Molecular Dynamics Study of Interactions Between Nano Crystals and Solid-Liquid Phase Equilibria

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Ingenieurwissenschaft (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Gernot Bauer

aus Ostfildern

Hauptberichter: Prof. Dr.-Ing. Joachim Groß
Mitberichter: Prof. Dr.-Ing. habil. Jadran Vrabec

Tag der mündlichen Prüfung: 29.11.2019

Institut für Technische Thermodynamik und Thermische Verfahrenstechnik der Universität Stuttgart

2020

Eidesstattliche Erklärung zu meiner Dissertation mit dem Titel:

Molecular Dynamics Study of Interactions Between Nano Crystals and Solid-Liquid Phase Equilibria

Hiermit erkläre ich, dass ich die beigefügte Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

_____   _____
Ort, Datum   Unterschrift

Contents

Journal publications

This thesis led to the following publications:

- Chapter 2: G. Bauer, A. Lange, N. Gribova, C. Holm and J. Gross: Effective potentials between gold nano crystals - functional dependence on temperature, Molecular Simulation, 41 (14), 2015, 1153-1158

- Chapter 3: G. Bauer, N. Gribova, A. Lange, C. Holm and J. Gross: Three-body effects in triplets of capped gold nanocrystals, Molecular Physics, 155(9-12), 2017, 1031-1040

- Chapter 4: G. Bauer and J. Gross: Phase Equilibria of Solid and Fluid Phases from Molecular Dynamics Simulations with Equilibrium and Nonequilibrium Free Energy Methods, Journal of Chemical Theory and Computation, 15 (6), 2019, 3778-3792

The chapters 2 to 4 present literal quotes of the published work. Any addition with respect to the published work is marked. Any deletion is indicated with square brackets as ,[...]'. Cross-references between chapters of this thesis, which are added to the published version of the text to increase readability, are marked by square brackets. The Supporting Information is presented in the Appendix of this thesis.

Kurzfassung

Nanokristalle sind Makromoleküle, die als Bausteine so genannter Superstrukturen dienen. Die Gitterstruktur einer solchen Superstruktur – und damit auch deren Eigenschaften – hängen von der Art und Weise ab, mit welcher Nanokristalle wechselwirken. Kenntnis der Mechanismen dieser Wechselwirkungen ermöglicht es, Superstrukturen zu entwerfen, die maßgeschneiderte Eigenschaften besitzen. Diese Wechselwirkungen sind abhängig von strukturellen Eigenschaften der Nanokristalle, wie dem Material, aus welchem die Kerne bestehen, der Kerngröße, der Art der Liganden, sowie der Menge an Liganden, die auf der Kernoberfläche adsorbiert sind. Weitere Einflussparameter sind Temperatur, Zusammensetzung, Druck sowie die Art des Lösungsmittels. Die Untersuchung der Bildung von Superstrukturen mittels atomistischer Simulationen ist nicht möglich, da ein einzelner Nanokristall aus mehreren tausend Wechselwirkungszentren bestehen kann und der Bildungsprozess auf Zeitskalen abläuft, die nicht in atomistischen Simulationen erreicht werden können. Es müssen somit andere Modellierungsstrategien angewendet werden.

Eine Strategie liegt darin, die Freiheitsgrade des Systems zu reduzieren, indem alle atomaren Wechselwirkungen in ein effektives Potential zusammengefasst werden, welches lediglich von den relativen Positionen der Kerne der Nanokristalle abhängt. Im ersten Teil dieser Arbeit nutze ich diese Strategie zur Untersuchung von Wechselwirkungen zwischen Nano-kristallen im Vakuum, welche aus einem Goldkern mit Alkylthiolen als Liganden bestehen. Ohne Lösungsmittel ist das System allein durch die Temperatur charakterisiert, während die Struktur der Nanokristalle durch Kerngröße, Anzahl an Liganden und Ligandenlänge definiert ist. Ich ermittle das effektive Paarpotential in Form des *potential of mean force* (PMF) als Funktion der Kern-Abstände aus einer Reihe von Molekulardynamik Simulationen. Ein Ergebnis der Studie ist, dass die Temperaturabhängigkeit des PMF mittels thermodynamischer Störungstheorie abgeleitet werden kann, basierend auf einer Aufteilung von Wechselwirkungen in rein repulsive und attraktive Beiträge. Diese beiden Beiträge können aus zwei PMF bei unterschiedlichen Temperaturen ermittelt werden. Sind die Beiträge bekannt, können PMF in einem großen Temperaturbereich extrapoliert werden, was den Simulationsaufwand deutlich verringert. Der entwickelte Ansatz zeigt gute Übereinstimmung mit Simulationsergebnissen, sowohl für Paare gleichartiger Nanokristalle als auch für Paare von Nanokristallen unterschiedlicher Größe.

Betrachtet man einen einzelnen Nanokristall, so bilden die auf der Kernoberfläche adsorbierten Liganden eine annähernd sphärische Schicht um den Kern. Diese Schicht wird deformiert, wenn sich ein zweiter Nanokristall in unmittelbarer Nähe befindet, da die Ligandenschichten beider Kristalle überlappen was zu Bereichen hoher Ligandendichte führt. Auf Grund dieser lokal hohen Dichte kann die Wechselwirkung mit einem dritten Kristall nicht allein durch eine Summe an Paarwechselwirkungen beschrieben werden und

es ist unabdingbar Dreikörper- (bzw. Mehrkörper-) Potentiale zu berücksichtigen. Des Weiteren beschäftigt sich diese Arbeit deshalb mit der Untersuchung von Nanokristall-Triplets für welche ich ein Korrekturpotential als Funktion von zwei Kernabständen und einem Winkel entwickle. Dieses Korrekturpotential ist vorwiegend repulsiv bei kleinen Abständen und Winkeln und klingt ab wenn die drei Nanokristalle linear angeordnet sind. Meine Simulationen ergeben, dass das Korrekturpotential nicht von der Temperatur abhängt und ich schlage auf Grund dieser Erkenntnis ein empirisches Model vor, welches lediglich auf den repulsiven Zweikörper-Potentialen basiert.

Im zweiten Teil dieser Arbeit untersuche ich fest-flüssig Phasengleichgewichte von atomistischen und molekularen Systemen mittels Molekulardynamik Simulationen. Während es für Dampf-Flüssigkeits-Gleichgewichte etablierte Methoden und Simulationsabläufe gibt, können fest-fest sowie fest-flüssig Gleichgewichte oftmals nur durch Kombination mehrerer unterschiedlicher Methoden bestimmt werden. In dieser Studie wird die Schmelz-kurve für drei Systeme zunehmender Komplexität bestimmt: Argon, Methanol und Wasser. Dazu werden die chemischen Potentiale (die absoluten freien Energien) der beteiligten Phasen ermittelt und gleich gesetzt. Für die absolute freie Energie der Feststoffphasen stelle ich eine neue Methode vor – eine Kombination aus der *extended Einstein crystal* Methode und der *Einstein molecule* Methode – welche eine effiziente und numerische stabile Berechnung ermöglicht. Ich verwende dabei unterschiedliche freie Energie-Methoden: Thermodynamische Integration (TI) mit zwei Integrationsschemata, die *multistate Bennett acceptance ratio* (MBAR) Methode sowie Nichtgleichgewichts-Simulationen. Für präzise Ergebnisse sind alle Methoden (außer TI mit einem Integrationsschema nach Simpson) ähnlich effizient. Ich diskutiere die Unterschiede der Methoden hinsichtlich Anwendungsfreundlichkeit, Konvergenzverhalten sowie der Bewertung von statistischen Unsicherheiten. Die Schmelzkurve wird durch Simulationen im isothermen-isobaren Ensemble für unterschiedliche Temperaturen und Drücke in Kombination mit Vorhersagen durch Umgewichtung bestimmt, was trivial parallelisiert werden kann. Die auf diese Weise ermittelten Schmelzkurven haben eine niedrige statistische Unsicherheit und zeigen gute Übereinstimmung mit Literaturdaten.

Abstract

Nano crystals are macro molecules that act as building blocks for so called super structures. The lattice configuration of such a super structure – and therefore its properties – depends on the way nano crystals interact with each other which means that knowledge of the nature of nano crystal interactions enables the design of structures with tailor-made properties. These interactions are functions of structural properties of the nano crystals themselves such as the core material and size, the ligand type and the amount of ligands adsorbed on the core surface but they also depend on system properties like temperature, composition, pressure and solvent type. Because a single nano crystal can consist of multiple thousand interaction sites, studying the formation of super structures from atomistic simulations is not feasible since it occurs on a time-scale that cannot be covered with atomistic models, so that other modeling strategies have to be applied.

One such strategy is reducing the degrees of freedom of the system by encoding all atomistic interactions within effective potentials that only depend on relative core positions of nano crystals. In the first part of this work I follow this strategy and study interactions between nano crystals in vacuum that consist of icosahedral gold cores on which alkyl thiol ligands are adsorbed. Without a solvent, the only thermodynamic system parameter is the temperature whereas structural parameters are the core size, amount of ligands, and ligand length. I obtain the effective pair potential – the potential of mean force (PMF) – as a function of the core to core distance from a sequence of molecular dynamics simulations. I find that the functional dependence of the PMF on temperature can be motivated from thermodynamic perturbation theory by dividing interactions into a repulsive and an attractive contribution. These two contributions can be estimated from two PMF at different temperatures which allows for extrapolation to a large temperature region which significantly reduces the simulation effort. This approach shows good agreement with simulation data both for pairs of identical nano crystals as well as for pairs of different sized nano crystals.

The adsorbed ligands form a corona around the core which is approximately spherical for a single nano crystal but is deformed when two nano crystals are in close vicinity to each other because their coronae overlap which leads to regions of high ligand density. As a consequence, a third nano crystal will interact with the two other crystals in a way that cannot be described by pair interactions alone and three-body (and in fact multi-body) interactions have to be considered. I therefore study triplets of nano crystals and develop a three-body correction as function of the triplet configuration (two distances and one angle). I find that these corrections are mostly repulsive for small core distances and angles and quickly decay when approaching a linear configuration of three nano crystals. My simulations suggest that the three-body correction does not depend on temperature

and I propose an empirical model which only uses the repulsive contributions of the two-body PMF based on this observation.

In the second part of this work I study solid-liquid phase equilibria of atomistic and rigid molecular systems by means of molecular dynamics simulations. As opposed to vapor-liquid equilibria where well-established methods and workflows for determining phase equilibria exist, solid-solid and solid-liquid equilibria require the combination of multiple different methods. I conduct simulations to locate the melting transition line for three systems with increasing complexity – argon, methanol and water – by computing and equating the chemical potentials (absolute free energies) of candidate phases. For the absolute free energy of a solid phase I propose a new calculation method by combining the extended Einstein crystal and the Einstein molecule method enabling an efficient and numerically stable free energy pathway. Free energies are determined using different methods, i.e. thermodynamic integration (TI) with two integration schemes, the multistate Bennett acceptance ratio (MBAR) method as well as nonequilibrium simulations. I find that for highly precise results all methods (save Simpson's rule for TI) are comparably efficient to use and I discuss differences regarding ease of application, assessment of convergence as well as estimation of statistical uncertainties. The coexistence line is traced by combining simulations performed in the isothermal-isobaric ensemble for a range of temperatures and pressures based on estimates from reweighting techniques which can be done in an embarrassingly parallel fashion. The resulting melting lines have low statistical uncertainties and show good agreement with data from literature.

Chapter 1

Introduction

A central property in the field of molecular thermodynamics and also an important property of this thesis is the free energy. Free energies are thermodynamic potential functions and so-called fundamental equations, which means that once the free energy of a system is known all thermodynamic properties are directly accessible from its partial derivatives. Knowledge of free energies allows calculating e.g. properties and phase equilibria, binding affinities and reaction barriers. Being based on principles of statistical mechanics by means of the partition function, free energies intrinsically encode microscopic information and therefore provide a tool to describe thermodynamical, chemical and biological processes in a systematic way and across different (length) scales.

1.1   Partition functions and free energies

In statistical mechanics, free energies are formulated in terms of their associated partition functions, e.g. the Helmholtz energy A is formulated in terms of the *canonical* partition function Q as

$$\beta A(N, V, T) = -\ln Q(N, V, T)\,, \tag{1.1}$$

where $\beta = 1/kT$, k is Boltzmann's constant, T denotes the temperature, N is the number of molecules and V represents the system volume. N, V and T are control variables and are defined for a system with constant values. For each control variable there is a response of the system; for the canonical ensemble these are

- the chemical potential $\mu = \left(\frac{\partial A}{\partial N}\right)_{T,V}$,

- the pressure $P = -\left(\frac{\partial A}{\partial V}\right)_{N,T}$,

- and the internal energy $U = \left(\frac{\partial \beta A}{\partial \beta}\right)_{N,V}$.

The type of control variables can be chosen to suit the system, experiment or property of interest. Changing control variables can be achieved using a Legendre transform towards a different thermodynamic potential or a Laplace transform towards a different partition function.[1] For example a transformation of the entropy as function of the amount of substance N, volume V and internal energy E, $S(N, V, E)$, to the Helmholtz energy, $A(N, V, \beta)$, can be performed either via Legendre Transformation by transformation of the internal energy into the inverse temperature, β, or via Laplace transformation of the microcanonical partition function $\Omega(U, V, N)$. In the same fashion one can derive the Gibbs energy as function of amount of substance, pressure and temperature, $G(N, P, T)$, as well as the Grand potential as function of the chemical potential, volume and temperature, $\Xi(\mu, V, T)$.

The partition function is the normalization constant of the probability density function of a microstate for a given set of control variables. The canonical probability density function, $\rho(x, N, V, T)$, and partition function, $Q(N, V, T)$, read

$$\rho(x, N, V, T) = \frac{C e^{-\beta \mathcal{H}(x)}}{Q(N, V, T)} \tag{1.2}$$

$$Q(N, V, T) = C \int e^{-\beta \mathcal{H}(x)} \mathrm{d}x \,, \tag{1.3}$$

with the phase space point $x = \{\mathbf{r}^N, \mathbf{p}^N\}$, the particle positions and momenta of the system, $\mathbf{r}^N$ and $\mathbf{p}^N$, respectively, the normalization constant of the partition function $C = C(N, V, T)$ and the Hamiltonian $\mathcal{H}$ which describes the total energy of the system. For any function of the phase space coordinates, $\alpha(x)$, the *ensemble average* can be computed by integration, as

$$\langle \alpha \rangle_{NVT} = \frac{C}{Q(N, V, T)} \int \alpha(x) e^{-\beta \mathcal{H}(x)} \mathrm{d}x = \frac{\int \alpha(x) e^{-\beta \mathcal{H}(x)} \mathrm{d}x}{\int e^{-\beta \mathcal{H}(x)} \mathrm{d}x} \,. \tag{1.4}$$

Free energies form the basis of a plethora of modern methods that are used to model and optimize technical processes and fluid properties but also to study mechanisms and driving forces on a molecular level. Modern equations of state such as the family of SAFT (Statistical Associating Fluid Theory) equations of state are formulated in terms of additive Helmholtz energy contributions, each of which is based on intra- or inter-molecular energies arising e.g. from dispersive, polar or associative interactions.[2] Macroscopic properties can be calculated by combining suitable contributions based on the topology of the substance or mixture one is interested in and – since all contributions are based on molecular principles – equations of state of this kind can be used for a wide range of substances and thermodynamic states in a transferable fashion.

These Helmholtz energy contributions can be derived from a series expansion by in-

troducing a perturbation – for example within the intermolecular energy function – to describe molecular fluids with respect to a preferably well-known reference. The expansion terms are integrals over pairs (first order) of particles, 4 particles (second order), and so forth and contain pair, four-body, etc. correlation functions of the reference fluid.[3] In general these correlation integrals cannot be solved analytically; therefore they are often supplemented by molecular simulations or parameterized using experimental data which limits transferability of the equation of state.

Another way to obtain free energies is by means of molecular simulations. Different from theoretical approaches, in molecular simulations one generates a large number of *representative* microstates for a given set of control variables and Hamiltonian. 'Representative' means that microstates are generated according to the probability density of the ensemble. The two most widely used methods are Monte-Carlo (MC) and Molecular Dynamics (MD).[4,5] MC simulations are stochastic processes where a Markov chain of microstates is created. In this work, molecular simulations are carried out via MD where microstates are created by temporal evolution of the system based on integration of an equation of motion. When microstates are generated according to the ensemble probability density, an ensemble or time average of a static property, $\alpha(x)$, can be determined by simple averaging, i.e.

$$\langle \alpha \rangle = \frac{1}{n} \sum_{i}^{n} \alpha(x_i) \, , \tag{1.5}$$

where n is the total number of microstates generated in the simulation. In contrast to eq. (1.4), the probability density function enters the average implicitly as it is used in the creation of microstates.

Similar to equations of state, in molecular simulations one needs a set of functions that describes energies between interaction sites, which can be atoms or groups of atoms, in the system. This set of energy functions is called a force field where forces are obtained by taking the negative spatial derivative of the (scalar) potential functions. In this work, classical potentials are used to describe intramolecular energies (such as covalent bonds, angles and torsion) as well as intermolecular interactions (van der Waals and electrostatic). These potentials are mathematically comparably simple, but combining multiple of these functions enables simulation of arbitrarily complex systems.

The actual (*a priori* unknown) probability density function is not needed for a simulation nor is it generally readily available as a result from a simulation – and as such neither is the free energy of the system. De facto it is not feasible to compute the *absolute* free energy except for very simple model systems. In practice this is rarely a problem as it is typically sufficient to have information about the free energy *difference* between two states. For example, phase equilibria are found where the difference of the Gibbs

energy between two phases vanishes and likewise the stability of protein conformations or different solid structures can be assessed based on relative free energies.

1.2 Estimating free energies in molecular simulations

In simulations, free energy differences are obtained by traversing a path between two systems or states of interest. Only if the absolute free energy of one of the two systems or states is known, the absolute free energy of the other system or state can be determined by evaluation of the free energy difference. In many cases, however it is sufficient to know the free energy difference of two systems. The path connecting both systems or states may be physically meaningful. For example going from a state with low temperature to a state with high temperature at constant volume and amount of substance may be realized in a laboratory experiment by transferring heat to the system. In simulations one can also choose hypothetical pathways which may be particularly suited or convenient to solve the problem at hand or to increase efficiency of a simulation. For example, to compute the free energy of solvation of a large molecule in a dense solvent one can incrementally grow the molecule or create (and later remove) a cavity in which the molecule can be inserted.[6] Another example are flat histogram Monte-Carlo simulations, where one utilizes free energy differences (which are estimated during the course of a simulation) to force the system to visit e.g. predefined volumes (in constant pressure simulations) or numbers of molecules (in constant chemical potential simulations) with equal probability, making it possible to observe a liquid and a vapor phase within a single simulation.[7–11]

One of the oldest and most widely used free energy method is the free energy perturbation (FEP) formalism also known as Zwanzig's high-temperature expansion or exponential averaging.[12] When using FEP one is interested in the free energy difference between a *reference* system or state and the *target*. For a perturbation of the potential energy one can formulate the target, U_1, as

$$U_1(\mathbf{r}) = U_0(\mathbf{r}) + \Delta U(\mathbf{r}) \,, \tag{1.6}$$

where U_0 denotes the reference potential energy and ΔU is the perturbation term. The free energy difference between both states can be written as a ratio of the configurational partition functions, $Z(N, V, T)$, as (omitting the control variables for brevity)

$$A_1 - A_0 = \Delta_{01} A = -kT \ln \frac{Z_1}{Z_0} \,, \tag{1.7}$$

where the configurational partition function reads

$$Z_i(N, V, T) = \int e^{-\beta U_i(\mathbf{r}^N)} \mathrm{d}\mathbf{r}^N . \tag{1.8}$$

Here, one assumes that the kinetic free energy contribution is the same in both states and hence the integrals over all particles' momenta cancel out when taking the ratio of the partition functions. Substituting eq. (1.6) into Z_1 of eq. (1.7) leads to the following reformulation

$$\beta \Delta_{01} A = -\ln\langle e^{-\beta \Delta U} \rangle_0 \tag{1.9}$$
$$= \ln\langle e^{\beta \Delta U} \rangle_1 . \tag{1.10}$$

These equations are exact only if both states' Boltzmann factors are finite within the same phase space volume. The index of the angular brackets denotes the ensemble's underlying potential energy function, i.e. $\langle \ldots \rangle_0$ denotes an ensemble average where configurations are generated according to the probability density distribution of the reference system.

FEP is a convenient method to compute the free energy difference because only a single simulation is required which – depending on the perturbation ΔU – may even be used to compute free energy differences for multiple target potentials. In practice, FEP is limited to small perturbations or more precisely, small variances of the perturbation.[13] A large variance indicates that reference and target only have a small number of representative configurations in common, or in other words, that they have small *overlap* of the respective probability densities. This problem may be alleviated either by combining simulations from both end states (as discussed below) or by introducing intermediate states granting a "slower/finer transition" between reference and target. Repeated evaluation of eq. (1.9) for all adjacent states can then be used to connect data from multiple intermediate simulations. However, one has to be careful when parts of the phase space of either reference or target are inaccessible from the other respective state which e.g. is the case for a hard sphere as a reference and a Lennard-Jones (LJ) fluid as a target fluid. For LJ-fluids, distances smaller than the hard sphere diameter may constitute a relevant contribution to the free energy whereas these configurations have a probability of zero to be observed within the hard sphere system.

FEP can be considered as limiting case of Bennett's acceptance ratio (BAR) method.[14] Bennett showed that the free energy difference between two states can be formulated in terms of ensemble averages of a finite function $W(\{\mathbf{r}^N\})$ as

$$\beta \Delta_{01} A = \ln \frac{\langle W e^{-\beta U_0} \rangle_1}{\langle W e^{-\beta U_1} \rangle_0} \,, \tag{1.11}$$

which can be optimized to yield the minimal statistical variance of the free energy estimate (with respect to W) resulting in two functions that can be iteratively solved, i.e.

$$e^{-\beta \Delta_{01} A} = \frac{\langle (1 + e^{\beta \Delta U - C})^{-1} \rangle_0}{\langle (1 + e^{-\beta \Delta U + C})^{-1} \rangle_1} e^{-C} \,, \tag{1.12}$$

$$C = \beta \Delta_{01} A + \ln \frac{n_1}{n_0} \,, \tag{1.13}$$

where n_0 and n_1 denote the number of uncorrelated samples as obtained from reference and target state simulations, respectively. These two equations constitute the acceptance ratio method and reduce to the FEP formalism when either n_0 or n_1 is zero. The optimal choice for $W = (1 + e^{\beta \Delta U - C})^{-1}$ is the Fermi function and it was shown to be the best asymptotically unbiased estimator for the free energy.[15,16]

A further generalization of BAR is the multistate Bennett acceptance ratio (MBAR) method.[17–19] I argued above that insufficient overlap between states may be alleviated by conducting additional simulations that act as bridge between the reference and target state. Chaining evaluations of adjacent states – either using FEP or BAR – allows calculating the desired free energy difference. However, information of non-adjacent states is not utilized in such a procedure. The MBAR approach is very similar to BAR where the free energy between two states, say i and j, is also formulated by introducing ensemble averages

$$\beta \Delta_{ij} A = \ln \frac{\langle W_{ij} e^{-\beta U_i} \rangle_j}{\langle W_{ij} e^{-\beta U_j} \rangle_i} \,, \tag{1.14}$$

but for a total of K states where W is a $K \times K$ matrix, encoding information between all pairs of states. Similar to BAR, the goal is to find an optimal solution for W *across all states*. The full derivation of the MBAR equation is outside the scope of this brief introduction but for the Helmholtz energy it reads

$$\beta A_i = - \ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e^{-\beta U_i(\mathbf{r}_{nj})}}{\sum\limits_{k=1}^{K} N_k e^{\beta A_k - \beta U_k(\mathbf{r}_{nj})}} \,, \tag{1.15}$$

where the indexes i and j denote thermodynamic states, n counts through single observa-

tions of a simulation and $\mathbf{r}_{nj}$ is the n'th configuration (from a total of N_j configurations) observed in simulation of state j. While equation (1.15) defines the absolute free energy it can only be solved (self-consistently) up to an additive constant, which ultimately leads to free energy differences.

