [124] H. Yang, et al., Molecular dynamics simulation studies of binary blend miscibility of poly(3-hydroxybutyrate) and poly(ethylene oxide), Polymer 45 (2004) 453–457.
- [125] S.L. Mayo, B.D. Olafson, W.A. Goddard, DREIDING: a generic force field for molecular simulations, J. Phys. Chem. 94 (1990) 8897–8909.
- [126] M.I. Chaudhari, L.R. Pratt, M.E. Paulaitis, Concentration dependence of the Flory-Huggins interaction parameter in aqueous solutions of capped PEO chains, J. Chem. Phys. 141 (2014) 244908.
- [127] H. Sun, COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds, J. Phys. Chem. B 102 (1998) 7338–7364.
- [128] Z. Luo, J. Jiang, Molecular dynamics and dissipative particle dynamics simulations for the miscibility of poly(ethylene oxide)/poly(vinyl chloride) blends, Polymer 51 (2010) 291–299.
- [129] I.M. de Arenaza, E. Meaurio, B. Coto, J.-R. Sarasua, Molecular dynamics modelling for the analysis and prediction of miscibility in polylactide/polyvinylphenol blends, Polymer 51 (2010) 4431–4438.
- [130] X. Chen, C. Yuan, C.K. Wong, G. Zhang, Molecular modeling of temperature dependence of solubility parameters for amorphous polymers, J. Mol. Model. 18 (2012) 2333–2341.
- [131] J. Gupta, C. Nunes, S. Vyas, S. Jonnalagadda, Prediction of solubility parameters and miscibility of pharmaceutical compounds by molecular dynamics simulations, J. Phys. Chem. B 115 (2011) 2014–2023.
- [132] S.S. Patnaik, R. Pachter, A molecular simulations study of the miscibility in binary mixtures of polymers and low molecular weight molecules, Polymer 43 (2002) 415–424.
- [133] T.-X. Xiang, B.D. Anderson, Molecular dynamics simulation of amorphous hydroxypropylmethylcellulose and its mixtures with felodipine and water, J. Pharm. Sci. 106 (2017) 803–816.
- [134] C.P. Callaway, et al., Molecular modeling approach to determine the flory-huggins interaction parameter for polymer miscibility analysis, ChemPhysChem (2018).
- [135] L. Wu, L. Chen, H. Sun, On accuracy of predicting densities and solubility parameters of polymers using atomistic simulations, Mol. Simul. 43 (2017) 510–518.
- [136] C. Zhong, H. Masuoka, Modeling of gas solubilities in polymers with cubic equation of state, Fluid Phase Equilibria 144 (1998) 49–57.
- [137] K. Gauter, R.A. Heidemann, Modeling polyethylene-solvent mixtures with the Sanchez-Lacombe equation, Fluid Phase Equilibria (2001).
- [138] C. Kiparissides, V. Dimos, T. Bouloutouka, A. Anastasiadis, A. Chasiotis, Experimental and theoretical investigation of solubility and diffusion of ethylene in semicrystalline PE at elevated pressures and temperatures, J. Appl. Polym. Sci. 87 (2003) 953–966.
- [139] Y. Sato, M. Yurugi, K. Fujiwara, S. Takishima, H. Masuoka, Solubilities of carbon dioxide and nitrogen in polystyrene under high temperature and pressure, Fluid Phase Equilibria 125 (1996) 129–138.
- [140] Y. Sato, M. Yurugi, T. Yamabiki, S. Takishima, H. Masuoka, Solubility of propylene in semicrystalline polypropylene, J. Appl. Polym. Sci. 79 (2001) 1134–1143.
- [141] Y. Sato, et al., Vapor–liquid equilibrium ratios for hexane at infinite dilution in ethylene + impact polypropylene copolymer and propylene + impact polypropylene copolymer, Fluid Phase Equilibria 170 (2000) 49–67.
- [142] Y. Sato, K. Fujiwara, T. Takikawa, S. Takishima, H. Masuoka, Solubilities and diffusion coefficients of carbon dioxide and nitrogen in polypropylene, high-density polyethylene, and polystyrene under high pressures and temperatures, Fluid Phase Equilibria 162 (1999) 261–276.
- [143] M.C. Kroon, E.K. Karakatsani, I.G. Economou, G.-J. Witkamp, C.J. Peters, Modeling of the carbon dioxide solubility in imidazolium-based ionic liquids with the tPC-PSAFT equation of state, J. Phys. Chem. B 110 (2006) 9262–9269.
- [144] M. Lísal, W.R. Smith, J. Kolafa, Molecular simulations of aqueous electrolyte solubility: 1. The expanded-ensemble osmotic molecular dynamics method for the solution phase, J. Phys. Chem. B 109 (2005) 12956–12965.
- [145] V. Kanellopoulos, D. Mouratides, P. Pladis, C. Kiparissides, Prediction of solubility of  $\alpha$ -olefins in polyolefins using a combined equation of state molecular dynamics approach, Ind. Eng. Chem. Res. 45 (2006) 5870–5878.
- [146] M. Fermiglia, S. Pricl, Equation-of-state parameters for pure polymers by molecular dynamics simulations, AIChE J. 45 (1999) 2619–2627.