The general form of the MBAR equation for the free energy, F, reads

$$\beta_i F_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e^{-u_i(\mathbf{r}_{nj})}}{\sum\limits_{k=1}^{K} N_k e^{\beta F_k - u_k(\mathbf{r}_{nj})}} \, , \tag{1.16}$$

where the reduced potential function, $u_i(\mathbf{r}_{nj})$, was introduced that reads

$$u_i(\mathbf{r}_{nj}) = \beta_i \left[U_i(\mathbf{r}_{nj}) + p_i V(\mathbf{r}_{nj}) - \mu_i N(\mathbf{r}_{nj}) \right] \, , \tag{1.17}$$

where U_i is the total potential energy, p_i is the pressure, V is the volume and μ_i and N are the chemical potential and the number of molecules in the system, respectively. Depending on the choice of terms considered in the reduced potential function $u_i(\mathbf{r}_{nj})$, different free energies can be evaluated, i.e. using $u_i = \beta_i U_i$ corresponds to control variables $\{N, V, T\}$ and yields the Helmholtz energy, $u_i = \beta_i U_i + \beta_i p_i V$ gives the Gibbs energy and $u_i = \beta_i U_i - \beta_i \mu_i N$ is used for defined $\{\mu, V, T\}$ and leads to the Grand potential.

To utilize the MBAR formalism, each configuration observed in state i has to be evaluated for each reduced potential j. This is trivial for the $\beta_i p_i V$ and $\beta_i \mu_i N$ terms since one can simply store a time series of volume and number of molecules from a simulation. Reevaluation for different temperatures, pressures and chemical potentials is then done by combining (i.e. multiplying) the new state's control variables with the sampled time series as a post-processing step. In contrast, the state dependence of a potential energy function U_i may not be as trivial (i.e. linearly dependent) and – in the worst case – one has to store all trajectories to facilitate reevaluation of the total potential energy if the simulation code does not provide mechanisms to perform reevaluations on the fly. For simple perturbations (or when a perturbation can be separated into linear basis functions[20,21]) however, reevaluation of trajectories is rarely needed and MBAR can be used exclusively as a post-processing step without further simulations.

Note that MBAR is very similar to the weighted histogram analysis method (WHAM).[22,23] In WHAM, simulation data is binned into histograms which are then combined. For the limiting case of zero bin width the WHAM equation reduces to the above shown MBAR equation (eq. (1.16)).

Another very common method to estimate free energies from molecular simulations is

thermodynamic integration (TI). Let us reformulate the perturbation potential by introducing a coupling parameter λ. As an example, one could use linear coupling, i.e.

$$U(\lambda) = (1 - \lambda)U_0 + \lambda U_1 \,, \tag{1.18}$$

but in general, any coupling can be used so that when λ runs from zero to unity the reference potential and the target potential are recovered, respectively. In fact, the way the coupling is introduced is a degree of freedom for optimizing the procedure and will vary depending on the system of interest. The free energy difference can be calculated via

$$\Delta_{01}A = \int_0^1 \frac{\mathrm{d}A}{\mathrm{d}\lambda}\mathrm{d}\lambda \,. \tag{1.19}$$

The derivative with respect to the coupling parameter is

$$\frac{\mathrm{d}A}{\mathrm{d}\lambda} = \frac{\partial}{\partial\lambda}\left[-kT\ln\tilde{C}(N,V,T)\int e^{-\beta U(\lambda,\mathbf{r}^N)}\mathrm{d}\mathbf{r}^N\right]\,, \tag{1.20}$$

$$= -\frac{1}{\beta Z(N,V,T,\lambda)}\int\frac{\partial}{\partial\lambda}e^{-\beta U(\lambda,\mathbf{r}^N)}\mathrm{d}\mathbf{r}^N\,, \tag{1.21}$$

$$= -\frac{1}{\beta Z(N,V,T,\lambda)}\left(-\beta\int\frac{\partial U(\lambda,\mathbf{r}^N)}{\partial\lambda}e^{-\beta U(\lambda,\mathbf{r}^N)}\mathrm{d}\mathbf{r}^N\right)\,, \tag{1.22}$$

$$= \left\langle\frac{\mathrm{d}U(\lambda,\mathbf{r}^N)}{\mathrm{d}\lambda}\right\rangle_\lambda\,, \tag{1.23}$$

so that

$$\Delta_{01}A = \int_0^1 \left\langle\frac{\mathrm{d}U(\lambda,\mathbf{r}^N)}{\mathrm{d}\lambda}\right\rangle_\lambda \mathrm{d}\lambda \,. \tag{1.24}$$

The constant $\tilde{C}(N,V,T)$ contains an integration over all particle's momenta which are unaffected by the perturbation and hence, cancels out. Different from the previously discussed methods, TI requires one to sample the derivative of the potential energy function (the slope) with respect to the coupling parameter. Integration can be done by introducing nodes (values of λ) where simulations are performed. The way the nodes are chosen depends on the integration scheme. Efficiency of TI depends on the graph of $\frac{\mathrm{d}U}{\mathrm{d}\lambda}$ versus λ. If the slope is a smooth, slowly changing function of λ, a low number of integration nodes is needed requiring very little simulation time. Otherwise additional simulations

15

have to be added in regions where the slope changes rapidly. Degrees of freedom for optimizing the efficiency of the TI application are: the way the coupling parameter λ leads to a (non-linear) transition between reference and target, equation (1.18), the number of integration nodes and the integration scheme.

1.3 Potential of mean force

The potential of mean force (PMF) is a free energy profile along a coordinate of interest.[13,24,25] This coordinate, R, often called reaction coordinate or collective variable, can be any function of the configurational degrees of freedom in the system, $R = R(\mathbf{r}^N)$, e.g. a distance or angle between molecules or interaction sites. The Helmholtz energy as a function of R is referred to as potential of mean force, with

$$\beta A(N, V, T; R') = -\ln Q(N, V, T; R') \tag{1.25}$$

$$= -\ln C(N, V, T) \int e^{-\beta \mathcal{H}(x)} \delta \left[R' - R(\mathbf{r}^N) \right] \mathrm{d}x\,, \tag{1.26}$$

where the accessible phase space is reduced to those states that satisfy $R(\mathbf{r}^N) = R'$ by means of the delta distribution. The PMF can be related to the Helmholtz energy by taking the exponent and integrating over all possible values of the reaction coordinate

$$\int e^{-\beta A(N, V, T, R)} \mathrm{d}R = C(N, V, T) \int e^{-\beta \mathcal{H}(x)} \mathrm{d}x = e^{-\beta A(N, V, T)}\,. \tag{1.27}$$

As before, only differences of the PMF are meaningful, that is one can write (omitting the control variables)

$$A(R'_1) - A(R'_0) = \int\limits_{R'_0}^{R'_1} \frac{\mathrm{d}A(R'')}{\mathrm{d}R''}\mathrm{d}R'' \tag{1.28}$$

$$= \int\limits_{R'_0}^{R'_1} \left\langle \frac{\mathrm{d}U(\mathbf{r}^N)}{\mathrm{d}R''} \right\rangle_{R''} \mathrm{d}R'' \tag{1.29}$$

$$= \int\limits_{R'_0}^{R'_1} \left\langle \frac{\mathrm{d}U(\mathbf{r}^N)}{\mathrm{d}\mathbf{r}^N} \cdot \frac{\mathrm{d}\mathbf{r}^N}{\mathrm{d}R''} \right\rangle_{R''} \mathrm{d}R'' \tag{1.30}$$

$$= -\int\limits_{R'_0}^{R'_1} \left\langle \mathbf{f}^N \cdot \frac{\mathrm{d}\mathbf{r}^N}{\mathrm{d}R''} \right\rangle_{R''} \mathrm{d}R'', \tag{1.31}$$

where $\mathbf{f}^N$ in the last line is the force (hence the name) which is projected along the reaction coordinate via the dot product with $\frac{\mathrm{d}\mathbf{r}^N}{\mathrm{d}R''}$. Note that the ensemble average is performed using the microstate probability density subject to the constraint, i.e.

$$\left\langle \alpha(x) \right\rangle_{R''} = \frac{\int \alpha(x) e^{-\beta\mathcal{H}(x)} \delta\left[R'' - R(\mathbf{r}^N)\right] \mathrm{d}x}{\int e^{-\beta\mathcal{H}(x)} \delta\left[R'' - R(\mathbf{r}^N)\right] \mathrm{d}x}. \tag{1.32}$$

In principle the PMF could be obtained from a "regular" simulation by simply sampling the reaction coordinate during the course of the simulation followed by an estimation of the probability, $P(R)$, e.g. using a histogram. More often than not such a naive approach cannot be used because the statistical uncertainty (for the entire PMF-range) is high. This issue may either arise due to slow dynamics or due to some values of the reaction coordinate being very improbable to be observed. Both causes can be circumvented by forcing the system to visit the entire range of the reaction coordinate either by introducing restraints (e.g. Umbrella sampling[25]) or constraints (e.g. SHAKE[26] or RATTLE[27]).

References

1. Zia, R. K., Redish, E. F. & McKay, S. R. Making sense of the Legendre transform. *Am. J. Phys.* **77,** 614–622 (2009).

2. Chapman, W. G., Gubbins, K. E., Jackson, G. & Radosz, M. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilib.* **52,** 31–38 (1989).

3. Hansen, J.-P. & McDonald, I. R. *Theory of simple liquids* (Elsevier, 1990).

4. Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 2001).

5. Tuckerman, M. *Statistical mechanics: theory and molecular simulation* (Oxford University Press, 2010).

6. Li, L., Totton, T. & Frenkel, D. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. *J. Chem. Phys.* **146,** 214110 (2017).

7. Fitzgerald, M., Picard, R. & Silver, R. Canonical transition probabilities for adaptive Metropolis simulation. *Europhys. Lett.* **46,** 282 (1999).

8. Fitzgerald, M., Picard, R. & Silver, R. Monte Carlo transition dynamics and variance reduction. *J. Stat. Phys.* **98,** 321–345 (2000).

9. Wang, F. & Landau, D. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86,** 2050 (2001).

10. Errington, J. R. Direct calculation of liquid–vapor phase equilibria from transition matrix Monte Carlo simulation. *J. Chem. Phys.* **118,** 9915–9925 (2003).

11. Hemmen, A., Panagiotopoulos, A. Z. & Gross, J. Grand Canonical Monte Carlo Simulations Guided by an Analytic Equation of State – Transferable Anisotropic Mie Potentials for Ethers. *J. Phys. Chem. B* **119,** 7087–7099 (2015).

12. Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **22,** 1420–1426 (1954).

13. Chipot, C. & Pohorille, A. *Free energy calculations* (Springer, 2007).

14. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **22,** 245–268 (1976).

15. Crooks, G. E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **61,** 2361 (2000).

16. Shirts, M. R., Bair, E., Hooker, G. & Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **91,** 140601 (2003).

17. Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **135,** 40–57 (2001).

18. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129,** 124105 (2008).

19. Shirts, M. R. Reweighting from the mixture distribution as a better way to describe the Multistate Bennett Acceptance Ratio (2017).

20. Naden, L. N., Pham, T. T. & Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 1. Removal of Uncharged Atomic Sites. *J. Chem. Theory Comput.* **10.** PMID: 26580188, 1128–1149 (2014).

21. Naden, L. N. & Shirts, M. R. Linear basis function approach to efficient alchemical free energy calculations. 2. Inserting and deleting particles with coulombic interactions. *J. Chem. Theory Comput.* **11,** 2536–2549 (2015).

22. Ferrenberg, A. M. & Swendsen, R. H. Optimized monte carlo data analysis. *Comput. Phys.* **3,** 101–104 (1989).

23. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13,** 1011–1021 (1992).

24. Trzesniak, D., Kunz, A.-P. E. & van Gunsteren, W. F. A comparison of methods to compute the potential of mean force. *ChemPhysChem* **8,** 162–169 (2007).

25. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **1,** 932–942 (2011).

26. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

27. Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52,** 24–34 (1983).

Chapter 2

Effective potentials between gold nano crystals - functional dependence on the temperature

The content of this chapter is a literal quote of the publication

> *G. Bauer, A. Lange, N. Gribova, C. Holm and J. Gross, Molecular Simulation, 41 (14), 2015, 1153-1158*

A. Lange was involved in preceding work concerning a density functional approach that led to this study. N. Gribova helped establishing the simulation workflow and performed some of the presented simulations. C. Holm had an advisory role in setting up simulations. J. Gross had the role of a daily supervisor and was involved in editing the manuscript. Additions or deletions compared to the published work are marked with angular brackets.

A method is presented that allows to combine the effective potential between two nano crystals (NC), the potential of mean force (PMF), as obtained from all-atomistic molecular dynamics simulations with perturbation theory. In this way, a functional dependence of the PMF on temperature is derived, which enables the prediction of the PMF in a wide temperature range. We applied the method to systems of capped gold NCs of different size. They show very good agreement with data from atomistic simulations.

2.1   Introduction

Nano crystals (NCs) are building blocks of newly engineered materials that combine optical and electrical properties in a custom made fashion.[1–3] Depending on their shape, the solvent, and the temperature, nano crystals can form superstructures, each of which is associated with a particular set of physical properties.[4–6]

In principle, it should be possible to accurately describe NCs with a capping layer of the so-called ligands and their corresponding superstructures via molecular simulation. Determining the superstructure of NCs from atomistic molecular simulations is, however difficult, because of the large number of interaction sites – up to several 1000 for a single NC – and most importantly the slow dynamics of the phase transition towards stable superstructures.

As a consequence, in practice coarse graining strategies have to be applied. We take the route to first determine the effective pair potential between two NC, the potential of mean force (PMF), from all-atomistic molecular dynamic (MD) simulations.[7] Calculating phase diagrams of various stable superstructures still becomes tedious, because the PMF for defined NCs (or mixtures of NCs) depends on the surrounding solvent and temperature.

In this work, we propose a method to predict the PMF between two NCs in vacuum for different temperatures. Using thermodynamic perturbation theory (TPT) of first order, we develop a correlation of the free energy as a function of the temperature that simplifies the practical application considerably. Parameterising the perturbation expression with results from Molecular Dynamics (MD) simulations enables the prediction of PMF for a wide range of different temperatures from only two simulated PMF curves.

The method is applied to various systems of gold NCs capped with alkyl chains with thiol head groups in vacuum. Variables are the core size and the number and length of attached ligands. The predicted potentials are compared with simulation data and show very good agreement.

The remainder of the paper is structured as follows. First, we describe the simulation setup and summarize the TPT. Then, the results for different systems are presented, followed by a brief conclusion.

2.2   Method

In this section we introduce a method that relates MD simulations to an analytic fluid theory to predict the temperature dependence of the PMF. First, we review how the PMF, representing the effective interactions of capped gold nano crystals at fixed distances, is obtained from molecular simulations. Then we summarize the relevant elements of thermodynamic perturbation theory (TPT) that are needed to model the underlying

temperature dependence. For the application of capped NCs, we derive a relation for calculating the PMF at an arbitrary temperature.

2.2.1   PMF from molecular dynamic simulations

There are different methods to obtain the PMF via computer simulations e.g. configurational-bias Monte Carlo, steered Molecular Dynamics or constraint Molecular Dynamics simulations. An overview is given by Trzesniak et al.[8]

In this work we use constraint Molecular Dynamics simulations. We align the centers of mass of the two NCs with the x-axis of the simulation box. The center of mass distance is kept constant during a simulation by freezing the gold atoms while we decrease the distance step-by-step in-between multiple simulations. By freezing the gold cores instead of only restraining their centers of mass we make the following assumptions: (i) The rotation of the NCs is negligible within the time frame of a simulation run especially at distances where the ligand spheres start to overlap.[7] (ii) The NC gold cores are modelled as rigid icosahedra, exposing only (111)-facets[9] but there is no orientational effect, i.e. the orientation in which we freeze the core does not matter and the contributions due to core-core interactions can be neglected compared to the contributions due to ligand interactions.[10]

According to a united-atom approach, we consider SH-, CH_2- and CH_3-groups in the ligands as single interaction sites with a force field described in the work of Schapotschnikow et al.[7] Our simulations were performed in the NVT ensemble using the GROMACS simulation software[11]. The integrator that also acts as a thermostat is a leap-frog stochastic dynamics integrator where we set the friction constant to $\gamma_i = 0.5\,\mathrm{ps}^{-1}$ according to the suggestion in the GROMACS manual. The integrator step size is $\Delta t = 0.002\,\mathrm{ps}$. A simulation (at a specific center of mass distance) consists of an energy minimization run and a production run. We conduct 5 to 10 runs for every temperature, each with an equilibration time of $1\,\mathrm{ns}$ and a total runtime up to $16\,\mathrm{ns}$.

In general, the size of the simulation box has to be sufficiently large to avoid interactions of the capping layers of the NCs with their periodic images. For our simulations, all axes of the box are at twice the initial center of mass distance. A convenient choice for the initial distance r_{12} is $r_{12} \geq d_c + 2L$, where d_c ist the diameter of the gold core and L is the length of an elongated ligand. At $r_{12} = d_c + 2L$ the two NCs start to touch each other. A larger separation distance ensures that the initial forces on the centers of mass of both NCs are zero.

To calculate the PMF we measure the average forces acting on the center of mass of both gold cores, $\mathbf{F_1}$ and $\mathbf{F_2}$ over the course of a simulation run. The mean force between

the two NCs is then obtained from

$$F_m(r_{12}) = \frac{1}{2} \langle (\mathbf{F_2} - \mathbf{F_1}) \cdot \mathbf{r_u} \rangle_{r_{12}},$$

where $\mathbf{r_u} = \mathbf{r_{12}}/r_{12}$ is the unit vector connecting two particles and the angular brackets denote the average in the restrained canonical ensemble. In the last step we integrate the obtained forces over all distances to get the potential of the mean force, as

$$\Phi_m(r_{12}) = \int\limits_{r_{12}}^{\infty} F_m(s) \, \mathrm{d}s.$$

2.2.2   Thermodynamic perturbation theory

The complex interactions between particles can successfully be modeled via perturbation theory.[12,13] The starting point is a pair potential between atomistic (or united-atom) interaction sites. This potential is split into two parts. The first part, representing short ranged interactions, is the reference and often chosen to describe repulsive interactions. The second part is referred to as perturbation.

Introducing the coupling parameter λ, the pair potential reads

$$u_\lambda = u^{\mathrm{ref}} + \lambda u^{\mathrm{per}}, \tag{2.1}$$

where $\lambda \in [0,1]$ switches the perturbation on and off. $u_{\lambda=1} = u$ is the full pair potential while $u_{\lambda=0} = u^{\mathrm{ref}}$ is the reference. For any parameter, a Taylor expansion with respect to λ represents the corresponding perturbation. In equation 2.1 and in the following equations we have assumed only one type of (united-atom) interaction site. In the current case it is an average of thiol-, CH_2- and CH_3-groups of the ligands. It is straight forward to develop all equations for different individual interaction sites, but this is not necessary here. With equation 2.1, the Helmholtz energy becomes

$$A^{\mathrm{tar}} = A^{\mathrm{ref}} + A^{\mathrm{per}}, \tag{2.2}$$

where the perturbation is expanded as

$$A^{\mathrm{per}} = \left(\frac{\partial A}{\partial \lambda}\right)_{\lambda=0} \lambda + \frac{1}{2}\left(\frac{\partial^2 A}{\partial \lambda^2}\right)_{\lambda=0} \lambda^2 + \mathcal{O}(\lambda^3), \tag{2.3}$$

with $\lambda = 1$. In first order,

$$\left(\frac{\partial A}{\partial \lambda}\right)_{\lambda=0} = -\frac{1}{\beta}\left(\frac{1}{Z}\frac{\partial}{\partial \lambda}\underbrace{\int \exp(-\beta U_{N,\lambda}(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N}_{=Z}\right)_{\lambda=0} \tag{2.4}$$

$$= \frac{1}{Z^{\mathrm{ref}}}\int U_N^{\mathrm{per}}(\mathbf{r}^N)\exp(-\beta U_N^{\mathrm{ref}}(\mathbf{r}^N))\mathrm{d}\mathbf{r}^N\,, \tag{2.5}$$

where $U_{N,\lambda}(\mathbf{r}^N)$ is the total potential energy of the system and Z denotes the configuration integral and U_N^{per} is equivalently the total potential energy of the perturbation part of the potential according to equation 2.1. With the definition of the pair correlation function $g_{\alpha\beta}^{\mathrm{ref}}(\mathbf{r}_1, \mathbf{r}_2)$, where α and β denote ligand segments of NC1 and NC2 respectively, one gets

$$A = A^{\mathrm{ref}} + \frac{1}{2}\sum_\alpha\sum_\beta\iint \rho_\alpha(\mathbf{r}_1)\rho_\beta(\mathbf{r}_2)g_{\alpha\beta}^{\mathrm{ref}}(\mathbf{r}_1, \mathbf{r}_2)u^{\mathrm{per}}(r_{12})\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\,. \tag{2.6}$$

Equation (2.6) is the Helmholtz energy (as a functional of the density of ligand segments) according to first-order perturbation theory. Previous studies showed, that ligand interactions dominate effective interactions between NCs.[10] Only at small distances, core interactions have to be considered. Therefore, equation 2.6 contains the ligand segment densities ρ_α and ρ_β only. Ligand interactions can be modeled using a Lennard-Jones potential as a target potential. We see that it is crucial to choose a suitable reference, since, to apply perturbation theory, it is necessary to have knowledge of the structure (i.e. $g_{\alpha\beta}^{\mathrm{ref}}$) of the reference fluid. Therefore, we have chosen a system, where all interaction sites are represented by hard-sphere potentials. The reference fluid then represents a hard-sphere chain (index 'hsc') fluid, which is well described by Tripathi and Chapman.[14] The Helmholtz energy is then

$$A = A^{\mathrm{hsc}} + \frac{1}{2}\sum_\alpha\sum_\beta\iint \rho_\alpha(\mathbf{r}_1)\rho_\beta(\mathbf{r}_2)g_{\alpha\beta}^{\mathrm{hsc}}(\mathbf{r}_1, \mathbf{r}_2)u^{\mathrm{per}}(r_{12})\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\,. \tag{2.7}$$

To describe the reference using hard-sphere chains we need to assure that this reference fluid provides the same Helmholtz energy contribution as the reference part of our target potential. One option to achieve this is to modify the hard sphere contact distance by defining an equivalent hard sphere diameter $d(T, \rho)$.[15] It is important to note, that the temperature behavior of d depends on the division of Lennard-Jones potential into reference and perturbation. Prominent separations were proposed by Barker and Henderson[16]

and Weeks, Chandler and Andersen[17] and are not shown here.

We will show later that for our method it is not necessary to actually choose a separation distance explicitly. For a fairly large temperature range, it is sufficient to assume a constant hard-sphere diameter.

We write the dimensionless total Helmholtz energy A/kT as

$$A/kT = \underbrace{A^{\mathrm{hsc}}/kT}_{\neq f(T)} + A^{\mathrm{per}}/kT \, . \tag{2.8}$$

In terms of our method, we can simplify this expression and define two functions a and b that depend on the number of molecules and the volume only.

$$A/kT = a(N,V) + b(N,V)/kT \, , \tag{2.9}$$

where $b(N,V)$ is a temperature independent correlation given by the last term of equation 2.7. The temperature dependence through intramolecular potentials is absorbed into an ideal gas contribution which can be assumed to be equal for every center-of-mass distance. Since we are interested in the Helmholtz energy difference, its contribution vanishes.

To summarize we see that using perturbation theory for a Lennard-Jones system enables us to formulate a very simple temperature dependence of the Helmholtz energy contributions. This expression assumes a first-order perturbation theory using a constant equivalent hard sphere diameter.

Up to now, we moved along two different paths. The first one yielded the effective potential using atomistic MD simulations, the second one provided us with a functional temperature dependence of the Helmholtz energy for systems with Lennard-Jones interactions. To motivate the connection between these two paths, we consider two distances 1 and 2 of two NCs in vacuum, where both configurations are in equilibrium. We then pull the centers of mass towards each other. This way, we transfer configuration 1 into configuration 2. The work that is needed is the potential of mean force and can be formulated as Helmholtz energy difference

$$\Phi_m = A_2 - A_1 \, . \tag{2.10}$$

This expression is a simple case of Jarzynski's non-equilibrium equality.[18,19] It holds under the condition, that the transfer between configurations happens adiabatically slow. In other words, every configuration along the path between 1 and 2 has to be in equilibrium which is exactly what we establish in atomistic MD simulations. We simplify the notation in 2.10

$$\Phi_m(r_{12}) = \Delta A(r_{12}) \,, \tag{2.11}$$

where r_{12} denotes the separation distance between centers of mass of the two NCs analogue to the previous description. Now, we use the derived formula from perturbation theory 2.8 to formulate the right hand side

$$\Phi_m(r_{12})/kT = \Delta A^{\mathrm{hsc}}(r_{12})/kT + \Delta A^{\mathrm{per}}(r_{12})/kT \tag{2.12}$$

$$= a(r_{12}) + b(r_{12})/kT \,. \tag{2.13}$$

Here, we have made a transition from the variables N, V of equation 2.9 to the separation distance of two NCs, r_{12}. That is possible, because in equilibrium conditions, a given r_{12} uniquely determines the average density field and thus $\langle N \rangle$ for a defined V.

$a(r_{12})$ and $b(r_{12})$ can be calculated directly from simulations at two different temperatures

$$a(r_{12}) = \frac{\Phi_m(r_{12}, T_2) - \Phi_m(r_{12}, T_1)}{k(T_2 - T_1)} \tag{2.14}$$

$$b(r_{12}) = \frac{T_1 \Phi_m(r_{12}, T_2) - T_2 \Phi_m(r_{12}, T_1)}{T_1 - T_2} \,. \tag{2.15}$$

The elegance of this method is that it is not necessary to concern oneself with the actual decomposition of the target potential or the calculation of the Helmholtz energy contributions while still being able to utilize the theoretical framework from perturbation theory to reduce the simulation effort drastically.

2.3   Results

We investigated NCs consisting of 147 and 1415 gold atoms that form the core. Attached are alkyl thiols with 8 to 12 carbon atoms. For all systems the cores are fully loaded with ligands.[20] For each PMF, the initial separation distance was chosen sufficiently large to assure that no significant interactions occur between the two NCs. In this region the PMF is zero. The simulation results were verified in comparison with the results obtained by Schapotschnikow et al.[7]

Figure 2.1 presents the PMF from simulations for a system of two NCs consisting of 147 gold atoms with 58 ligands. The ligands consist of alkyl chains with 8 (Figure 2.1a) and 12 (Figure 2.1b) carbon atoms (notation $\mathrm{Au}_{147}(\mathrm{SC8})_{58}$ and $\mathrm{Au}_{147}(\mathrm{SC12})_{58}$). The diagrams present the PMF for varying center-of-mass distances for different temperatures.

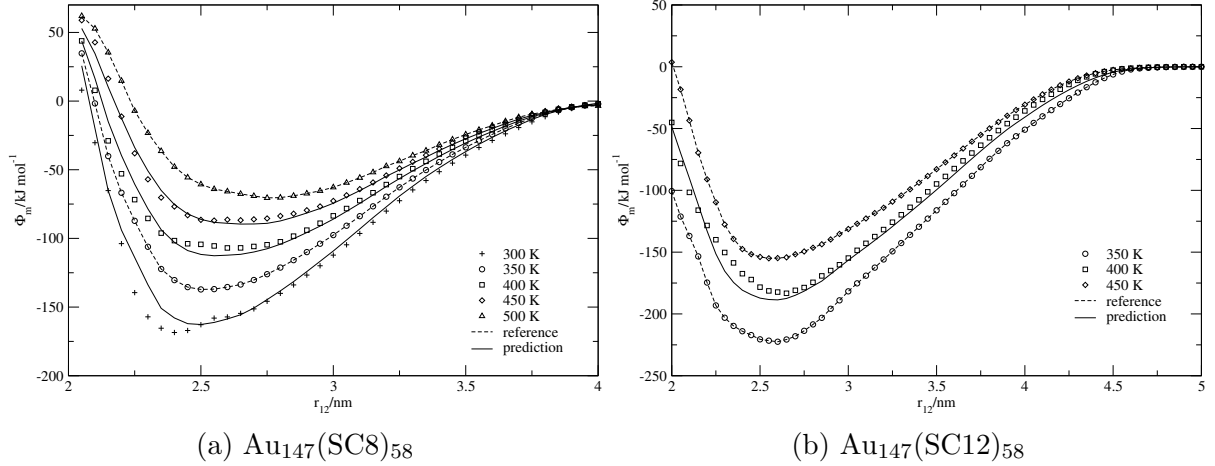(a) $Au_{147}(SC8)_{58}$ (b) $Au_{147}(SC12)_{58}$

Figure 2.1: PMF as function of center of mass separation r_{12} between two nano crystals for different temperatures in vacuum. Symbols represent data from constraint MD simulations, dashed lines represent reference PMF and solid lines denote predicted PMF with the presented method.

Every symbol denotes a total number of 10 simulations. Solid lines represent predicted potentials using the method presented in section 2.2.2. Dashed lines represent the PMF used as reference.

The predicted PMF show very good agreement in the whole region of the simulations. Even though there are minor deviations from simulation data near the minima, all predicted values lie within error bars. Error bars are exemplarily shown in Figure 2.2. For the remaining figures we omit error bars for improved clarity. The attractive region of the potential, where error bars from simulations are small, is predicted with good agreement. It is noteworthy, that the references should be chosen at temperatures where ligands have no preferred orientations which occurs in melting and freezing transitions.[21,22] Upon freezing, ligands align themselves which leads to rugged potential curves. Such behavior can be seen in Figure 2.1a for the lowest temperature 300 K. Extrapolated potentials should therefore be treated with caution since one has to make sure that no phase transition occurs within the temperature boundaries.

Moving to bigger systems, NCs consisting of 1415 gold atoms and 242 ligands with 12 carbon atoms were simulated. Figure 2.2 shows results for a system of two of these NCs. Again, the predicted PMF agrees with simulated data.

In order to assess the method for mixed NCs, we investigate a system of strongly different NC-sizes. Figure 2.3a provides results for the PMF between $Au_{1415}(SC12)_{242}$ and $Au_{147}(SC12)_{58}$. In this system, good agreement with simulation data can be observed up to approximately 4.25 nm which is the distance where perfectly elongated ligands would start to touch the core of the neighbor NC. In the simulated systems though, bending of ligands towards regions of lower density starts at slightly higher distances. The position
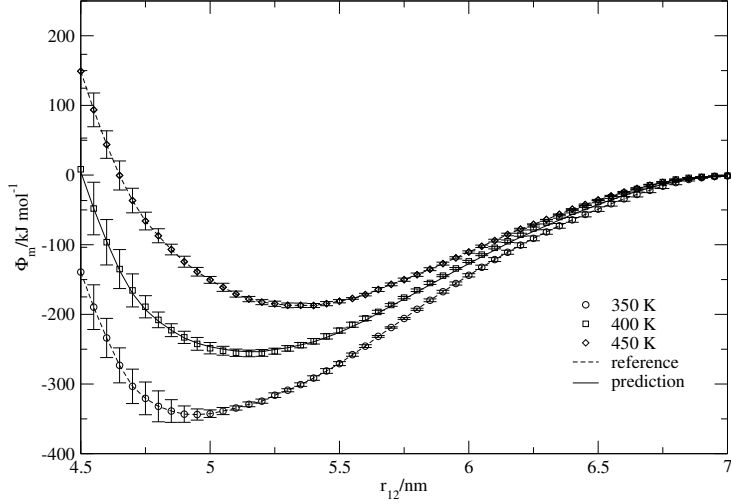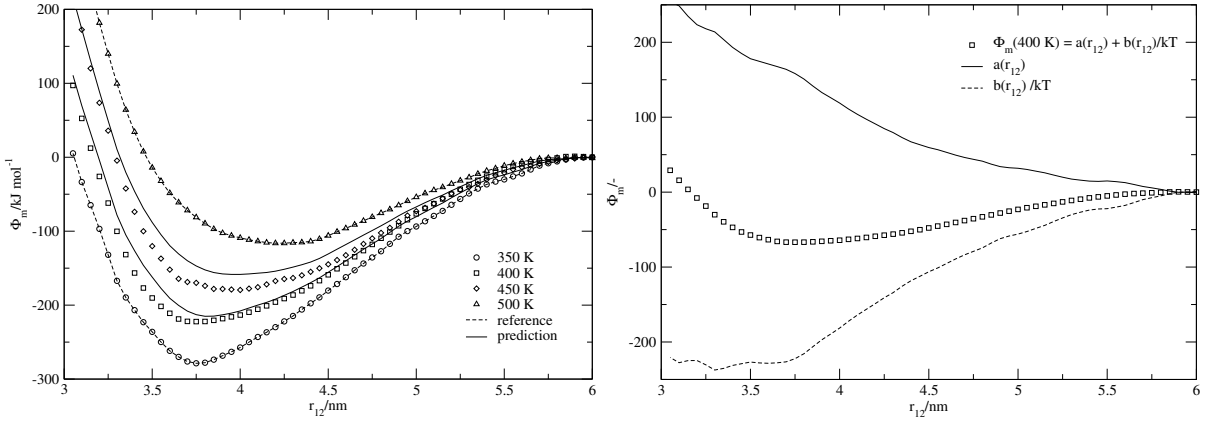
Figure 2.2: PMF for $Au_{1415}(SC12)_{242}$ as function of center of mass separation r_{12} between two nano crystals for different temperatures in vacuum.

(a) $Au_{1415}(SC12)_{242}$ - $Au_{147}(SC12)_{58}$

(b) Decomposition of the PMF into reference $(a(r_{12}))$ and perturbation $(b(r_{12}))$ contribution.

Figure 2.3: PMF as function of center of mass separation r_{12} between two nano crystals of different size (2.3a). According to the theory, the decomposition of the PMF shows a purely repulsive reference contribution (2.3b).

of the minimum in the PMF is predicted in good agreement at all temperatures.

The proposed method of predicting PMFs is based on a repulsive contribution and an attractive part to the PMF. Figure 2.3b illustrates both contributions depending on the NC-distance. The resulting repulsive contribution increases monotonically when reducing distances, as one would expect. The simulation time of this system was 32 ns for every distance. A sequence of 61 simulations was conducted to calculate the PMF, which leads to a simulation time of 1952 ns for a single temperature.

The proposed formalism has here been applied to NCs in vacuum, but should equally be applicable to treat the PMF for systems with explicit solvent. We would expect the

repulsive part of the PMF (see Figure 2.3b) to only be mildly changed by a solvent. The attractive part, in contrast, should strongly be influenced by the presence of a solvent.

2.4 Conclusion

In this article, we propose a simple method to predict the PMF between capped gold NCs in vacuum at any temperature using data from only two MD simulations. The underlying dependency of the PMF on the temperature was derived from first-order perturbation theory where we simplified the reference to be only a linear function of the temperature. The predicted PMF was compared with results from constraint MD simulations and showed very good agreement for systems of different sized and shaped NCs.

References

1. Wang, J. Carbon-Nanotube Based Electrochemical Biosensors: A Review. *Electroanalysis* **17,** 7–14 (2005).

2. West, J. & Halas, N. Engineered nanomaterials for biophotonics applications: Improving sensing, imaging, and therapeutics. *Annu. Rev. Biomed. Eng.* **5,** 285–292 (2003).

3. Shevchenko, E., Talapin, D., Kotov, N., O'Brien, S. & Murray, C. Structural diversity in binary nanoparticle superlattices. *Nature* **439,** 55 (2006).

4. Whetten, R. L. *et al.* Crystal Structures of Molecular Gold Nanocrystal Arrays. *Acc. Chem. Res.* **32,** 397–406 (1999).

5. Patel, N. & Egorov, S. A. Interactions between sterically stabilized nanoparticles in supercritical fluids: A simulation study. *J. Chem. Phys.* **126** (2007).

6. Landman, U. & Luedtke, W. D. Small is different: energetic, structural, thermal, and mechanical properties of passivated nanocluster assemblies. *Faraday Discuss.* **125,** 1–22 (2004).

7. Schapotschnikow, P., Pool, R. & Vlugt, T. J. H. Molecular Simulations of Interacting Nanocrystals. *Nano Lett.* **8,** 2930–2934 (2008).

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

9. Wang, Y., Teitel, S. & Dellago, C. Melting of icosahedral gold nanoclusters from molecular dynamics simulations. *J. Chem. Phys.* **122,** - (2005).

10. Tay, K. & Bresme, F. Computer Simulations of Two Dimensional Gold Nanoparticle Arrays: the Influence of Core Geometry. *Mol. Simul.* **31,** 515–526 (2005).

11. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).

12. Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **22,** 1420–1426 (1954).

13. Barker, J. A. & Henderson, D. Perturbation Theory and Equation of State for Fluids: The Square-Well Potential. *J. Chem. Phys.* **47,** 2856–2861 (1967).

14. Tripathi, S. & Chapman, W. G. Microstructure of inhomogeneous polyatomic mixtures from a density functional formalism for atomic mixtures. *J. Chem. Phys.* **122** (2005).

15. Rowlinson, J. The statistical mechanics of systems with steep intermolecular potentials. *Mol. Phys.* **8,** 107–115 (1964).

16. Barker, J. A. & Henderson, D. Perturbation Theory and Equation of State for Fluids. II. A Successful Theory of Liquids. *J. Chem. Phys.* **47,** 4714–4721 (1967).

17. Weeks, J. D., Chandler, D. & Andersen, H. C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **54,** 5237–5247 (1971).

18. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78,** 2690–2693 (14 Apr. 1997).

19. Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **56,** 5018–5035 (5 Nov. 1997).

20. Pool, R., Schapotschnikow, P. & Vlugt, T. J. H. Solvent Effects in the Adsorption of Alkyl Thiols on Gold Structures: A Molecular Simulation Study. *J. Phys. Chem. C* **111,** 10201–10212 (2007).

21. Luedtke, W. D. & Landman, U. Structure and Thermodynamics of Self-Assembled Monolayers on Gold Nanocrystallites. *J. Phys. Chem. B* **102,** 6566–6572 (1998).

22. Badia, A., Lennox, R. & Reven, L. A dynamic view of self-assembled monolayers. *Acc. Chem. Res.* **33,** 475–481 (2000).

Chapter 3

Three-body Effects in Triplets of Capped Gold Nanocrystals

N. Gribova was involved in the simulation setup and performed some of the presented simulations. A. Lange was involved in the development of the empirical three-body model. C. Holm had an advisory role in setting up simulations. J. Gross had the role of a daily supervisor and was involved in editing the manuscript. Additions or deletions compared to the published work are marked with angular brackets.

Three-body interactions constitute an important part of the effective potential between nanocrystals (NCs). In this study, molecular dynamics simulations are conducted on gold NCs capped with alkyl thiol ligands in vacuum. Over the course of a simulation performed in two- and three-body systems, we measure the forces acting on the cores of the NCs. These forces are then used to calculate the two- and three-body potentials of mean force (PMF). The influence of the ligand length, the size of the core, and the temperature are studied. We find that three-body effects are mainly repulsive. Longer ligand lengths and bigger core sizes further increase the strength of repulsion. According to our simulation data, the three-body contribution is independent of the temperature. Furthermore, we propose an empirical model of the three-body contribution based on the repulsive part of the two-body PMF.

3.1 Introduction

Nanocrystals (NCs) are small crystals of metals or semiconductors, often coated with ligands to prevent aggregation in a solution. They can self-assemble to a wide range of so-called superstructures. Different superstructures — even if formed from mono disperse NCs — may have completely different physical properties. The superstructure and thus the properties depend on the parameters that control the assembling process. It is fundamental to understand the assembling process and its parameters when aiming at creating materials with customized properties. Many studies of the last decades, both experimental[1–5] and theoretical[6,7], as well as studies based on computer simulations[8–16], reflect the need for a better understanding of the self-assembling process of NCs.

Because of the large number of atoms that form a NC and because of the slow dynamics of the self-assembling process one cannot reproduce a regular lab-experiment on the formation of superstructures within an atomistic simulation. A common method to circumvent the problem of too large sizes and too long simulation times is to determine effective pair potentials from atomistic simulations of two isolated NCs (referred to as two-body system). The resulting potential of mean force (PMF) represents each NC as a single interaction site, allowing many-particle simulations that can reach much longer time scales. However, there are phenomena that cannot be explained by two-body interactions alone. For example, from experiments[5] it is known that inter-particle distances in superstructures are different for varying ligand lengths, yet simulations[11] of two-body systems had shown that the equilibrium distance does not depend on the ligand length. One possible explanation of these findings is the strong influence of three-body or in general – multi-body – interactions. Schapotschnikow et al.[13] investigated the effect of the ligand length on the effective three-body potential of gold NCs in vacuum. In their simulations, they considered three NCs on the corners of equilateral triangles and evaluated the effect of three-body interactions on the two-body potential. They found that the corrected two-body potentials have higher equilibrium distances and higher energy minima. According to their observations, the alignment of three NCs into a chain is energetically favored over a more compact structure. They argue that a triangular configuration has a disadvantage due to areas of high ligand density or an interpenetration of overlapping ligand spheres. Also, they present the so-called Overlap Cone Model (OCM) which is a comparably simple geometric model to estimate the equilibrium distance between multiple NCs based on idealized ligand packing.

Very recently, Boles and Talapin investigated the role of ligand arrangement concerning equilibrium distances within NC superstructures.[17] They argue, that many-body effects arise due to deformations of the ligand layers, an effect that is not captured within the OCM. The degree of deformation depends on the ratio of ligand length and core size. Also

the total number of neighbors plays a determining role towards the equilibrium distance.

These studies clearly show that multi-body effects are an important factor concerning NC systems. Similar to the simulations of Schapotschnikow et al., we present a new method for obtaining the three-body PMF of gold NCs in vacuum. Furthermore, we suggest how to parameterize the three-body correction from the two-body PMF which can then be used within coarse grained simulations in the spirit of conventional three-body potentials. We discuss, how parameters defining shape and size of NCs including the ligand layer as well as the temperature affect the PMF. A detailed study of the formation and stability of superstructures including three-body effects is thus possible and will be subject of future examinations.

3.2 Methods

In this section we present the general methods of our study. We start with a molecular description of the NCs. Then, details of the simulation workflow as well as the simulation parameters are presented. We give a short recapitulation of how two-body PMF are obtained and expand the theory to systems with three NCs.

3.2.1 NC model

The NCs studied here consist of a gold core and a capping layer of alkyl thiol ligands (cf. Fig. 3.1). They are defined by three parameters: the core size (i.e., the number of gold atoms or the core radius or diameter, r_c and d_c respectively), the number of ligands and the alkyl type, which can either be specified by the number of carbons or the ligand length (denoted by l, the length of a ligand when all bond angles are at their equilibrium value). In literature, the capping layer is often described as spherical corona and interactions between NCs arise when their coronae interpenetrate.[10][13]

The gold cores are modeled as rigid icosahedrons formed by a well-defined number of gold atoms. Determining stable structures of the gold cores is still an active field of research. We chose icosahedral structures because they were found to be stable both, in experiments[18] and in simulations[19] for a wide range of core sizes.

In our study, the cores consist of 147 ($d_c \approx 1.8$ nm) or 1415 gold atoms ($d_c \approx 3.7$ nm). The number of ligands adsorbed onto the gold core is maximal; the onset of saturation depends on the solvent and the temperature. Throughout the paper, we adopt the notation of Schapotschnikow et al.[11] abbreviating $Au_{1415}(SC12)_{242}$, which refers to a core consisting of 1415 gold atoms with 242 1-dodecanethiol ligands (Fig. 3.1).

To describe the interactions between the ligands and between the ligands and the gold core we use the united atoms model where the chemical groups SH, CH_2, Au and CH_3 are
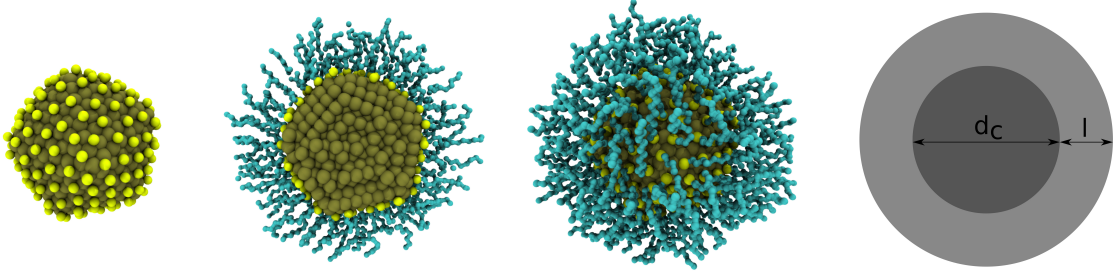
Figure 3.1: United atom and schematic representation of an $Au_{1415}(SC12)_{242}$ NC at temperature $T = 350$ K (molten ligand layer). Brown spheres represent gold atoms, yellow spheres the thiol head group and cyan lines the carbon chains. From left to right: gold core with thiol head groups, full NC cut in half, full NC and schematic spherical representation in terms of core diameter and ligand length.

Table 3.1: Lennard-Jones force field parameters for ligand-ligand as well as ligand-gold interactions used in this study. All parameters were taken from Ref.[12]

ϵ_{ij}[K]	CH_3	CH_2	SH	Au
CH_3	108	78	117	108
CH_2	78	56	84	88
SH	117	84	126	2795
Au	108	88	2795	-

σ_{ij}[Å]	CH_3	CH_2	SH	Au
CH_3	3.76	3.86	4.11	3.54
CH_2	3.86	3.96	4.21	3.54
SH	4.11	4.21	4.45	2.65
Au	3.54	3.54	2.65	-

represented as single interaction sites. Intermolecular interactions between said groups are modeled via a truncated ($r_c = 12.0$ Å) and shifted Lennard-Jones potential. The force field parameters are taken from Pool et al. (see Table 3.1). Since we model the gold cores as rigid, we do not include interactions between the gold atoms. Consequently, there are no interactions between the gold cores. These interactions are small compared to the ligand interactions and therefore we do not consider them which is an approximation that is valid as long as the capping layer is intact, with thiols staying on the core surface during the whole simulation.

Our simulations are performed within a certain temperature range — between 350 K and 450 K. It is noteworthy that, due to the large surface-to-volume ratios, small cores (147 gold atoms) may undergo a melting transition at those temperatures, which, according to the studies of Wang and authors[20], results in a structural transition from icosahedral to spherical shapes. We discuss the role of the core shape with respect to the

PMF in the following section, but we note that a rigid model for the gold core formally excludes any melting transitions.

3.2.2 Simulation details

The PMF is obtained by integration over the forces between the NCs evaluated at different center of mass distances and is therefore a function of the core-to-core separation distance. There are different simulation methods for computing the PMF. For systems of gold NCs, a comparison of methods is given by Schapotschnikow et. al.[11] For a more general discussion we refer to the work of Trzesniak et. al.[21] In the following we give a brief overview.

An unconstrained method, such as umbrella sampling[22], controls the separation distance of the NCs by introducing a biasing function acting on the NCs' centers of mass. Often, the whole region of interest (the whole range of separation distances) is divided in so-called windows. That way, each window can be simulated separately and the whole region can be restrained to small ranges.[23]

Monte Carlo methods lead to comparably simple implementations for constraint simulations. In constraint simulations, the distance of the NCs is not varied (within a sampling period). Within the Monte Carlo framework, only the configurational degree of freedom is sampled. Configurations are generated via translational displacement and regrow moves of the ligands via configurational bias methods[24] while the whole crystal is able to rotate around the fixed center of mass position of the core. In MD simulations one enforces (holonomic) constraints through a set of Lagrange multipliers that must be solved during the integration. Two widely used methods for determining the Lagrange multipliers by iteration are the SHAKE[25] and the RATTLE[26] algorithms, which are both implemented in many simulation packages. An alternative MD method is steered MD. The basic idea of steered MD is the transformation of the system between two states by performing work on the system. It is a non-equilibrium method, where the work performed on the system is not reversible. The relation between the performed work and the free energy between the two states of the system is given by Jarzynski's equality.[27,28] For the system of this study, one could perform simulations where the NCs are pulled towards each other and the work is sampled over the separation distance. A very good overview of this method is given in Ref.[29].

We now describe the method used to obtain the three-body PMF in this study. The general work flow is comparable to that of the two-body system described in our earlier work.[30] All of our simulations are performed in a NVT ensemble with the GROMACS package, version 4.6.[31] The setup of a simulation sequence consists of three steps.

First, we create a NC-triplet configuration by manually placing three NCs into a simulation box. In our work flow, we set NC 1 and NC 2 to a desired center of mass distance and place the third NC so that the corresponding two-body PMF with NC 1

and NC 2 are essentially zero [see chapter 2.3].We will refer to the distance, where the two-body PMF is zero as $r_{ij,0}$. The triplet is arranged in the center of the box, and the size of the box is chosen so that all distances to periodic images are greater or equal than $r_{ij,0}$. The initial orientation of the NCs is obtained by random rotations. Only after performing these rotations the NCs are placed into the simulation box.

The second step is an energy minimization of the configuration conducted with a steepest descent algorithm.[31] Furthermore, we perform short (2 ns) MD runs in the NVT ensemble to equilibrate the system. In the equilibrated configuration we then reduce the center of mass distance between the gold cores of NC 2 and NC 3. This is done manually by shifting all positions of NC 3 along the connecting vector while the orientation of all NCs and the box stays the same (without rotation when shifting). We then relax the configuration again using energy minimization and short equilibration. The procedure is repeated for multiple distances, and the resulting configurations are used as starting configurations for the production runs.

The last step is the actual MD simulation (in a NVT ensemble). As the starting configurations are created beforehand, all simulations (for every distance) can be performed by parallel computation (the number of distances equals the number of simulations). For the production simulation, we use a leap-frog stochastic dynamics integrator which also acts as a thermostat.[31] We use the inverse friction constant $\tau_t = 2$ ps and the time step $\Delta t = 0.002$ ps in our simulations, in accordance to the GROMACS manual. A simulation run consists of 5 ns for small systems (Au_{147}) and 10 ns for large systems (Au_{1415}). We capture the forces within 5 independent runs (one run or sequence consists of steps one to three) and determine the mean value at every distance. The PMF is then calculated via integration of the mean values.

In our simulations, the center of mass distances between the three gold cores are kept constant by freezing the gold atoms. In doing so, we constrain their positions and their momenta so that, effectively, the equations of motion are only solved for the ligands. Such an approach is justified by the following assumptions: (a) rotations of the gold cores happen on a timescale larger than the simulation time; in practice, rotations are inhibited due to steric interactions of ligands at small separation distances.[11] (b) the relative orientation of the two icosahedral cores does not influence the resulting PMF. In fact, the work of Tay and Bresme[14] suggests that — at least for small ratios of core diameter over ligand length — the shape does not influence the NC-NC interaction potential. This simulation method was tested in earlier work [see chapter 2] for two-body systems and led to similar results compared to those in Ref.[11]
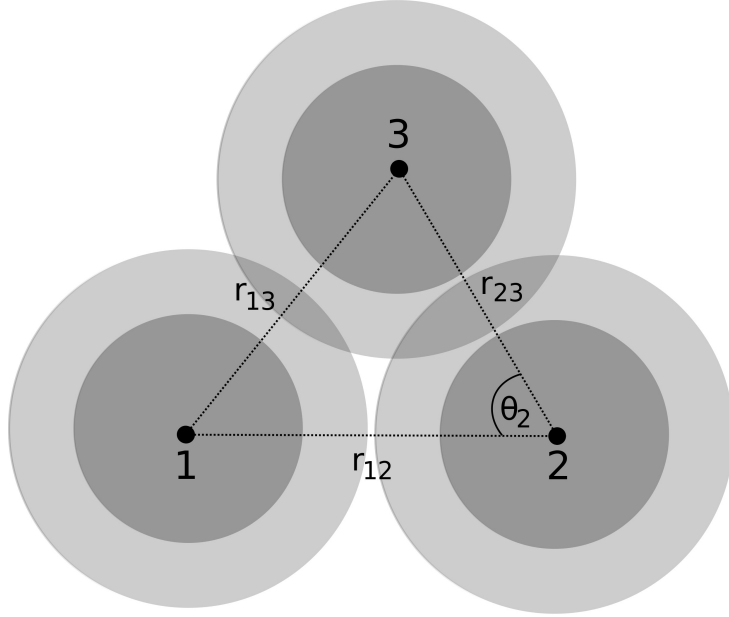
Figure 3.2: Schematic representation of a NC triplet. The configuration within the simulation setup is characterized by the separation distances r_{12}, r_{23}, and the angle θ_2.

3.2.3 Three-body configurations and PMF from forces

In a system with two NCs (denoted '1' and '2') the two-body PMF, ϕ^{2b}, is calculated by integration of the mean force $\vec{F}_m^{2b}$ acting along the separation distance $r_{12} = |\vec{r}_2 - \vec{r}_1|$,[32]

$$\phi^{2b}(r_{12}) = \int\limits_{r_{12}}^{\infty} -\vec{F}_m^{2b} \cdot \vec{e}_{12} \mathrm{d}r'_{12} \,, \tag{3.1}$$

where $\vec{e}_{12} = \vec{r}_{12}/r'_{12}$ is the unit vector connecting the centers of mass of the two NCs. The PMF is a function of the center of mass separation distance r_{12}. To obtain $\phi^{2b}(r_{12})$ from simulations, we fix two NCs at different separation distances and sample the forces acting on their centers of mass. The ensemble average of the mean force from simulations is then integrated according to Eq. (3.1).

We also introduce a third NC (denoted '3'). To describe the configuration of three NCs one needs two additional parameters: two separation distances, two angles, or a combination thereof. We start employing the first alternative (i.e., we utilize three separation distances all together).

The *total* force acting on '3' in a NC triplet is given by the sum of two-body forces and a three-body force,

$$\vec{F}_3 = \vec{F}_{3,1}^{2b} + \vec{F}_{3,2}^{2b} + \vec{F}_{3,12}^{3b} \,. \tag{3.2}$$

$\vec{F}_{3,1}^{2b}$ and $\vec{F}_{3,2}^{2b}$ denote the two-body forces on '3' induced by '1' and '2', respectively. $\vec{F}_{3,12}^{3b}$ denotes the three-body force on '3' in the presence of '1' and '2'. The three-body force is given by the negative gradient of the three-body potential, $\phi^{3b}(r_{12}, r_{13}, r_{23})$. Note that the three-body potential as we defined it above is a 'correction' potential. It does not describe the total interaction energy of the triplet configuration, but only an additive three-body contribution (in addition to the pair potentials). In the following, we refer to $\vec{F}_{3,12}^{3b}$ and $\phi^{3b}(r_{12}, r_{13}, r_{23})$ as three-body force and potential, respectively while $\vec{F}_{3}^{3b}$ and its integration, ϕ_3, we refer to as *total* three-body force and potential, respectively. By applying the chain rule this force can be represented in terms of the two forces, $\vec{F}_{31,2}$ and $\vec{F}_{32,1}$, on '3' (exerted by '1' in presence of '2' and exerted by '2' in presence of '1', respectively), according to

$$\vec{F}_{3,12}^{3b} = -\frac{\partial \phi^{3b}}{\partial \vec{r}_3} \tag{3.3}$$

$$= -\frac{\partial r_{13}}{\partial \vec{r}_3}\frac{\partial \phi^{3b}}{\partial r_{13}} - \frac{\partial r_{23}}{\partial \vec{r}_3}\frac{\partial \phi^{3b}}{\partial r_{23}} \tag{3.4}$$

$$= -\vec{e}_{13}\frac{\partial \phi^{3b}}{\partial r_{13}} - \vec{e}_{23}\frac{\partial \phi^{3b}}{\partial r_{23}} \tag{3.5}$$

$$= \vec{F}_{31,2} + \vec{F}_{32,1} \, , \tag{3.6}$$

where $\vec{e}_{13}$ denotes the unit vector between the centers of mass of '1' and '3'. Finally, the three-body potential is obtained by integration of Eq. (3.3),

$$\phi^{3b} = \int\limits_{\vec{r}_3}^{\infty} -\vec{F}_{3,12}^{3b}(\vec{s}) \, \mathrm{d}\vec{s}. \tag{3.7}$$

Conceptually, the three-body case is very similar to the two-body case. However, as configurations are determined by three parameters, the computational effort largely increases.

In our simulations, we characterize the configuration utilizing another set of parameters. We employ the center of mass separation distances, r_{12}, r_{23}, and the angle θ_2 (cf. Fig. 3.2). During a simulation run, θ_2 and $\vec{r}_{12}$ are constant. As in the two-body case, we achieve this by freezing the cores of the NCs. Successively, '3' approaches '1' and '2' within a sequence of simulations. The three-body force on '3' is given by

$$\vec{F}_{3,12}^{3b}(r_{23}; r_{12}, \theta_2) = -\left.\frac{\partial r_{23}}{\partial \vec{r}_3}\frac{\partial \phi^{3b}}{\partial r_{23}}\right|_{r_{12}, \theta_2} \tag{3.8}$$

$$= -\left.\vec{e}_{23}\frac{\partial \phi^{3b}}{\partial r_{23}}\right|_{r_{12}, \theta_2} , \tag{3.9}$$

where '3' moves along the separation distance r_{23} while r_{12} and θ_2 are kept constant; the partial derivatives with respect to r_{12} and θ_2 thus vanish. The integration is performed along $\mathrm{d}\vec{r}_{23} = \mathrm{d}(\vec{e}_{23}|\vec{r}_{23}|) = \vec{e}_{23}\mathrm{d}r_{23}$,

$$\phi^{3b}(r_{23}; r_{12}, \theta_2) = \int_{r_{23}}^{\infty} -\vec{F}_{3,12}^{3b} \cdot \vec{e}_{23}\mathrm{d}r'_{23}\,. \tag{3.10}$$

$\vec{F}_{3,12}^{3b}$ is the force caused by the three-body potential. This force on '3' is not directly accessible from simulations as one can only calculate the *total* force $\vec{F}_3$ including two-body forces. Therefore, we subtract the two-body forces obtained in simulations of two-body systems,[30]

$$\phi^{3b} = \int_{r_{23}}^{\infty} -\left(\vec{F}_3 - \vec{F}_{13}^{2b} - \vec{F}_{23}^{2b} \right) \cdot \vec{e}_{23}\mathrm{d}r'_{23} \tag{3.11}$$

$$= \int_{r_{23}}^{\infty} -\vec{F}_3 \cdot \vec{e}_{23}\mathrm{d}r'_{23} \tag{3.12}$$

$$+ \int_{r_{23}}^{\infty} \vec{F}_{23}^{2b} \cdot \vec{e}_{23}\mathrm{d}r'_{23} \tag{3.13}$$

$$+ \int_{r_{23}}^{\infty} \vec{F}_{13}^{2b} \cdot \vec{e}_{23}\mathrm{d}r'_{23}\,. \tag{3.14}$$

The integrand in Eq. (3.11) is the three-body contribution projected onto the connecting vector $\vec{r}_{23}$. After calculating the *total* three-body force, two-body forces are subtracted and the resulting force is projected onto the direction given by $\vec{r}_{23}$. Then the projected force is integrated.

A brief summary of our methodology reads as follows. The two- and total three-body forces are obtained from sequences of simulations. The center of mass positions of the NC cores are constraint by freezing the gold atoms. The three-body PMF is calculated and integrated according to Eq. (3.11). The whole procedure is repeated for different configurations (values of r_{12} and θ_2) and NC shapes (core size and ligand length) as well as temperatures.

3.3 Results and discussion

We start with the two-body PMF. In vacuum, the PMF can be divided into two parts: a short-range repulsion and a long-range attraction. When increasing the temperature, we observe a shift of the equilibrium distance towards larger separation distances. The depth of the potential well decreases for higher temperatures, leading to lower slopes of the attractive and repulsive regions, as Fig. 3.3 (a) shows. However, when changing the ligand length we do not observe a change of the equilibrium distance (Fig. 3.3 (b)). The

range of the potential increases towards longer ligands while the slopes of the attractive and repulsive regions are similar. When comparing NCs of different core size we find that bigger cores lead to stronger repulsive potentials while the range of the potential is smaller (in dimensionless scale).
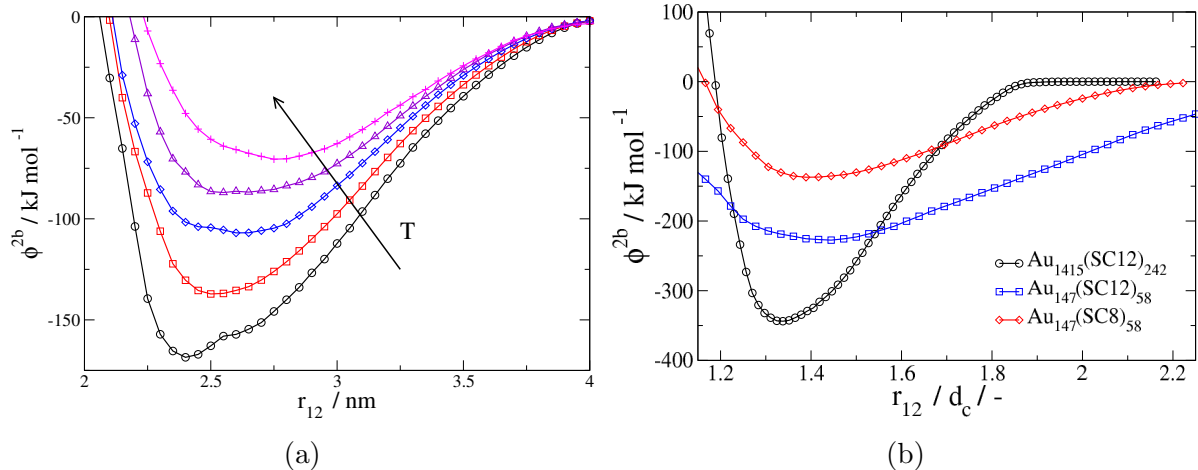

(a)

(b)

Figure 3.3: Potentials of mean force between two NCs in vacuum. Here we consider $Au_{147}(SC8)_{58}$. The left-hand side panel shows the PMFs for different temperatures. With increasing temperature (by 50 K), the equilibrium distance of the potential increases while the steepness of the attractive and repulsive regions decreases. At the lowest temperature (300 K) the chains of the NC begin to freeze, leading to rugged PMF curves. The right-hand side panel shows the PMF for different ligand lengths and different core sizes. Increasing the ligand chain lengths lead to longer ranged PMFs yet similar equilibrium distances.

For the three-body systems, we explain the generation of the potential curves using the example in Figure 3.4, which shows simulation data for a triplet of $Au_{147}(SC12)_{58}$ at $T = 350\,\text{K}$. The separation distance r_{23} (abscissa) of the NCs is plotted against the PMF (ordinate). Every symbol represents a particular triplet configuration. Symbols connected by a line represent a specific separation distance between two NCs and a specific angle at which the third NC approaches the two NCs. The distance between the two NCs affects the strength of the three-body interaction. The corresponding relationship is illustrated in Figure 3.4a, where different curves represent different values of r_{12} at constant angle $\theta_2 = 60°$. For large distances r_{12}, we observe small and purely repulsive potentials. For smaller distances r_{12}, we observe weak attractive potentials and a steeper profile, which lead to stronger repulsion for small r_{23}. A similar relationship can be observed for the approaching angle θ_2. For small angles ($\theta_2 = 60°$), we observe mildly attractive and steeply repulsive profiles. For larger angles ($\theta_2 = 90°$), the PMF decreases, and for $\theta_2 = 120°$ we see a non-vanishing contribution only at very small distances r_{23} (cf. Fig. 3.4b).

We can now examine the influence of NC core size and of ligand length on the three-
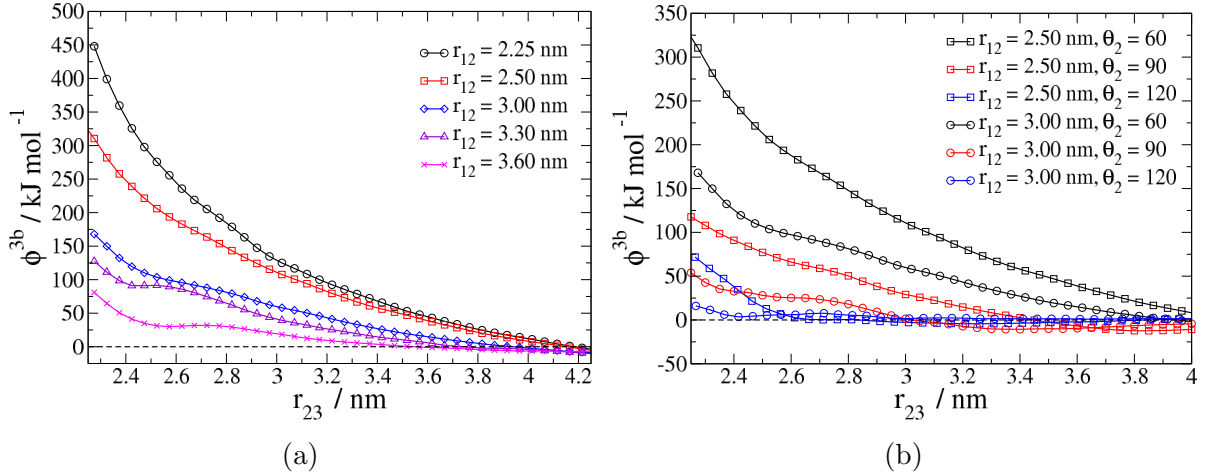
Figure 3.4: Triplets of $Au_{147}(SC12)_{58}$ simulated for different angles θ_2 ($= 60°$ left, $= 60°$, $= 90°$ and $= 120°$ right) at T = 350 K. Different curves represent varying center of mass distances r_{12}. By increasing r_{12} or the angle θ_2 one decreases the repulsive three-body energy. For $\theta_2 = 60°$ and $r_{23} = 2.50\,nm$, the three NCs form equilateral triangle. In this configuration, the contribution of three-body potential to the potential energy of the triplet is almost 30 %. (For the two-body system, the energy at the equilibrium distance $r_{12} = 2.50\,nm$ reads $\phi_{eq} = -225\,^{kJ}/_{mol}$.)

body potential. Figure 3.5 shows the integrated *total* force and the three-body PMF. Considered are two NCs of different size covered by equal ligands. The dimensionless separation distances on the abscissa are scaled as multiples of the core size. One observes that the scaled range of the potential is smaller for the NCs with the larger core — a behavior similar to the two-body system. As opposed to the two-body PMF, however, the resulting three-body PMF has the same magnitude. Consequently, three-body effects are stronger in a system with larger $\frac{l}{d_c}$ ratio, because their two-body PMFs are comparably smaller. The influence of the ligand length is shown in Fig. 3.6. Shorter ligands (smaller $\frac{l}{d_c}$ ratio) lead to weaker three-body effects.

We continue exploring how the three-body potentials depend on the configuration $(r_{12}, r_{23}, \theta_2)$ of the three NCs. At separation distances smaller than the core diameter plus two times the length of the ligands, $r_{12} < d_c + 2l$, the capping layers start to interpenetrate. We observe that the coronae of the NCs remain almost spherical (ligands pack themselves without bending) until distances smaller than the core diameter plus the length of the ligands, $r_{12} < d_c + l$. Starting at this distance, ligands heavily bend away from the connecting axis, as Fig. 3.7 shows. For small angles (here 60 degrees), we observe that ligands form bundles in the region enclosed by the NCs. Going to even more compact configurations (decreasing r_{12}), ligands are pressed out of the enclosed region between the NCs completely as can be seen in the rightmost panel of Fig. 3.7.

Fig. 3.8 shows that three-body potentials lead to higher equilibrium distances than can be explained by only two-body potentials. We analyse the PMF by decomposing the
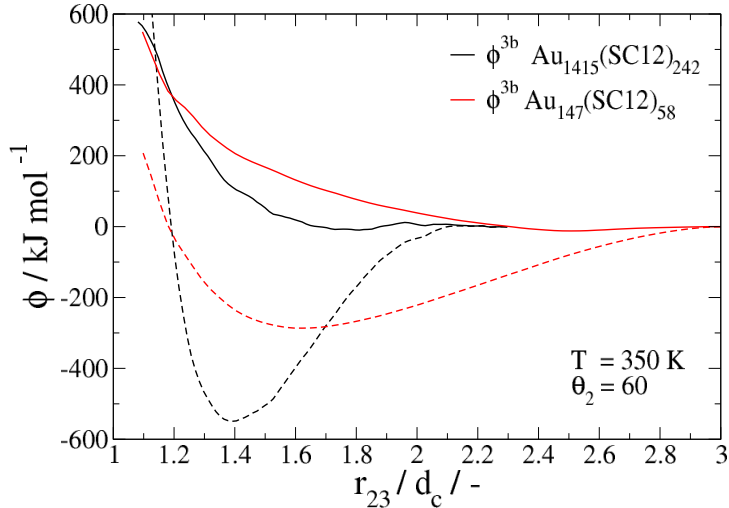
Figure 3.5: Three-body potential (solid lines) and *total* integrated forces (dashed lines) for two systems with different core size plotted over the separation distance divided by the core size ($d_{c,1415} = 3.7\,\mathrm{nm}$, $d_{c,147} = 1.8\,\mathrm{nm}$). For both potentials, the angle is $\theta_2 = 60\,^\circ$ and the separation distances between '1' and '2' are set to the two-body equilibrium distance ($r_{12} = r_{eq}$). Like in two-body systems, bigger core sizes reduce the range of the PMF. The three-body PMF shows a steep repulsion.

different contributions. The solid curve in Fig. 3.8 represents the PMF, ϕ^{3b}. Symbols are the projected and integrated forces along $\vec{r}_{23}$ (denoted by the index ",23"). The integrated *total* forces on NC3, $\phi_{3,23}$ (including the two-body contributions, according to Eq. (3.10)), are given by squares. The integrated two-body forces (the two-body PMFs) between NC2 and NC3 are given by diamonds and between NC1 and NC3 by triangles. This decomposition shows that ϕ_3 and the two-body PMF between NC1 and NC 3, $\phi^{2b}_{13,23}$, start attaining finite values at the same distance while the two-body PMF between NC2 and NC3 (which is evaluated along $\vec{r}_{23}$, hence no projection), ϕ^{2b}_{23}, remains zero. This leads to an effective attraction as $\phi_{3,23}$ has a steeper slope.

We investigated the PMF for different temperatures and had an important finding. Results for simulations at $T = 350$ K and $T = 450$ K are shown in Figure 3.9. The results confirm that the three-body PMF ϕ^{3b} is independent of the temperature within the range to good approximation. We note, that the integrated *total* force strongly varies with temperature. But this temperature dependence is due to the two-body contributions, according to Eq. (3.11), whereas the ϕ^{3b} contribution does not vary with temperature. A temperature independent three-body contribution is a great simplification in modeling various stable forms of NC-superstructures, because it makes atomistic molecular simulations of three-NC configurations for every temperature dispensable. We can make an attempt to provide rational for the nearly temperature independent three-body corrections. The three-body corrections are a result of increased ligand-density between three particles approaching one another. One can consider the three-body corrections a result
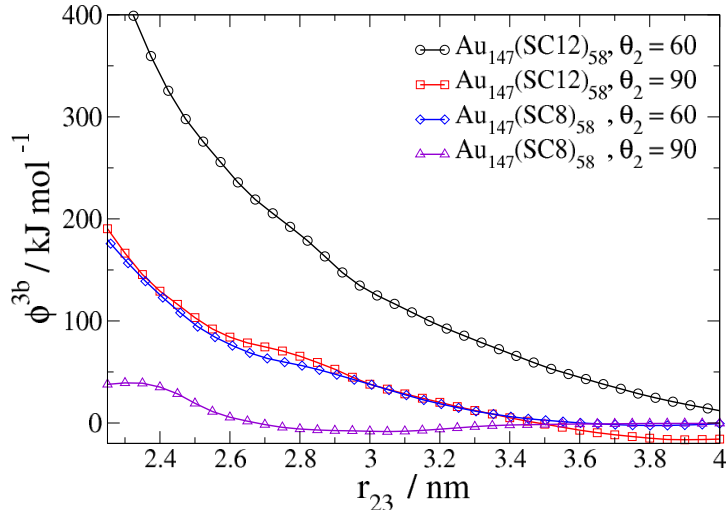
Figure 3.6: PMFs for NCs with different ligand lengths for the configuration $r_{12} = 2.25\,\mathrm{nm}$ and different angles. Shorter ligands (SC8) lead to smaller and shorter ranged contributions to the three-body PMF. For $\theta_2 = 90\,^\circ$ the PMF for the small ligands deviates from zero only for small distances.

of ligand structure. Like other structural properties, say the radial distribution function or three-particle distribution functions in simple fluids, the structure of the ligand cloud at sufficiently high density is weakly temperature dependent.

3.4   Parameterization of the three-body PMF

The determination of three-body potentials from atomistic simulations is a tedious task especially if one is also interested in covering the dependence of the different (molecular) parameters. We thus strive for a simple analytical model that represents the three-body PMF ϕ^{3b}.

We found that the three-body PMF is dominantly repulsive. This observation is important not only because it relaxes us from the need to determine three-body potentials for every temperature, but also because it guides us in proposing an empirical model for the three-body PMF. We assume the three-body PMF term scales with the repulsive part of the two-body PMF, ϕ^{2b}_{rep}. In our previous work [see chapter 2], we proposed to decompose the two-body potentials ϕ^{2b} into a repulsive part ϕ^{2b}_{rep} and an attractive portion ϕ^{2b}_{att}, by applying a first order expansion of the Helmholtz energy with respect to the inverse temperature. For our three-body model, we propose a symmetric combination of those repulsive contributions,
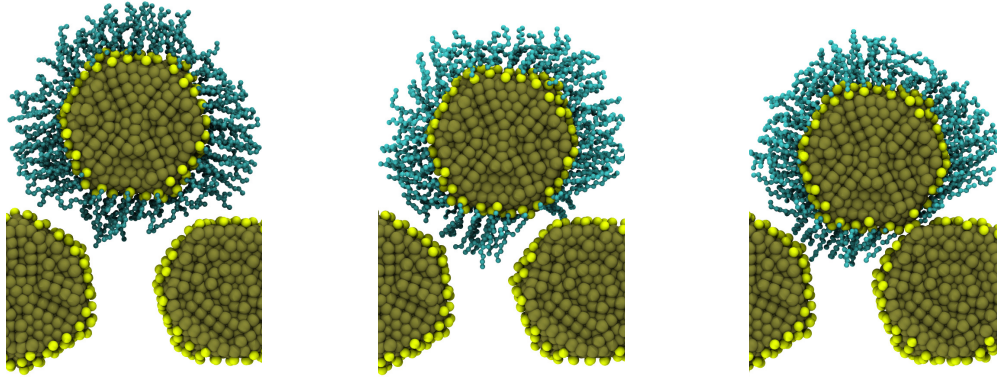
Figure 3.7: The deformation of the ligand layer increases when NC3 approaches the other two (fixed) NCs. For clarity, the whole configuration is sliced at the center of mass position of the cores and ligands of NC1 and NC2 are faded out. All snapshots are taken from simulations at $T = 350$ K and $r_{12} = 5.50$ nm. From left to right, the distance r_{23} is 5.00 nm, 4.50 nm, 4.00 nm respectively.

$$\phi^{3b} = A \cdot \sin\theta_1 \sin\theta_2 \sin\theta_3 \cdot$$
$$\times \left(\phi_{rep}^{2b}(r_{12})\phi_{rep}^{2b}(r_{13})\phi_{rep}^{2b}(r_{23}) \right)^{\alpha} . \tag{3.15}$$

Parameters A and α can be determined from three-body simulations. Because with this ansatz, Eq. (3.15), we treat the three-body PMF as being solely repulsive, we are not able to reproduce the small attractive contributions we observed within the simulations.

The simulation results presented above were used to determine the model parameters, A and α. They were determined from the simulated PMF curves using a least squares method. Some deviation of the model, Eq. (3.15), from the three-body PMF from molecular simulations becomes apparent in the r_{23}-range of 2.5 nm to 2.9 nm. The non-smooth course of the model is due to the non-smooth repulsive two-body potential ϕ_{rep}^{2b}. Given the wide range of different $(r_{12}, r_{23}, \theta_2)$-configurations covered in Fig. 3.10, however, we assess the agreement of the model with the molecular simulation data as good. We note also, that the two parameters A and α can be determined from two simulations. We observed that the results of Eq. (3.15) (and the values of A and α) are not overly sensitive to the choice of the selected $\phi^{3b}(r_{ij})$-curve for $\theta_2 = 60°$.

3.5 Conclusion

In this work we analysed the existence and influence of three-body effects in the system of capped gold NCs in vacuum. Our study shows that the three-body effects are mainly
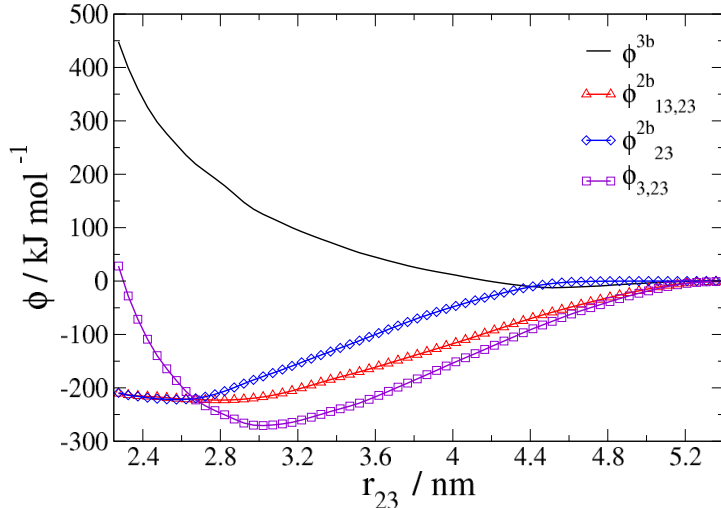
Figure 3.8: Contributions to the three-body PMF for the configuration $r_{12} = 2.25$ nm, $\theta_2 = 60°$ at $T = 350$ K. The black solid curve without symbols represents the resulting three-body PMF (i.e. the result of equation (3.11)) while the magenta curve with squares is the result of the integration of the *total* force measured in the simulation. The red(triangle) and blue(diamond) lines are the integrated two-body forces, which we subtract from the *total* force to get the three-body force. The second subscript '23' indicates that the forces from simulations are projected onto $\vec{r}_{23}$. There is a small region ($r_{12} > 4.50$ nm) where we observe the three-body PMF to be attractive.

repulsive and arise for short separation distances. We further find that the magnitude scales with the ligand length and size of the NC core while it is independent of the temperature. A model to describe the PMF of NC triplets was presented. It is based on a combination of repulsive potentials which can be extracted from two-body PMF. Additionally, two model parameters have to be determined by three-body simulations. To obtain the three-body PMF, we fixed three NCs in appropriate arrangements and measured the forces acting on their centers-of-mass. We then subtracted two-body PMF to obtain a correction potential which can be used to describe triplet configurations in coarse-grained simulations in addition to two-body potentials. We hope that our findings will facilitate the coarse-grained simulations of superstructure assembly with effective two- and three-body potentials.

The present study was limited to monodisperse NCs and we did not discuss the influence of a solvent as well as PMFs between different sized NCs, but the simulation method can be applied to such systems analogously. Solvents can strongly affect the PMF, and consequently the resulting superstructure and simulations of NCs in vacuum only provide a scenario for very poor solvent conditions. Simulations with explicit solvent, however, are computationally more demanding and the number of parameters grow (solvent type, density) while implicit solvent simulations lack information such as the arrangement of solvent molecules around NCs. Simulations for binary PMF in vacuum show to be promis-
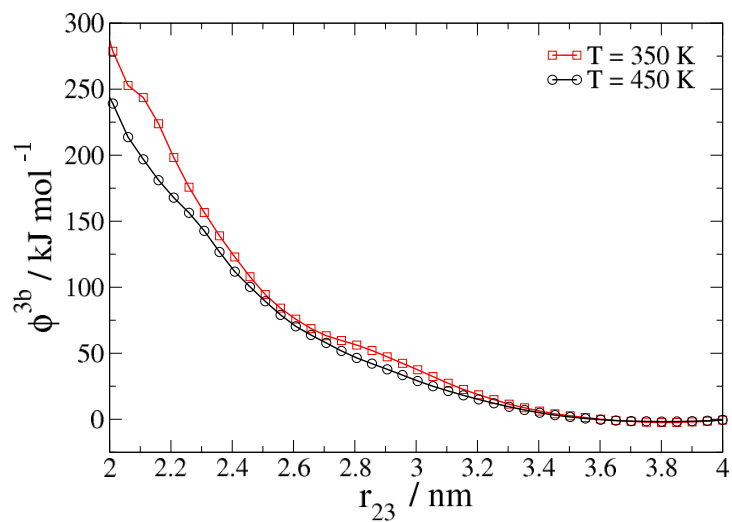
Figure 3.9: Three-body potentials of an $Au_{147}(SC8)_{58}$ triplet for two different temperatures. The PMF is temperature independent. The integrated *total* forces show similar temperature behavior like the two body system (not shown). Subtracting the two-body forces eliminates this dependence.

ing in that the two-body PMF can be obtained quite simply from two-body PMFs of pure components via mixing rules.
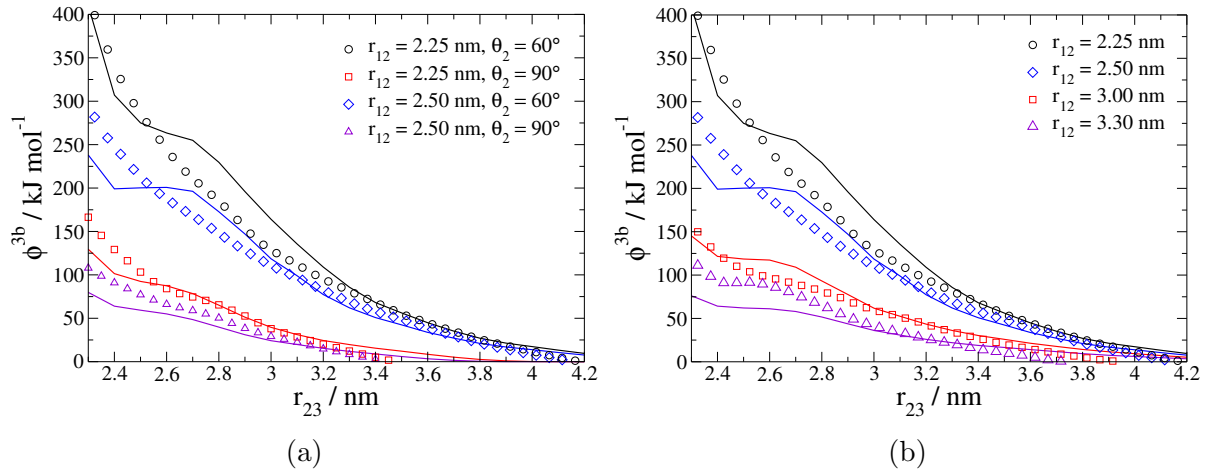
Figure 3.10: Three-body PMF from atomistic simulations (symbols) and modeled via equation (3.15) (solid lines) for the system $Au_{147}(SC12)_{58}$ at T = 350 K for varying angles (3.10a) and distances (for $\theta_2 = 60°$ between NC1 and NC2 (3.10b). To calculate the model parameters, data from two simulations were used. Model parameters: $A = 1.0339 \cdot 10^{-7}$ kJ/mol, $\alpha = 1.32297$.

References

1. Miszta, K. *et al.* Hierarchical self-assembly of suspended branched colloidal nanocrystals into superlattice structures. *Nat. Mater.* **10,** 872–876 (Nov. 2011).

2. Shevchenko, E., Talapin, D., Kotov, N., O'Brien, S. & Murray, C. Structural diversity in binary nanoparticle superlattices. *Nature* **439,** 55 (2006).

3. Schaeffer, N., Wan, Y. & Pileni, M.-P. Hierarchy in Au Nanocrystal Ordering in Supracrystals: III. Competition between van der Waals and Dynamic Processes. *Langmuir* **30,** 7177–7181 (2014).

4. Wan, Y., Goubet, N., Albouy, P. & Pileni, M. Hierarchy in Au nanocrystal ordering in supracrystals: A potential approach to detect new physical properties. *Langmuir* **29,** 7456–7463 (2013).

5. Wan, Y., Goubet, N., Albouy, P.-A., Schaeffer, N. & Pileni, M.-P. Hierarchy in Au nanocrystal ordering in a supracrystal: II. Control of interparticle distances. *Langmuir* **29,** 13576–13581 (2013).

6. Rabani, E. & Egorov, S. Interactions between passivated nanoparticles in solutions: Beyond the continuum model. *J. Chem. Phys.* **115,** 3437–3440 (2001).

7. Egorov, S. & Binder, K. Effect of solvent quality on the dispersibility of polymer-grafted spherical nanoparticles in polymer solutions. *J. Chem. Phys.* **137,** 094901 (2012).

8. Luedtke, W. D. & Landman, U. Structure, dynamics, and thermodynamics of passivated gold nanocrystallites and their assemblies. *J. Phys. Chem.* **100,** 13323–13329 (1996).

9. Luedtke, W. D. & Landman, U. Structure and Thermodynamics of Self-Assembled Monolayers on Gold Nanocrystallites. *J. Phys. Chem. B* **102,** 6566–6572 (1998).

10. Landman, U. & Luedtke, W. D. Small is different: energetic, structural, thermal, and mechanical properties of passivated nanocluster assemblies. *Faraday Discuss.* **125,** 1–22 (2004).

11. Schapotschnikow, P., Pool, R. & Vlugt, T. J. H. Molecular Simulations of Interacting Nanocrystals. *Nano Lett.* **8,** 2930–2934 (2008).

12. Pool, R., Schapotschnikow, P. & Vlugt, T. J. H. Solvent Effects in the Adsorption of Alkyl Thiols on Gold Structures: A Molecular Simulation Study. *J. Phys. Chem. C* **111,** 10201–10212 (2007).

13. Schapotschnikow, P. & Vlugt, T. J. Understanding interactions between capped nanocrystals: three-body and chain packing effects. *J. Chem. Phys.* **131,** 124705 (2009).

14. Tay, K. & Bresme, F. Computer Simulations of Two Dimensional Gold Nanoparticle Arrays: the Influence of Core Geometry. *Mol. Simul.* **31,** 515–526 (2005).

15. Verso, F. L., Yelash, L., Egorov, S. A. & Binder, K. Interactions between polymer brush-coated spherical nanoparticles: the good solvent case. *J. Chem. Phys.* **135,** 214902 (2011).

16. Patel, N. & Egorov, S. A. Interactions between sterically stabilized nanoparticles in supercritical fluids: A simulation study. *J. Chem. Phys.* **126** (2007).

17. Boles, M. A. & Talapin, D. V. Many-Body Effects in Nanocrystal Superlattices: Departure from Sphere Packing Explains Stability of Binary Phases. *J. Am. Chem. Soc.* **137.** PMID: 25773648, 4494–4502 (2015).

18. Koga, K. & Sugawara, K. Population statistics of gold nanoparticle morphologies: direct determination by HREM observations. *Surf. Sci.* **529,** 23–35 (2003).

19. Nam, H.-S., Hwang, N. M., Yu, B. D. & Yoon, J.-K. Formation of an Icosahedral Structure during the Freezing of Gold Nanoclusters: Surface-Induced Mechanism. *Phys. Rev. Lett.* **89,** 275502 (27 Dec. 2002).

20. Wang, Y., Teitel, S. & Dellago, C. Melting of icosahedral gold nanoclusters from molecular dynamics simulations. *J. Chem. Phys.* **122,** - (2005).

21. Trzesniak, D., Kunz, A.-P. E. & van Gunsteren, W. F. A Comparison of Methods to Compute the Potential of Mean Force. *ChemPhysChem* **8,** 162–169 (2007).

22. Torrie, G. & Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23,** 187–199 (1977).

23. Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13,** 1011–1021 (1992).

24. Siepmann, J. I. & Frenkel, D. Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.* **75,** 59–70 (1992).

25. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

26. Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52,** 24–34 (1983).

27. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78,** 2690–2693 (14 Apr. 1997).

28. Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E* **56,** 5018–5035 (5 Nov. 1997).

29. Tuckerman, M. *Statistical mechanics: theory and molecular simulation* (Oxford University Press, 2010).

30. Bauer, G., Lange, A., Gribova, N., Holm, C. & Gross, J. Effective potentials between gold nano crystals–functional dependence on temperature. *Mol. Simul.* 1–6 (2014).

31. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).

32. Ciccotti, G., Ferrario, M., Hynes, J. T. & Kapral, R. Constrained molecular dynamics and the mean potential for an ion pair in a polar solvent. *Chem. Phys.* **129,** 241–251 (Jan. 1989).

Chapter 4

Phase Equilibria of Solid and Fluid Phases from Molecular Dynamics Simulations with Equilibrium and Nonequilibrium Free Energy Methods

Additions or deletions compared to the published work are marked with angular brackets.

In this work we present a methodology to determine phase coexistence lines for atomic and rigid molecular systems with an emphasis on solid-fluid and on solid-solid equilibria. Phase coexistence points are found by computing the absolute free energy for each candidate phase separately. For solid phases a combination of the extended Einstein crystal and the Einstein molecule method is presented which constitutes a convenient way to compute the absolute free energy with fixed center of mass. We compare results from equilibrium methods – thermodynamic integration and reweighting using the multistate Bennett acceptance ratio estimator (MBAR) – with simulations using a non-equilibrium method and discuss their advantages and disadvantages. Once absolute free energies of different phases are available, they are combined with simulations performed in the isothermal isobaric ensemble and MBAR which enables efficient, iterative tracing of coexistence lines. The method is applicable to both liquid-solid as well as solid-solid transitions and is comparably simple and convenient to apply since the same method (MBAR) is used to compute

free energies and to trace the coexistence line. Furthermore, statistical uncertainties can readily be computed in a transparent manner. We apply the method to an atomic solid (fcc argon) as well as small molecular systems (methanol and water) using the LAMMPS simulation package. Our study shows that all methods can be used to reliably compute the absolute free energy of solid phases while MBAR is the most flexible method with high statistical efficiency. We find the non-equilibrium method is an attractive choice since it is simple to setup and to post-process and is, hence, less prone to errors. The presented workflow provides a flexible, efficient and robust way to compute phase diagrams using openly available software.

4.1 Introduction

The knowledge of phase diagrams is of great interest for separation processes, pharmacology and material design. In the last decades, the molecular simulation community made significant advances in computing phase diagrams of vapor and liquid phases.[1] Today, there are established methodologies to compute vapor-liquid and liquid-liquid equilibria of pure substances as well as multi-component mixtures, namely Gibbs ensemble Monte Carlo, grand canonical Monte Carlo with multiple histogram reweighting as well as the Grand equilibrium method and various thermodynamic integration techniques.[1–7] In the field of molecular simulations, phase diagrams are used not only to model and predict quantities such as phase transition temperatures, pressures and coexistence properties such as densities and enthalpies but also as a means to evaluate and improve force fields via parameterization, because phase behavior is very sensitive to the underlying intermolecular potentials.[4] In comparison to the numerous studies of phase diagrams including liquid and vapor phases, studies of phase diagrams including crystalline solids are comparably scarce.[8,9]

There are multiple reasons why this is the case. Prior to the computation of a phase diagram the researcher has to have knowledge about the candidate crystal structure (or multiple structures) of the components of interest. By itself, the prediction of crystal structures is a difficult task and a vibrant field of research. Perfect crystals, to which we simply refer to as *solids*, consist of periodic repetitions of (often non orthogonal) unit cells. In comparison to methods used to compute vapor-liquid equilibria, this long range periodicity induces an inherent difficulty in constructing a reversible path between a solid and liquid phase. Also, since solids contain multiples of unit cells, the number of molecules cannot be arbitrarily chosen by the researcher. When comparing two solid structures, especially for small numbers of unit cells, the number of molecules will usually be different and it is not easily possible to directly connect systems of different size.

Computational methodologies for phase diagrams including solids (knowing the structure) can be divided into two steps, (1) finding initial phase coexistence and (2) tracing the coexistence line starting from the initial point. Despite the above mentioned difficulties there exist a great number of methods to approach the first task. These methods can be further divided into approaches relying on the computation of free energies which are used to explicitly solve phase coexistence conditions (equal temperature, pressure and chemical potential in both phases) using reference state models[10–13] or using simulations where a direct path between both phases is traversed[14–17] and direct simulation methods where both phases are brought into contact within a single simulation. For a comprehensive overview of literature concerning free energy methods, we refer to the review of Vega and Sanz.[8] For the second task, i.e. tracing the coexistence line, mostly the Gibbs-Duhem

integration method as well as histogram reweighting methods are used.[18–21]

For both the initial coexistence point as well as the coexistence line, multiple separate simulations and analysis methods are needed and while there exist established methods for each step, combining all steps is in practice not always an easy task. In this study we alleviate these difficulties by choosing a unified simulation and analysis method for all steps for computing coexistence lines. The proposed method fulfills the following criteria for determining phase coexistence properties: First, the method is applicable to both solid-fluid as well as solid-solid equilibria. Second, the method works for both mono-atomic crystals as well as molecular crystals. Third, the method can be applied with existing, freely available simulation software without the need to modify the codes. And finally, the analysis method predicts statistical uncertainties in a robust way. Statistical uncertainties play an important role in this field of research, because small changes to free energy differences between phases can lead to very pronounced shifts in the resulting phase coexistence conditions, so that precision and accuracy of the methods have to be analyzed carefully.

We use the multistate Bennett acceptance ratio (MBAR) estimator to compute free energy differences for the reference states and also to efficiently trace phase coexistence lines.[22,23] Like histogram reweighting, MBAR is used to combine data from multiple simulations to determine free energies continually in parameter space but without the need to bin data and with comparably higher accuracy. As noted before, histogram reweighting methods were successfully used to compute phase diagrams for solids. Eike et al. used histogram reweighting to find initial solid-liquid coexistence points by combining data from simulations in the NpT ensemble with free energies from constrained fluid λ-integrations[17,21] which they then used in conjunction with Gibbs-Duhem integration to trace the phase coexistence line. Chang et al. utilized reweighting techniques to directly trace phase coexistence lines by combining absolute free energies from Monte-Carlo simulations in the extended ensemble with data from NpT simulations.[19,20] Only recently, Schieber et al. used MBAR to study the phase diagram of benzene polymorphs.[24] In their work, they construct the free energy difference between two polymorphs as a function of temperature and pressure. This is done by combining data from separate simulations of the respective polymorphs of interest with so called pseudo-supercritical path[17,21,25] simulations to compute the free energy difference between the polymorphs at the same temperature and pressure.

In this study, we first compute the absolute free energies of candidate phases at a reference state (which can be different for both phases). For solid phases, we propose a combination of the extended Einstein crystal[26] and the Einstein molecule method[12,27–29] both of which are modifications of the method originally published by Frenkel and Ladd.[10] We conduct a comprehensive study of different ways to evaluate free energy differences using

several equilibrium and a non-equilibrium methods. Free energies at reference states are combined with data from NpT simulations to find the point of phase coexistence. Using this initial point, we predict coexistence for new states where no simulation data is present by reweighting existing data using MBAR. In an iterative fashion, short NpT runs at the new states are performed to improve initial estimates. We demonstrate the methods for argon (i.e. a Lennard-Jones fluid), as a widely studied system that is simple to reproduce, as well as for methanol as modeled using the united-atom OPLS (OPLS-UA) force field (with two solid phases), and for an ice phase of the TIP4P/2005 water model for which data is available in literature.[27,29,30] While the method is presented for rigid molecules, the work of Li et al. suggests that it should be also applicable to flexible molecules without modification assuming that free energy contributions due to intramolecular configurations are identical in all phases.[26] The emphasis of this work is to provide (1) a robust free energy approach to compute phase coexistence lines with (2) a transparent discussion of possible free energy methods and associated statistical uncertainties, (3) accompanied with detailed input and post-processing files. In that way, our study is meant to serve as valuable source for both, experienced practitioners but more importantly novices in this exciting field.

4.2 Methods

Two phases are in thermodynamic equilibrium if their temperatures, pressures and chemical potentials are equal. Temperature and pressure are control variables that can conveniently be imposed on a system so that the task of determining phase coexistence in molecular simulations really requires computing and equating the chemical potential in all trial phases. We start with a presentation of the methods we use to calculate free energies with a focus on the free energy of solid phases. Detailed information about the free energy path as well as different simulation and evaluation methods is given. Since the free energies are usually not determined at the phase coexistence state, we then discuss how phase coexistence is found by combining our results for the reference states with additional simulations in the isobaric-isothermal ensemble. Using an initial phase coexistence point, we explain how the coexistence line can be iteratively traversed combining free energy estimates utilizing MBAR in conjunction with additional short simulations.

4.2.1 Absolute free energy of the solid phase

We propose to compute the absolute free energy of a solid by combining the extended Einstein crystal and the Einstein molecule method.[12,26] The extended Einstein crystal method makes use of a reference state, namely a system of "central atoms" tethered to

their respective lattice positions with a harmonic potential. The central atoms can be single atoms or virtual sites of each molecule. In this reference state, no intermolecular interactions are present and – for rigid molecules – the free energy can be calculated analytically because each molecule resembles an independent harmonic oscillator. The extended Einstein crystal method entails a thermodynamic path from this reference state to the solid of interest in multiple steps. The first step along the pathway is to restrain the orientation of the molecules by gradually tethering additional atoms of the molecules to their lattice positions. The path to the solid of interest is completed by turning on intermolecular interactions and switching off the tethering potential. Different from the original Einstein crystal method of Frenkel and Ladd[10], the system's center-of-mass is not constrained along the path. The center-of-mass constraint, originally introduced to suppress divergence along the integration path, is not needed within Molecular Dynamics simulations given that (a) the systems' initial momentum is zero and (b) the propagation algorithm does not change the systems' momentum.[26] Eliminating the center-of-mass constraint greatly reduces the complexity to compute the free energy of the solid, especially for rigid molecules. However, by introducing a harmonic tethering potential using the same tether constant for each atom type the second assumption, (b), may not always hold true.

We modify the extended Einstein crystal approach by introducing the center-of-mass restraint of the Einstein molecule method.[12] In the Einstein molecule method, a single atom of one of the molecules, called the "carrier", is frozen along the free energy path which suppresses translational movement of the crystal as a whole. This way of restraining the center-of-mass, as compared to a conventional removal of translation, is advantageous because the contribution to the free energy due to freezing the carrier is easy to compute. In our view, a shortcoming of the original Einstein molecule method is the reference state in which all molecules are fully tethered, i.e. their translation as well as their orientation is restrained. To compute the free energy of this reference state, integration over the phase space of the tethered molecules is necessary, which can be accomplished e.g. using Monte-Carlo integration. To do this efficiently however, comparably large tether constants are needed which then makes the following free energy computations more cumbersome. As such, we prefer the reference state of the extended Einstein crystal method in combination with the Einstein molecule method to restrain the center-of-mass. Note that freezing the carrier only restrains the movement of the system's center-of-mass, i.e. it will still change during the course of a simulation. For brevity we will still use the term "fixed center-of-mass" in what follows.

The absolute free energy of the solid phase A_{sol} in this work is determined in three steps, with

$$A_{\mathrm{sol}} = A_0 + \Delta A_0 + \Delta A_1 \tag{4.1}$$

We start with a system of non-interacting molecules where each molecule is tethered to its lattice position by adding a harmonic tether potential to one of its central atoms (it is also possible to restrain the center-of-mass of the molecule with a harmonic potential). A single atom from one of the central atoms in the system acts as carrier and is fixed at its lattice position $\mathbf{r}_0$, instead of having a harmonic potential towards this position. That keeps the center-of-mass of the system fixed. In this state, the molecules can rotate freely. The free energy of the so-defined reference state A_0 is calculated as

$$\frac{\beta A_0}{N} = \frac{1}{N}\left[\ln\frac{N\Lambda^3}{V} - \frac{3}{2}(N-1)\ln\left(\frac{\beta k \Lambda^2}{2\pi}\right)\right],\tag{4.2}$$

where $\beta = \frac{1}{k_B T}$, k_B is the Boltzmann constant, T is the temperature, N is the number of molecules, V is the volume, Λ is the de Broglie thermal wavelength and k is the tethering constant of the harmonic potential. While the actual value of Λ is a function of temperature (and determines ideal gas properties such as the heat capacity), we can use arbitrary values when computing phase coexistence, provided the same value is used for all reference states. A_0 is the free energy of the Einstein molecule for an atomistic system.[12]

We then gradually tether additional atoms within each molecule to their respective lattice positions. That constrains the orientation of each molecule, but it also influences the strength of the positional constraints. As noted by Li et al.[26] preferably one would solely restrain the orientation of molecules but this functionality is usually not implemented in simulation packages. This free energy difference is equivalent to ΔA_0 as presented in the extended Einstein crystal method. Then, we gradually turn off all tethering potentials while simultaneously switching on intermolecular interactions, giving a free energy contribution ΔA_1. The term $A_0 + \Delta A_0$ in this work corresponds to the contribution of A_0 of the Einstein molecule method[28] and the term ΔA_1 in our study is analogous to the terms $\Delta A_1 + \Delta A_2$ of the extended Einstein crystal and the Einstein molecule method in literature.

To calculate A_{sol} we have to choose a value for the harmonic tether constant k, which defines the harmonic tethering potential

$$u_{\mathrm{tether}}(r) = \frac{k}{2}r^2 \tag{4.3}$$

where $r = |\mathbf{r} - \mathbf{r}_0|$ is the instantaneous distance of an atom to its lattice position $\mathbf{r}_0$. In this notation, we omit an index for the individual (central) atoms. The resulting value for the free energy A_{sol} does – in theory – not depend on the value of the harmonic constant k. In practice however, the choice of k affects computational expense as well as

numerical stability. In this work, we follow the route of Eike and Maginn[25] to derive k, i.e. we fit k to match the probability distribution observed during a NVT simulation of the fully interacting (but not tethered) crystal, with probability density $P(r)$ for position coordinate r around the lattice position $\mathbf{r}_0$ (i.e. the ensemble averaged position) of a regarded central atom, as

$$P(r) = \left(\frac{\beta k}{2\pi}\right)^{3/2} 4\pi r^2 \exp(-\beta k/2 \cdot r^2)\,, \tag{4.4}$$

which is the normalized probability density (as function of the distance of an atom from its lattice position, r) of a harmonic oscillator. Generally, we can sample probability distributions for all atom types that are used to tether the molecule and fit separate values for the tether constants but for convenience we use the same value for all atoms. Using equation (4.4), we treat every atom as an independent harmonic oscillator, which is strictly justified only for atomic crystals. In molecules, especially when they are described with rigid bonding constraints, atoms do not move independently and as such using the resulting (mean) k value for each atom will not lead to the correct probability distribution. It is possible to improve k by conducting additional NVT simulations where atoms are tethered to their lattice positions and no intermolecular potentials are present. We can then iteratively adjust k so that the resulting probability distribution of the tethered ideal gas simulation matches that of the fully interacting system. As we will discuss in the results, in practice, however, it is completely sufficient to get an estimate of the magnitude of k, which does not require an iterative procedure.

4.2.2 Free energy methods to compute ΔA_0 and ΔA_1

We use three different methods to compute the two contributions ΔA_0 and ΔA_1 of equation (4.1), namely thermodynamic integration (TI), the multistate Bennett acceptance ratio estimator method (MBAR) and non-equilibrium simulations. In all methods we introduce a coupling parameter λ which is used to drive a system from an initial state (0) to a final state (1). For example, a possible way to introduce λ could read

$$U(\lambda) = (1 - \lambda)U_0 + \lambda U_1\,, \tag{4.5}$$

where λ describes a linear change of the intermolecular energy function, U, going from the initial energy function, U_0, to the final energy function, U_1. The free energy difference due to changing λ (for a generic function $U(\lambda)$) reads

$$\Delta A = A_1 - A_0 = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda\,, \tag{4.6}$$

where $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$ denotes the ensemble average for a given value of λ. We can compute this free energy difference by choosing a distribution of λ values (depending on the integration scheme), conducting separate NVT simulations at each λ value and sampling the integrant of equation (4.6), the ensemble average. This first method for determining free energy differences is referred to as 'thermodynamic integration'.[31–33] We can then use an integration rule, here Simpson's rule and Gauss-Legendre quadrature, to compute ΔA. The free energy difference depends on the number and the spacing of λ values as well as the integration method.

A second way to compute free energy differences is to use reweighting techniques. In this work, we use MBAR.[22,23] The general working equation of MBAR reads

$$\Delta f_i = - \ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-u_i(\mathbf{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp[f_k - u_k(\mathbf{x}_{jn})]}, \tag{4.7}$$

where $f_i = \frac{F}{k_B T}$ is the reduced free energy, F stands for the free energy, the indices i, j and k denote the states and n counts the observations of a simulation (i.e. the instantaneous values observed along the Markov chain of a Monte Carlo simulation or the instantaneous value observed during a molecular dynamics simulation for a configuration $\mathbf{x}_{jn}$). Further, u_i is the reduced potential function which depends on the variables defining a state. In the example above, a state is entirely defined by the value of λ and the reduced potential function reads

$$u_i(\mathbf{x}_{jn})) = \beta U(\lambda_i, \mathbf{x}_{jn}). \tag{4.8}$$

This equation requires the evaluation of a configuration $\mathbf{x}_{jn}$ observed in a simulation performed in the isothermal-isochoric ensemble at λ_j using the potential function $U(\lambda_i)$. If the potential function has a complex λ dependence, one either has to perform this evaluation during the simulation (for all other states of interest) or by storing configurations which we then reevaluate using the respective potential function. If we, however, choose the potential function so that it is linearly dependent on λ, i.e. we can write

$$U(\lambda_i, \mathbf{x}_{jn}) = f(\lambda_i) U(\mathbf{x}_{jn}), \tag{4.9}$$

reevaluation of the configuration is no longer necessary as $f(\lambda_i)$ is merely a scaling factor.[34,35] Using the latter equation, we completely circumvent reevaluation of configurations in this study. For example, for the computation of ΔA_1 for the solid phase and as shown in detail below, we introduce λ as linear scaling factor of both the harmonic tethering potentials as well as the intermolecular potentials, i.e. we turn off the tethers using $(1 - \lambda)$ and simultaneously turn on the intermolecular potential using λ as linear scaling factor. Within a simulation it is then sufficient to sample the total, unscaled tethering

energy as well as the total, unscaled intermolecular energy. To reevaluate a simulation conducted with λ_i at a new λ_j, we evaluate the sampled energies using the respective scaling factors but with λ_j instead of λ_i. In LAMMPS it is directly possible to sample potential functions that are different from those used to propagate the system and hence sampling of unscaled energies is trivial. If this is not possible and only scaled energies are available in a simulation software, rescaling still can be used but additional treatment of the end states is necessary. Because in the end states either the total scaled tethering energy ($\lambda = 1$) is zero or the total scaled intermolecular energy (($\lambda = 0$) is zero, it is not possible to directly reweight the energies. For these two states separate reevaluation of configurations is then necessary.

In general, we can compute free energy differences by choosing suitable control variables. In the above example we introduce a coupling parameter, using a constant temperature and density along all states. As such, the resulting free energy difference, is the Helmholtz energy, i.e $\Delta F = \Delta A$. We will later discuss how MBAR can also be used to trace the coexistence lines by defining states in terms of temperature and pressure, where the analogous free energy difference in equation (4.7) then is the Gibbs free energy difference ($\Delta F = \Delta G$).

Equation (4.7) can be iteratively solved to yield free energy differences with respect to a reference state given that there is sufficient overlap of configuration space between (at least) adjacent states, i.e. there are configurations in one state that are likely to occur in the neighboring states (and vice versa). Measuring the amount of phase space overlap between states provides a systematic way to select the number and distribution of λ values.

For the third method for determining free energy differences, the non-equilibrium approach, equation (4.6) can be interpreted as limiting case of a non-equilibrium process, i.e. as work that is performed to infinitely slowly switch the system from state (0) to (1) along a switching path given by λ.[36] For finite switching rates, the work is no longer reversible and we can write[37]

$$\Delta A = \bar{W} - \bar{W}^{\text{diss}}, \tag{4.10}$$

where $\bar{W}$ is the mean value of the irreversible work for multiple switching realizations, that is multiple independent runs using the same switching time and path. The mean dissipated work $\bar{W}^{\text{diss}}$ is unknown, however, following[37,38] we can compute the free energy difference by conducting simulations in both directions

$$\Delta A = \frac{1}{2}[W_{01}^{\mathrm{rev}} - W_{10}^{\mathrm{rev}}] \tag{4.11}$$

$$= \frac{1}{2}[\bar{W}_{01} - \bar{W}_{10} - \bar{W}_{10}^{\mathrm{diss}} + \bar{W}_{01}^{\mathrm{diss}}] \tag{4.12}$$

$$= \frac{1}{2}[\bar{W}_{01} - \bar{W}_{10}] \, , \tag{4.13}$$

where the indices (01) and (10) stand for the forward and backward direction in which the path is traversed, respectively. Analogously to the equilibrium case we introduce a λ parameter which continuously changes with simulation time and rate $\dot{\lambda}$. Then, we can write

$$W^{01} = \int_0^{t_s} \dot{\lambda} \left(\frac{\partial U}{\partial \lambda} \right)_t \mathrm{d}t = \sum_n^N \Delta\lambda_n \left(\frac{\partial U}{\partial \lambda} \right)_{n\Delta t} \, , \tag{4.14}$$

where t_s is the switching time, Δt is the time step, the index n stands for the simulation time step number and N denotes the total number of time steps. In this work we follow the simulation protocol developed and implemented in LAMMPS by Freitas et al.[37] That is, a simulation consists of an equilibration phase in the initial state, the forward switch, another equilibration in the final state, followed by a backward switch. During the switching process, where the coupling parameter λ is gradually switched from 0 to 1 or vice versa, the summand of equation (4.14) is sampled. The entire procedure is repeated multiple times to determine the mean values for the forward and backward work so that the final free energy can be calculated via equation (4.13). The parameters for the non-equilibrium method are the switching function, $\lambda(t)$, and the switching time, t_s.

To compute ΔA_0 we use the following expression for the λ-dependent potential function

$$\Delta A_0 : \quad U(\lambda) = \lambda \sum_i^o \frac{k}{2} \Delta r_i^2 + \sum_i^c \frac{k}{2} \Delta r_i^2 \, , \tag{4.15}$$

where the sum $\sum_i^o \ldots$ runs over all atoms that are tethered to restrain the orientation. The central atoms are coupled using the full tethering potential, $\sum_i^c \ldots$, irrespective of the λ value. Δr_i is the distance of atom i from its equilibrium lattice position. We remind that during this stage, no intermolecular interactions are present; only the tethering potentials and intramolecular constraints have to be integrated during the simulation which is computationally very cheap. Simulation times above $600\,\mathrm{ns/day}$ can be achieved on a desktop computer (for 320 methanol molecules, $\Delta t = 1.0\,\mathrm{fs}$, i7-4770 CPU 3.40GHz, LAMMPS

(22 Feb 2018) compiled with Intel 16.0) using a single core. Effectively, computation of ΔA_0 is about two orders of magnitude faster than ΔA_1.

The potential function for ΔA_1 reads

$$\Delta A_1: \quad U(\lambda) = (1-\lambda) \sum_i^{\text{o}+\text{c}} \frac{k}{2} \Delta r_i^2 + \lambda U_{\text{inter}}, \tag{4.16}$$

where the sum $\sum_i^{\text{o}+\text{c}} \dots$ contains all tethered atoms with tethering potentials that are gradually eliminated while the intermolecular potential, U_{inter}, is gradually turned on. Scaling intermolecular interactions (both Lennard-Jones and coulombic) linearly is sufficient because the harmonic tether potentials prevent atoms from overlapping for small values of λ and it allows us to perform the reweighting for MBAR (see equation (4.9)) as post-processing step.

Having determined the Helmholtz energy as described above, the final step is to compute the chemical potential which is the property of interest for phase equilibria. The chemical potential of the solid is determined from

$$\beta \mu_{\text{sol}} = \frac{\beta A_{\text{sol}}}{N} + \frac{\beta p V}{N} = \beta a_{\text{sol}} + \frac{\beta p V}{N}. \tag{4.17}$$

Throughout the following discussion of the results we will report the dimensionless Helmholtz energy per molecule, $\beta a = \frac{\beta A}{N}$.

4.2.3 Free energy of the liquid phase

The free energy or chemical potential of a liquid can be determined in various ways.[39] Most methods rely on insertion of an additional molecule into the bulk phase in the spirit of the method of Widom[40] but other options such as a supercritical path from the ideal gas are also possible.[8] Generally, the method of choice depends on the system and state point of interest, e.g. for large molecules it may be necessary to traverse a direct path from the solid[21,41] or to create a cavity before inserting a molecule[26] to yield accurate results for the liquid phase.

In this work we use different methods to compute the chemical potential of the liquid phases. For the argon system, we simply use the analytic Lennard-Jones equation of state of Thol et al.[42]. For methanol, we gradually insert a single methanol molecule into the bulk phase. Starting from a non-interacting molecule we first add van der Waals interactions using a soft-core Lennard-Jones potential[43] before coulombic interactions are turned on. This is realized using a coupling constant λ just as discussed above. The free energy is evaluated via MBAR from multiple independent equilibrium NpT simulations.

For water, we follow the route of Vega et al.[8] by computing the free energy difference between a Lennard-Jones system (for which we can calculate the free energy using the analytic equation of state of Thol et al.) and the liquid water system by gradually turning on coulombic interactions using separate equilibrium simulations in conjunction with MBAR. As a conservative estimate for the uncertainty of the analytic equation of state we assume the same uncertainty as obtained for the solid phase.
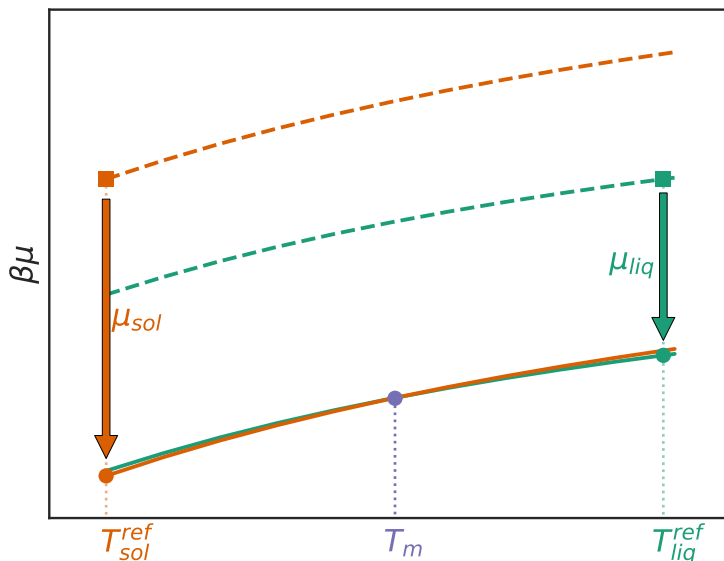
4.2.4 Initial coexistence point

Figure 4.1: Construction of the coexistence point (purple circle) at constant pressure. The relative free energy differences (dashed lines) are determined for each phase individually in regular NpT simulations with respect to reference states (square symbols), here solely defined by the temperatures $T_{\mathrm{sol}}^{\mathrm{ref}}$ and $T_{\mathrm{liq}}^{\mathrm{ref}}$. We use the chemical potentials, μ_{sol} (eq. (4.17)) and μ_{liq} (sphere symbols), as computed from the absolute free energies of the reference states to shift the curves (solid lines). The intersection of the solid lines denotes the coexistence point.

A schematic representation of the method to find the initial phase coexistence point is shown in FIG. 4.1. In this study, we determine absolute free energies for each phase at the same pressure which is convenient because that reduces the search space to only the temperature coordinate. We proceed in three steps. First, we conduct NpT simulations – separately for each phase – for various temperatures (along a predefined grid of temperatures), choosing the temperature-range to cover the region where phase coexistence is expected. During each simulation, we sample the system's internal energy and volume and evaluate the reduced potential function

$$u_i(\mathbf{x}_{jn})) = \beta_i \left[U(\mathbf{x}_{jn}) + pV(\mathbf{x}_{jn}) \right] , \qquad (4.18)$$

as a post-processing step. Index i denotes a state point along the grid of predefined temperatures. Using equation (4.7) with the above reduced energy function, the resulting free energy differences along these states are differences in the Gibbs free energy (or residual chemical potential) for varying temperatures along the predefined grid of temperatures (visualized as dashed lines in FIG. 4.1). Up to this point, the differences in the Gibbs energy are formulated with respect to the reference states (square symbols, FIG. 4.1). The next step is to shift the free energies to match the values of the absolute free energies calculated at the reference states (sphere symbols) using the methods we described before. Finally, the phase coexistence temperature (here the melting temperature T_m) is found at the intersection of the solid lines.
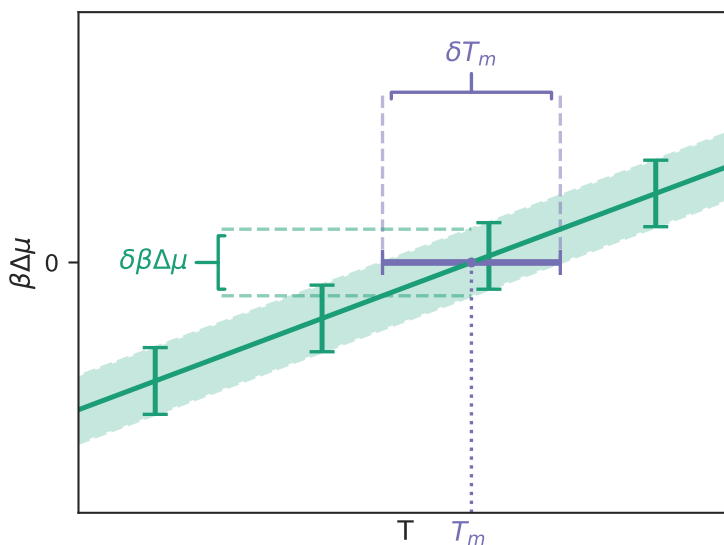

Figure 4.2: The statistical uncertainty of the melting temperature, δT_m, (purple) is found by intersection of the uncertainties from free energy calculations (green) with the zero line.

The statistical uncertainty of the coexistence temperature is determined from the statistical uncertainty of the difference of chemical potentials (between both phases) which consists of multiple contributions. For brevity of the following discussion we introduce the abbreviation $\hat{\mu} = \beta\mu$. The first contribution is the statistical uncertainty of the chemical potential of each reference state, $\delta\hat{\mu}_I^{\mathrm{ref}} = \delta\beta_I^{\mathrm{ref}}\mu_I^{\mathrm{ref}}$, where I denotes the respective phase, i.e. the liquid phase or a solid phase. Note that the reference temperatures as well as the number of molecules can be different for each phase. The second contribution is the statistical uncertainty of the differences of chemical potentials along the path from the reference state to the equilibrium state, $\delta\Delta\hat{\mu}_{\mathrm{I}} = \delta(\hat{\mu}_I - \hat{\mu}_I^{\mathrm{ref}})$. These uncertainties are directly accessible from MBAR. The statistical uncertainty of the difference of chemical potential between two phases, $\delta\beta\Delta\mu$ (here Δ denotes the difference between both phases, not temperatures), at any (inverse) temperature β can be calculated as

$$(\delta \beta \Delta \mu)^2 = (\delta \Delta \hat{\mu}_{\text{liq}})^2 + (\delta \hat{\mu}_{\text{liq}}^{\text{ref}})^2 + (\delta \Delta \hat{\mu}_{\text{sol}})^2 + (\delta \hat{\mu}_{\text{sol}}^{\text{ref}})^2 \,, \tag{4.19}$$

where we used a solid and a liquid phase as example. The resulting uncertainty of the coexistence temperature is then determined from the intersections of the statistical uncertainties over temperature where the difference of the chemical potential between both phases, $\beta \Delta \mu$, is zero as depicted visually in FIG. 4.2.

4.2.5 Construction of phase coexistence lines

Knowing an initial coexistence point, (T_0, p_0), we can construct the coexistence line by changing temperature (or pressure) and searching for the conjugated pressure (or temperature) where the chemical potentials of both phases are equal. This process can be repeated iteratively to trace the entire coexistence line. The question is, once we choose a new temperature, how can we sample the pressure-μ space efficiently to find a new coexistence point? We propose using the simulation data of the NpT simulations already conducted to construct the initial phase coexistence point. More specifically, we use MBAR to estimate free energies at additional states for which no simulation data is present. We select a new temperature and for both phases generate free energies for a range of pressures to find the equilibrium pressure at equal chemical potentials. The whole procedure can be performed for several new temperatures always using the initial coexistence point as reference to estimate the free energy differences from. At the new coexistence states (temperatures and pressures) found this way, additional NpT simulations for both phases are performed and used to improve the initial estimate. Since estimated states (especially for the first iteration) are different from the true coexistence states, simulations are conducted in the metastable region for at least one phase which could lead to phase transition occurring during a simulation. It is very important to make sure that no phase transitions are observed, e.g. by inspecting the radial distribution function of each phase for the different states.

Uncertainties for new coexistence states are determined using the bootstrap method. Starting from an initial coexistence point and simulation data at the estimated states, we use MBAR to construct the phase coexistence line as described above. Next, we generate bootstrap samples for every state point. This is done by randomly picking data points (with replacement) from the time series of potential energies and volumes for each state. These new samples are then evaluated using MBAR to generate a "bootstrapped coexistence line". The whole procedure is repeated 100 times and confidence intervals are obtained from the resulting distribution. Finally, we add the uncertainty of the initial coexistence state. Note that these statistical uncertainties are meaningful only if at least neighboring states overlap. If this is not the case, additional simulations have to be

performed to fill these gaps.

As a consequence of seeding the procedure with an initial coexistence point, its accuracy and precision are critically important. Clearly, the predictive efficiency of this method is limited to a certain range of temperature (because the phase spaces of the predicted and the simulated conditions become disjoint) and going to temperatures far away from the initial phase coexistence point requires multiple iterative prediction and simulation steps. As shown in the results section, we find that – for the systems studied in this work – already short NpT simulations of some hundred picoseconds are sufficient (without the need of a second refinement step) and because NpT simulations at multiple states can be performed in parallel the approach allows for rapid estimation of the coexistence line. As statistical uncertainties are determined with respect to the initial coexistence state they increase when moving away from the initial coexistence state, which is a downside our approach shares with the Gibbs-Duhem integration method.[44]

Our method of finding phase coexistence is similar in spirit to the work of Schieber et al. where MBAR was used to combine simulations at different temperatures and pressures to find phase coexistence of benzene polymorphs.[24] They also present an analytical alternative to the bootstrap method to assess statistical uncertainties of phase coexistence line by directly utilizing MBAR uncertainties which is cheaper than the bootstrap method although more difficult to apply.

4.3 Simulation details

Simulations are performed using the simulation package Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS, version 22 Feb 2018).[45] All input files are provided in the supplementary information.

The argon systems contain 5324 particles in both the liquid and solid phase. Lennard-Jones parameters are $\epsilon = 0.238122 \, \text{kcal}/\text{mol}$, $\sigma = 3.405 \, \text{Å}$. We use a truncated potential with a cut-off radius of $6\sigma = 20.43 \, \text{Å}$ and add standard long-range corrections.[33] We choose the large system size and cut-off radius to allow for comparison to the work of Mastny et al.[46] in which they showed that both – choosing too small cut-off radii ($< 4.5\sigma$ for a fcc Lennard-Jones solid) as well as small system sizes – can lead to systematic errors of the melting temperature.

The methanol systems contain 300 molecules for the α solid and the liquid phase and 320 molecules for the β solid phase. Methanol is modeled using the united atom version of the Optimized Potential for Liquid Simulations (OPLS-UA) force field where the parameters are taken from Saldago et al.[29] Similar to argon, we use a truncated potential with a cut-off radius of $r_c = 10.0 \, \text{Å}$ and add long-range corrections to both pressure and energy in accordance with Saldago et al.[29] Electrostatic interactions are

calculated using Particle Particle Particle Mesh Ewald.[47] The real-space contribution is truncated at a cut-off distance equal to that of the van der Waals cut-off. Bond lengths and angles are constrained using the SHAKE algorithm with a tolerance of 10^{-4} and a maximum of 50 iterations.[48] For the simulations with fixed center-of-mass of the entire simulation box, the CH3 united-atom site is used as "carrier". That is, the CH3 site of exactly one molecule is frozen in place during the course of the simulation. CH3 is also used as "central atom", i.e. it is the tethered site for a_0. We also performed simulations using the oxygen atom which yielded identical results.

The water systems contain 432 molecules for both the Ih-solid and the liquid phase. We use the TIP4P/2005 force field with a cut-off radius of $r_c = 8.5\,\text{\AA}$ for both the van der Waals and Coulombic interactions in accordance with the work of Abascal et al.[49] We use the oxygen atom as "carrier" and also as "central atom" for the simulation of Δa_0.

To keep the temperature constant, a Langevin thermostat is used with a damping constant of $100\,\text{fs}$.[50] We also conducted simulations using stochastic velocity rescaling and found the results to be identical.[51] In constant pressure simulations, we use a Nosé-Hoover barostat.[52] For the liquid phase, pressure coupling is performed isotropically while for the solid we use anisotropic coupling. In all simulations a Velocity-Verlet algorithm is applied to propagate the system using time steps of $2.0\,\text{fs}$ for the liquid and $1.0\,\text{fs}$ for the solid phase.[33] Smaller time steps for the free energy computations of the solid phases are necessary to properly sample the oscillations of tethered atoms especially when large constants for the tethering potential are used. As we will discuss later, the tether constant used in this study are comparatively low and as such, time steps of $1.0\,\text{ns}$ are sufficient to properly sample oscillations.

Initial configurations for simulations of the solid phases were created by placing atoms on the respective lattice sites of the perfect crystal. Then, anisotropic simulations in the isothermal isobaric ensemble at temperature $T_{\text{sol}}^{\text{ref}}$ and target pressure are performed from which the mean density is obtained. We then perform a simulation at this density (and shape of the simulation cell) and - after an equilibration phase - pick a random configuration as reference lattice. If configurations of the perfect crystal are available for the target pressure and temperature these configurations can also be used. The same initial configuration (for the respective phase) is used for all state points. I.e. all simulations for the free energy computations as well as the NpT simulations for the coexistence lines start from the lattice positions for the solid phases. For the liquid phase, initial configurations were created by randomly placing molecules into the simulation cell followed by an energy minimization and an equilibration in the NpT ensemble using an isotropic barostat as detailed above.

Besides MBAR, free energy differences for $\beta \Delta a_0$ and $\beta \Delta a_1$ from equilibrium simulations are computed using thermodynamic integration where we compare two integration

schemes, Gauss-Legendre quadrature (GLQ) and the method of Simpson. The distribution of λ-values for GLQ solely depends on the total number of values (the order of the quadrature). We add additional simulations for $\lambda = 0$ and $\lambda = 1$ which are needed for MBAR and the method of Simpson. Additionally, we use a distribution of λ-values similar to the switching function of Freitas et al. where differences between adjacent values are smaller near $\lambda = 0$ and $\lambda = 1$ similar to the GLQ scheme. We varied the number of λ-values (up to 100) to verify convergence of the calculated free energies. Utilizing overlap information from MBAR it is possible to iteratively adjust the number and distribution of λ-points so that the statistical uncertainties between states are minimized.[53] In this work we did not follow such an optimization approach. Instead, we use overlap information simply to ensure that all states are connected when the free energy is determined via MBAR.

As noted before, for non-equilibrium simulations no distribution of λ-values has to be provided. Instead, we have to specify a switching rate, i.e. the change of λ over the simulation time. We use the switching function reported and implemented in LAMMPS by Freitas et al. which results in slow changes of λ at the end states.[37]

As part of data post-processing, we remove non-equilibrated samples from all equilibrium simulations.[54] We then extract uncorrelated data evaluating the reduced potential functions for free energy calculations using the *timeseries* module of the *pymbar* python package (available at https://github.com/choderalab/pymbar). Consequently, the number of data points used for calculating averages and statistical uncertainties differ between simulations. As noted before, when evaluating free energies and statistical uncertainties using MBAR, it is pivotal to make sure that all states have sufficient overlap with neighboring states. For cross validation we also computed standard errors of the free energies using bootstrapping (500 bootstrap samples) and found errors to be consistent with those computed via MBAR.[55] For free energies of equilibrium simulations computed using Simpson and GLQ we estimate the error also using bootstrapping with 500 bootstrap samples. For non-equilibrium simulations, we follow the protocol of Freitas et al. and perform multiple independent runs for both forward as well as backward integration. Statistics are then calculated from the distribution of the free energies according to equation 4.13. The reported errors (numbers in parenthesis or bars in graphs) represent the 95% confidence interval. For MBAR this corresponds to twice the estimated statistical uncertainties (two standard deviations) directly reported from *pymbar* while the confidence intervals from bootstrap samples are constructed using the percentile method.[55] For the non-equilibrium simulations we perform $n = 10$ independent runs to get 10 independent values for the free energy, $x = \{\beta a_{\mathrm{sol},i}\}$, $i = 1, \ldots, 10$. We then compute the standard error of the mean of the free energy as $s(\bar{x}) = \frac{s(x)}{\sqrt{n}}$, where $s(x)$ is the unbiased estimate of the standard deviation as obtained from the independent runs. The confidence interval is then calculated

as $\bar{x} \pm 2.262 \cdot s(\bar{x})$ where the factor 2.262 comes from the t-distribution using 9 $(= n - 1)$ degrees of freedom.[56]

4.4 Results

In the following section we present our results. To validate our procedures, comparison to literature data is shown. We note that a proper comparison to published data concerning statistical uncertainties and computational efficiency is a non-trivial task because it is not always clear what error is reported and hence a meaningful comparison is difficult to ensure. As such we present statistical uncertainties of literature data as published in the respective studies which might not have the same confidence levels as those used in this work (and generally not the same simulation times were used) so that uncertainties are not intended to be directly comparable.

4.4.1 Argon

We start our discussion with the results for the fcc phase of argon. For the solid state, a reference temperature of $T = 83.88\,\mathrm{K}$ was chosen. First, an NpT simulation at the reference temperature and $p = 420.0\,\mathrm{bar}$ was performed to determine the equilibrium number density of $\rho = 0.02513(1)\,\text{Å}^{-3}$. In reduced Lennard-Jones units, this state corresponds to $T^* = \frac{Tk_B}{\epsilon} = 0.7$, $p^* = \frac{p\sigma^3}{\epsilon} = 1.0$, which is the reference point of the work of Eike et al.[21]

Subsequently, to determine the tether constant, we conducted a short simulation in the NVT ensemble and sampled the probability distribution of atom-displacements with respect to their equilibrium lattice positions. The simulated probability distribution was used to fit the value of the tether constant, $k = 3.0\,\mathrm{kcal}/\mathrm{mol\text{Å}^2}$. With the so-determined tether constant the absolute free energy of the solid phase can be calculated. Because argon consists only of a single Lennard-Jones interaction site, there is no orientational contribution, i.e. $\beta\Delta a_0 = 0$, and the absolute free energy is calculated as $\beta a_{\mathrm{sol}} = \beta a_0 + \beta\Delta a_1$.

With the non-equilibrium method applied to determine $\beta\Delta a_1$, only a single simulation (including equilibration, forward and backward direction) is required. The influence of the switching time on the free energy is assessed by repeating the procedure using the same switching function $\lambda(t)$ for different simulation times.

When equilibrium simulations are used to determine $\beta\Delta a_1$, we distribute up to 50 λ-values according to the GLQ scheme and add two additional simulations at $\lambda = 0$ and $\lambda = 1$, which are needed for the integration using Simpson's rule and for MBAR. As opposed to the ten runs from the non-equilibrium method, only a single free energy is computed and the statistical uncertainty is estimated using 500 bootstrap resamples for

GLQ and Simpson while for MBAR, statistical uncertainties are readily available from *pymbar*.

Table 4.1: Contributions to the absolute free energy of argon fcc as computed via GLQ, Simpson's rule, MBAR and from non-equilibrium simulations at $T = 83.88\,\mathrm{K}$ and $\rho = 0.02513\,\text{Å}^{-3}$ using a tether constant of $k = 3.0\,\mathrm{kcal}/\mathrm{mol\text{Å}^2}$ and a de Broglie wavelength of $\Lambda = 3.405\,\text{Å}$.

Method	βa_0	$\beta \Delta a_1$	$\delta \beta \Delta a_1$	$\beta \Delta a_{\mathbf{sol}}$	$\delta \beta \Delta a_{\mathrm{sol}}$
GLQ	5.2533	-11.4831	0.0004	**-6.2298**	0.0004
Simpson	5.2533	-11.4827	0.0008	**-6.2294**	0.0008
MBAR	5.2533	-11.4833	0.0002	**-6.2299**	0.0002
Non-eq.	5.2533	-11.4832	0.0003	**-6.2299**	0.0003

Table 4.1 summarizes the results for the absolute free energy of argon. The reported value for the non-equilibrium method is taken from a switching time of $5\,\mathrm{ns}$.
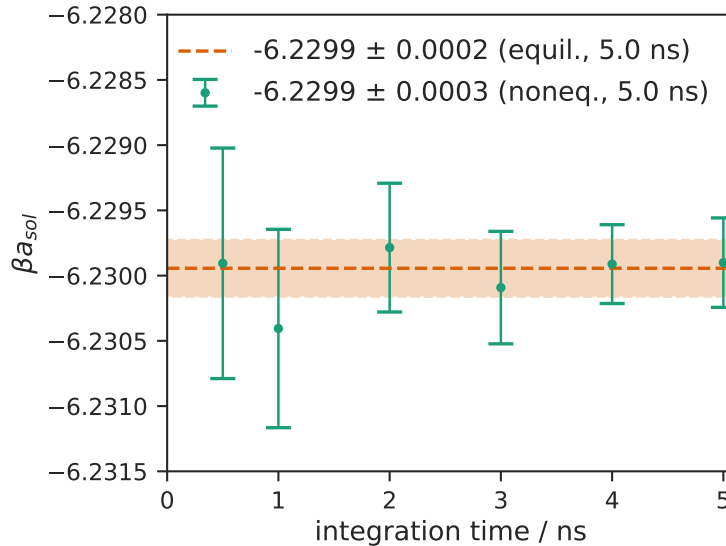

Figure 4.3: Absolute free energy of the argon fcc phase as computed from non-equilibrium simulations for different switching times (green symbols) and equilibrium simulations using MBAR (orange) at $T = 83.88\,\mathrm{K}$, $\rho = 0.02512\,\text{Å}^{-3}$ and a tether constant of $k = 3.0\,\mathrm{kcal}/\mathrm{mol\text{Å}^2}$.

Fig. 4.3 visualizes how the free energy changes with switching time. For comparison we also show the results obtained from MBAR. For argon, small switching times already yield good results for the free energy.

The liquid phase of argon at $T = 83.88\,\mathrm{K}$ and $p = 420.0\,\mathrm{bar}$ was computed using the equation of state of Thol et al. which is formulated in terms of the Helmholtz energy and consequently as function of temperature and density. We obtain the free energy (or the chemical potential) for a given temperature and pressure by iteratively adjusting the density until we find the target pressure using Newton's method. The free energy

can then directly be obtained for the temperature and the iterated number density ($\rho = 0.02269\,\text{Å}^{-3}$) as sum of the ideal gas contribution, $\beta a_{\mathrm{ig}} = -1.11025$ using $\Lambda = 3.405\,\text{Å} = \sigma$ for the de Broglie wavelength, and the residual contribution, $\beta a_{\mathrm{res}} = -5.10447$ to yield $\beta a_{\mathrm{liq}} = \beta a_{\mathrm{ig}} + \beta a_{\mathrm{res}} = -6.21472$.
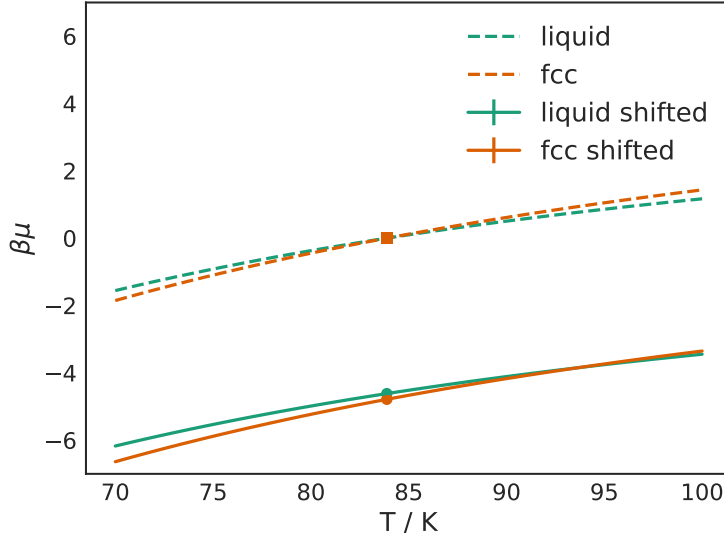

Figure 4.4: Chemical potential of argon at $p = 420.0\,\text{bar}$ for different temperatures with respect to reference states (dashed lines) and shifted (solid lines) according to the absolute free energies of the reference states.

To determine the equilibrium between fcc-solid and liquid argon, we performed NpT simulations at $p = 420.0\,\text{bar}$ and a range of temperatures around the reference states for both phases separately. Using MBAR, the free energy differences for the temperature range were constructed. The free energy of each phase is obtained by adding the absolute free energies of the reference states as detailed above and visualized in Fig. 4.4.

Fig. 4.5 shows the difference of the dimensionless chemical potential between both phases for varying temperature. The shaded region is the statistical uncertainty of $\beta\Delta\mu$, as calculated by combining the uncertainties from the absolute free energies of the references and the uncertainties from the free energy differences over multiple temperatures (with respect to the reference states) for both phases. The reported melting temperature and its statistical uncertainty are found at the zero-intersection of the mean value as well as the upper and lower bound of the confidence interval.

We reuse simulation data that was generated to find the initial phase coexistence point (data to construct the lines in Fig. 4.4) for calculating a first estimate of the phase coexistence line, see cross symbols in Fig. 4.6. We performed 50 additional simulations for each phase for these estimates (1 ns including equilibration, cross symbols in Fig. 4.6), which can be done in parallel, to correct the first estimate (see green line in Fig. 4.6). The number of additional simulations can be guided by inspecting overlap between adjacent states e.g. using overlap matrices[57] where the degree of overlap depends on the thermodynamic
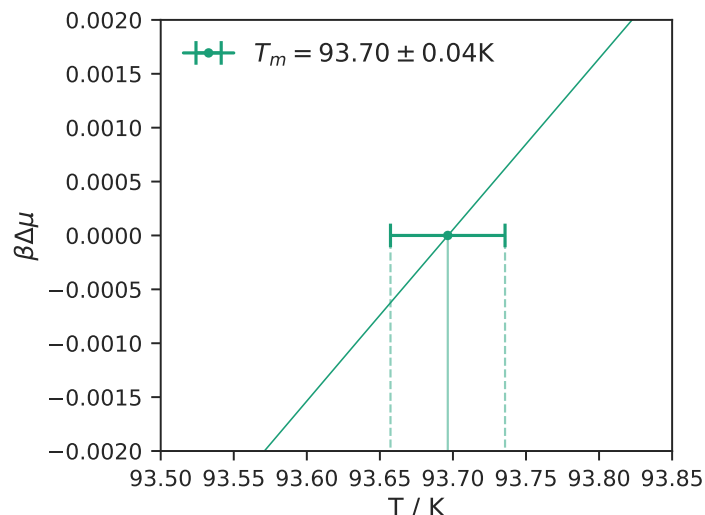
Figure 4.5: Melting temperature of argon fcc-solid at $p = 420.0\,\mathrm{bar}$. The phase equilibrium (green sphere) is located where both phases, fcc and liquid, have the same chemical potential. Uncertainty in the melting temperature is obtained from uncertainties of the differences of chemical potentials ($\delta\beta\Delta\mu$) between both phases (dashed vertical lines).

state, the system size[24] as well as the phase. Since we use a large system for argon – which leads to a narrower energy distribution – a comparably large number of simulations is needed to achieve overlap. The temperature increment was chosen such that there is sufficient overlap at low temperatures and pressures. For convenience, we chose the same increment for both phases and for the whole range of thermodynamic states which leads to large overlap at higher temperatures. It is possible to increase the computational efficiency of this method by selecting thermodynamic states as a function of temperature and pressure for each phase individually based on overlap information by iteratively adding or redistributing states. In this work we did not follow such an optimization approach because we found that already comparably small numbers of uncorrelated data points – and therefore short simulation times – are sufficient and a stage wise refinement (i.e. an iterative procedure) is not required. A single iteration already yields precise results, as confirmed by comparison of the resulting line with literature data[46] (shown by orange circles in Fig. 4.6).

4.4.2 OPLS-UA Methanol

We now present the results of our study of the α and β phases of methanol using the OPLS-UA forcefield. We chose methanol because, as opposed to argon, it shows features of more complex molecular structures and at the same time it is well documented in literature.[28,29] We followed the same procedure as presented for argon, so that for brevity we limit the discussion to the central results whereas detailed information of all intermediate steps is
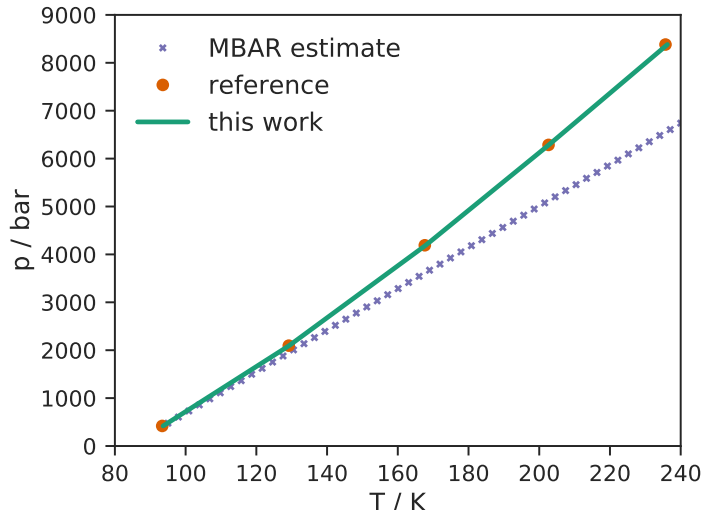
Figure 4.6: Phase equilibrium line of fcc-solid and liquid for argon. Results from this work using MBAR (green line) compared to values from literature (orange circles).[46] The purple crosses denote estimations based on MBAR using NpT data only from the initial coexistence point at $p = 420.0$ bar and represent the states where additional simulations were conducted. Error bars are smaller than line width and symbol sizes.

given in the supplementary information.

We use the same reference state for the α and β phase as Saldago et al.[29], namely $T_\alpha^{\text{ref}} = T_\beta^{\text{ref}} = 150$ K, $p = 1$ bar. NpT simulations yield number densities of $\rho = 0.01878(2)\,\text{Å}^{-3}$ and $\rho = 0.01882(2)\,\text{Å}^{-3}$ for the α and β phase, respectively. We use a tether constant of $k = 8.0\,\text{kcal}/\text{molÅ}^2$ for both phases which is considerably lower than the constant used in literature ($k = 3974.187\,\text{kcal}/\text{molÅ}^2$).[28] The chemical potential of the liquid phase was computed at $T_{\text{liq}}^{\text{ref}} = 298$ K, $p = 1$ bar and found to be $\beta\mu_{\text{liq}} = -12.157(9)$ using the value for the de Broglie wavelength $\Lambda = 1\,\text{Å}$. Comparing our results to REF.[29], we find good agreement for the β-phase ($\beta a_{\text{sol}}^{\text{ref}} = -29.44(1)$) and for the α-phase ($\beta a_{\text{sol}}^{\text{ref}} = -29.55(1)$). Detailed results for the absolute free energies for the α-solid and the β-solid phase for the different free energy methods are presented in table 4.2.

The choice of the tether constant plays an important role and as mentioned before the optimal value with regard to statistical efficiency as well as numerical stability depends on the free energy path that is traversed. In the original Einstein molecule method, a high value for k is preferable because it allows for efficient sampling of the reference state (which is done via Monte Carlo integration, see Appendix) and leads to convenient computation of the free energy change due to switching on intermolecular interactions since molecules are tightly coupled to their lattice positions. On the other hand, high tether constants limit the time-step that can be used within the simulation since molecules oscillate around their lattice positions with high frequency which can lead to numerical instabilities. To investigate the dependence on the time-step and the value of the tether

Table 4.2: Contributions to the absolute free energy of the α and β phases of methanol at $T = 150\,\mathrm{K}$ and $\rho_\alpha = 0.01878\,\mathring{A}^{-3}$, $\rho_\beta = 0.0194\,\mathring{A}^{-3}$. The tether constant is $k = 8.0\,^{\mathrm{kcal}}/_{\mathrm{mol}\mathring{A}^2}$ and the de Broglie wavelength is set to $\Lambda = 1.0\,\mathring{A}$.

Method	βa_0	$\beta\Delta a_0$	$\delta\beta\Delta a_0$	$\beta\Delta a_1$	$\delta\beta\Delta a_1$	$\beta\Delta a_{\mathbf{sol}}$	$\delta\beta\Delta a_{\mathrm{sol}}$
α-phase							
GLQ	2.157	8.270	0.004	-39.999	0.002	**-29.571**	0.005
Simpson	2.157	8.271	0.003	-39.990	0.008	**-29.562**	0.008
MBAR	2.157	8.271	0.002	-40.000	0.001	**-29.572**	0.003
Non-eq.	2.157	8.272	0.004	-40.007	0.001	**-29.578**	0.004
β-phase							
GLQ	2.159	8.294	0.007	-39.879	0.003	**-29.426**	0.008
Simpson	2.159	8.283	0.006	-39.868	0.012	**-29.427**	0.013
MBAR	2.159	8.282	0.004	-39.880	0.002	**-29.439**	0.005
Non-eq.	2.159	8.276	0.003	-39.879	0.001	**-29.444**	0.003

constant, we conduct additional simulations of the α methanol phase using a tether constant of $k = 3974.187\,^{\mathrm{kcal}}/_{\mathrm{mol}\mathring{A}^2}$ as published by Aragones et al. for different time-steps and compare them with results using k as derived from the probability distribution of atomistic displacements.[28]
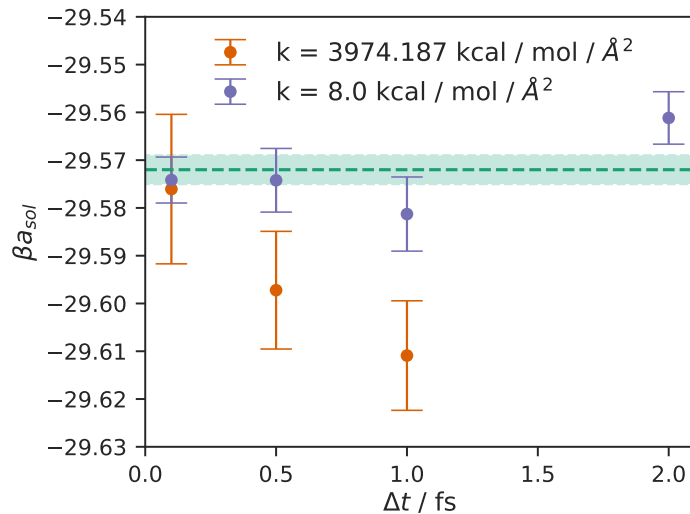

Figure 4.7: Absolute free energy of the methanol α-phase as computed from non-equilibrium simulations with switching time of $2\,\mathrm{ns}$ for different time-steps and tether constants, k. For comparison, the result from equilibrium simulations using MBAR with a simulation time of $5\,\mathrm{ns}$ is shown (green dashed line) with uncertainties as shaded region.

In FIG. 4.7 we analyze the effect of the chosen time-step for a given tether constant. For this purpose, we performed additional, separate simulation runs for the non-equilibrium method using a switching time of $2\,\mathrm{ns}$. These results suggest that for our method a low tethering constant is favorable. We find consistent results for time-steps up to $\Delta t = 1\,\mathrm{fs}$

while conserving decent precision. For a larger time-step of $\Delta t = 2\,\mathrm{fs}$ the resulting absolute free energy no longer overlaps with the uncertainties of the equilibrium simulation for the low tethering constant. Using the tether constant of Aragones et al. with our method we find larger statistical uncertainties and see a stronger dependence of the free energy on the time-step. Only with a time-step of $\Delta t = 0.1\,\mathrm{fs}$ (leading to rather excessive computation times) we were able to match the value for $a_0 + \Delta a_0$ of the original Einstein molecule method (see Appendix). We conclude that, for the presented free energy path, it is advisable to use comparably soft tether constants as obtained from adjusting the harmonic tether potential to resemble the probability distribution of atomistic displacements in the solid.
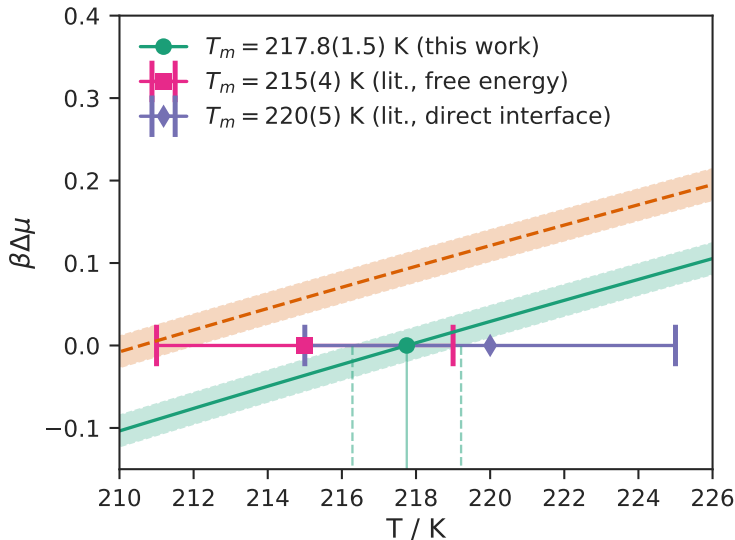

Figure 4.8: OPLS-UA methanol: Melting point of the α-phase (green line) as compared to data from literature.[29] The orange dashed line represents the difference of chemical potentials between liquid and the β-solid phase. Uncertainties in the chemical potentials are visualized as shaded regions.

We calculate the melting point using the absolute free energies of our reference states as computed via MBAR. In Fig. 4.8 we show the difference of chemical potentials between α and liquid (green) and β and liquid (orange) phases. Our results predict an α-liquid transition at $T_m = 217.8 \pm 1.5\,\mathrm{K}$ at ambient pressure which is in full agreement with the melting temperatures published by Saldago et al. ($T_m = 215 \pm 4\,\mathrm{K}$ using the Einstein molecule method, $T_m = 220 \pm 5\,\mathrm{K}$ from direct coexistence simulations).[29] Note that from experiments the β phase was found to be the stable phase (with the α phase being stable at lower temperatures). While the OPLS-UA force field is not able to properly describe the stable phase, Saldago et al. recently presented an new force field for methanol that correctly predicts the liquid β transition.[58] As before, the reported uncertainty of the melting temperature is given by the intersections of the uncertainties with the zero line of $\beta\Delta\mu$ (visualized as dashed lines). The uncertainty of T_m consists of the uncertainties

from the reference states as well as the uncertainties by reweighting for temperatures away from the reference states. Note that since the slopes of $\beta\Delta\mu$ are small, already small uncertainties of the free energies result in large uncertainties of the melting temperature.
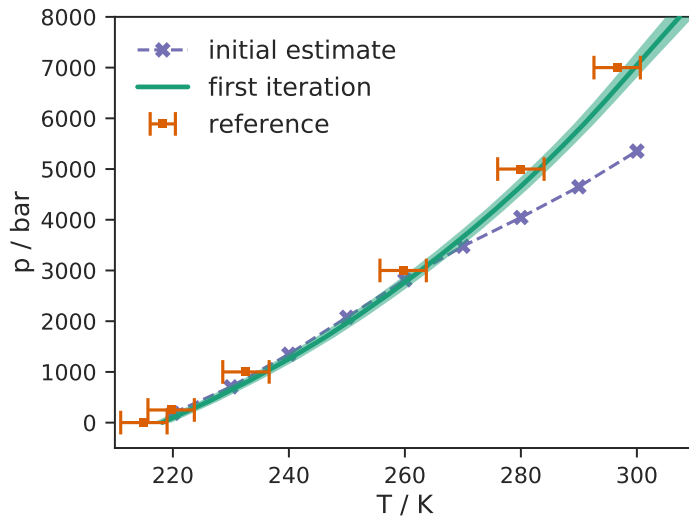

Figure 4.9: OPLS-UA methanol: Phase coexistence curve for α-solid and liquid using a single iteration of simulations (green line) at states estimated via MBAR (purple crosses) compared to data from literature (orange squares).[29] Uncertainties of the melting temperatures as obtained from the bootstrap method are visualized as shaded region.

Using the initial phase coexistence, we trace the coexistence line by estimating new temperatures for a given set of pressures and subsequently performing 9 additional NpT simulations in both phases (a total of 18 simulations) to refine the results. All additional NpT simulations are conducted in parallel, without a second refinement step (i.e. iteration). The results of a single iteration are shown in Fig. 4.9 together with results of Saldago et al. (the error bar for the literature data is ± 4 K, i.e. the error of the initial melting temperature). Due to slightly different initial melting temperatures there is a small gap between both coexistence lines underpinning the importance to start from a good initial value. Note also that statistical uncertainty in coexistence temperatures (δT) increases the farther we move away from the initial value.

4.4.3 Ice (Ih) and liquid phase equilibrium of TIP4P/2005 water

In our third study, we compute the absolute free energy of ice Ih at $T_{\rm sol}^{\rm ref} = 200$ K and $p = 1$ bar ($\rho = 0.03103\text{Å}^{-3}$) using the TIP4P/2005 force field. The absolute free energy of ice Ih at this state point is well documented in literature via Monte-Carlo as well as Molecular Dynamics simulations.[28]

Water adds another aspect of complexity for the calculation of free energies because it is a symmetric molecule. The oxygen atom acts as "central atom", i.e. it is initially

restrained to its lattice position through a harmonic tether potential, while both hydrogens are tethered to restrain the orientation (for the calculation of ΔA_0). Because we are tethering each hydrogen to just one lattice position, they become distinguishable. Furthermore, configurations that would have identical probability (180 degree rotations about the axis of symmetry) are suppressed and as such an additional contribution to the free energy A_0 is required. For detailed discussion of this topic we refer to the work of Aragones et al.[28]
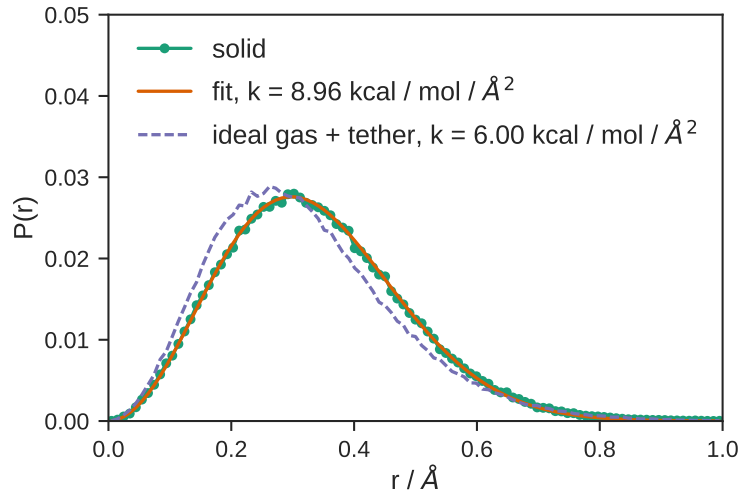

Figure 4.10: TIP4P/2005 water: probability distribution of oxygen displacement for the ice Ih phase from unrestrained NVT simulations (green symbols) at $T = 200\,\mathrm{K}$ and $\rho = 0.03103\,\mathring{A}^{-3}$. Resulting distribution from fitting k, the constant of the harmonic tether potential, to match the displacement probability of data (orange line). Probability distribution from a simulation of non-interacting molecules where all atoms were restrained using a constant of $k = 6.0\,\mathrm{kcal}/\mathrm{mol\mathring{A}^2}$ (purple dashed line).

Table 4.3: TIP4P/2005 water: contributions to the absolute free energy of ice Ih at $T = 200\,\mathrm{K}$ and $\rho = 0.03103\mathring{A}^{-3}$ using a tether constant of $k = 6.0\,\mathrm{kcal}/\mathrm{mol\mathring{A}^2}$.

Method	βa_0	$\beta\Delta a_0$	$\delta\beta\Delta a_0$	$\beta\Delta a_1$	$\delta\beta\Delta a_1$	$\beta\Delta a_{\mathrm{sol}}$	$\delta\beta\Delta a_{\mathrm{sol}}$
GLQ	0.206	6.761	0.005	-33.233	0.003	**-26.266**	0.006
Simpson	0.206	6.762	0.003	-33.233	0.005	**-26.265**	0.005
MBAR	0.206	6.763	0.003	-33.232	0.003	**-26.264**	0.004
Non-eq.	0.206	6.762	0.001	-33.236	0.003	**-26.269**	0.003

As before, we first determine a suitable value of the tethering constant. Fig. 4.10 illustrates the displacement probability of an oxygen atom obtained from an unrestrained simulation in the NVT ensemble. Fitting the tether constant according to equation (4.4) yields a value of $k = 8.96\,\mathrm{kcal}/\mathrm{mol\mathring{A}^2}$ (orange line). Next, we conduct simulations of non-interacting molecules that are tethered to their lattice positions using the same k-values for both oxygen as well as hydrogen. For the final value of $k = 6.0\,\mathrm{kcal}/\mathrm{mol\mathring{A}^2}$ (purple

dashed line) both oxygen as well as hydrogen probability distributions are in good enough agreement with those from unrestrained simulations.

Results for the free energies are reported in table 4.3. The analytical reference, βa_0, contains two additional contributions due to symmetry, $\beta \Delta a_0^{\mathrm{sym}} = -\ln 2$, and due to proton disorder of ice Ih, $\beta \Delta a_0^{\mathrm{pd}} = -\ln(3/2)$.[8,28] As before, we find agreement across all methods.

The absolute free energy of the liquid phase is computed at $T = 298\,\mathrm{K}$ and $p = 1\,\mathrm{bar}$ ($\rho = 0.03332(1)\mathrm{\mathring{A}}^{-3}$) as $\beta a_{\mathrm{liq}} = -15.6000(13)$ (see supplementary information for details). We follow the procedure of Vega et al.[8] and use a Lennard Jones system (using oxygen Lennard-Jones parameters) as reference system for which the free energy is readily available from the equation of state of Thol et al.[42] We then perform multiple equilibrium simulations in the isobaric isothermal ensemble in which we gradually turn on electrostatic interactions using linear scaling. The free energy difference is then computed via MBAR.
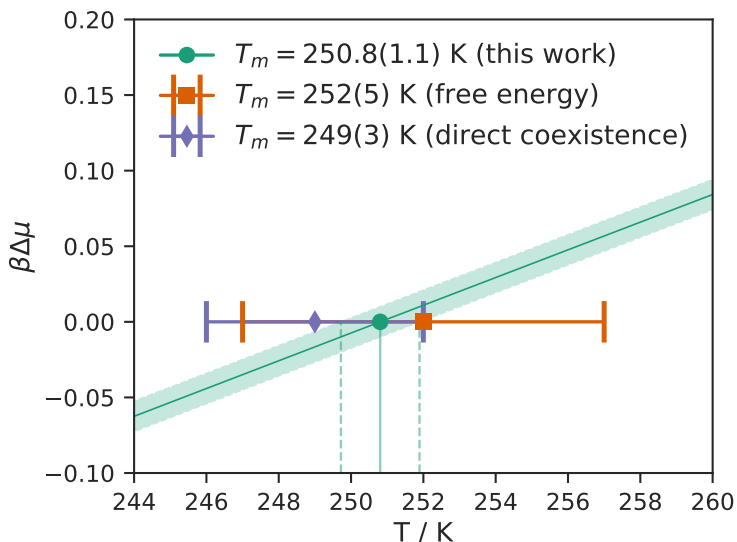

Figure 4.11: Melting point of TIP4P/2005 water: solid Ih and liquid phase equilibrium at $p = 1\,\mathrm{bar}$ from this study compared to data from literature. The melting temperature (green sphere) is located where both phases have the same chemical potential. For comparison, literature data (table II of REF.[59]) from free energy calculations (orange) and direct coexistence simulations (purple) is shown.

Fig. 4.11 shows the melting point as compared to data from literature. Good agreement is found for our results to literature references, whereby a rather low statistical uncertainty was here achieved. Fig. 4.12 presents the coexistence line as constructed from estimates using MBAR and a single iteration of 5 additional NpT simulations for both the liquid and the solid phase.
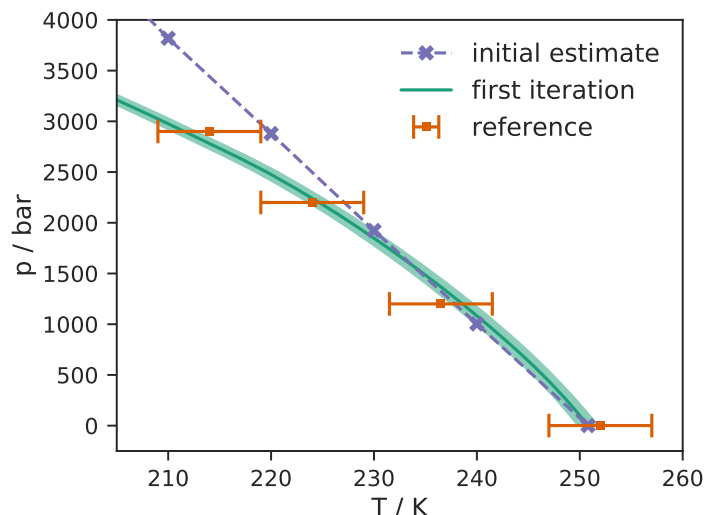
Figure 4.12: Melting line of TIP4P/2005 water: Ih-liquid phase equilibrium from this study (green line) compared to data from literature. Orange squares represent data from free energy methods and Gibbs-Duhem integration and are obtained from digitalizing figure 4 of reference[59]. Uncertainties of the melting temperatures as obtained from the bootstrap method are visualized as green shaded region. Purple crosses denote states where additional NpT were performed.

4.4.4 Discussion of free energy methods

Generally, for the systems we studied in this work, all methods yield similar results to within statistical uncertainty. This finding is in accordance with the studies of Moustafa et al. and Tan et al. in which the authors compared equilibrium methods to determine the absolute free energy of atomistic solids.[60,61] The advantage of the GLQ scheme is that a comparably small number of λ points already yields good estimates of the free energies. Furthermore in the GLQ scheme the states $\lambda = 1$ and $\lambda = 0$ are not needed. This is a big advantage when used in conjunction with the Einstein molecule method. While freezing the carrier constrains the center-of-mass of the system over the course of a simulation, the crystal as a whole wiggles quite considerably about its initial center-of-mass position. This is noticeable in particular when turning off the tethers completely during the calculation of ΔA_1 (see supplementary information). As a consequence considerably longer simulations are needed at this state to collect a sufficient number of uncorrelated samples. Using the GLQ scheme with a small (< 25) number of λ points alleviates this problem since the tethers are always active and hence the motion of the crystal is suppressed. For highly precise results however, the number of integration nodes has to be increased and the problem also arises using the GLQ scheme. In contrast, using MBAR and Simpson, endpoints have to be sampled. We found that a distribution of λ points according to the equation of the switching rate given in the work of Freitas et al. leads to consistent phase space overlap but typically more intermediate states (> 50) are needed.[37] Regions of

poor overlap can easily be enhanced by adding additional simulations whereas one has to repeat all simulations when changing the order of the GLQ scheme. We favor MBAR here, because convergence can be tested more easily by successively adding more states and also by comparing results with thermodynamic integration via Simpson for which no additional simulations are needed. In conclusion though, for the same total simulation time (but different numbers of λ points) we find that statistical uncertainties of MBAR and GLQ are practically identical. The non-equilibrium method yielded statistical uncertainties comparable to MBAR for the solids investigated in this study and it turned out to be a very robust method that is convenient to set up and post-process while convergence as well as statistical uncertainties are straightforward to determine.

4.5 Conclusion

In this work we study solid-fluid phase equilibria of atomic and rigid molecules. We propose a method to compute the absolute free energy of a solid phase that combines the center-of-mass constraint of the Einstein molecule method with the free energy path of the extended Einstein crystal method. The value for the tether constant (of the harmonic tether potential) was derived from the probability distribution of atomistic displacements of the target crystal phase which enables efficient and numerically stable tracing of the free energy path because comparably large time-steps can be used.

The free energy of the solid phase was evaluated using three different methods: thermodynamic integration, reweighting using MBAR and non-equilibrium simulations. All methods yield the same results within statistical uncertainty. We found MBAR to be the most flexible method with low statistical uncertainties for all systems investigated in this study since it can be used to compute free energies of the reference states and can also be used to construct the coexistence lines using simulations in the isothermal-isobaric ensemble. In conjunction with an analysis of phase space overlap it provides a systematic approach to distribute λ values, where λ is the coupling parameter enabling a transition between an Einstein crystal and a solid structure of full intermolecular interactions. In literature, thermodynamic integration using a Gauss-Legendre quadrature is often used which is a good strategy to get an estimate the free energy with comparably small numbers of λ values. Since the distribution of λ-values is fixed for a given number of points, testing for convergence is more cumbersome because old simulations cannot be reused. Our study showed that - using the same total simulation time - non-equilibrium simulations (for the systems considered in this study) are statistically competitive for calculating free energies of solids to equilibrium methods. They provide a robust route to perform free energy computations. Non-equilibrium simulations are appealing because in practice they are convenient to setup and moreover post-processing and statistical analysis are comparably

simple to perform. We believe that these are valuable and often underestimated features of a free energy method where typically several manual (subtle) steps are required.

Combining the absolute free energies of reference states we calculated melting points for fcc-solid argon, the α-solid phase of OPLS-UA methanol and the ice-Ih phase of TIP4P/2005 water. All computed melting points have high precision and show good agreement with literature data. Using initial melting points we presented an effective method to trace the coexistence line: Based on NpT simulations already available for identifying an initial melting point, we used MBAR to estimate new coexistence states at other temperatures and pressures. At the estimated conditions additional simulations are conducted to refine the results. Formally the procedure is iterative, however, in our application a single refinement step (using rather short simulation times) yielded good estimates for phase coexistence lines and allowed moving considerably far along a melting line.

Appendix

The contributions $A_0 + \Delta A_0$ of our method are identical to the reference, A_0, of the original Einstein molecule (EM) method.[28] In their publication, Aragones et al. compute the reference by integration over the configurational space of tethered ideal gas molecules and the carrier which has one frozen atom. The partition function of this system reads (for rigid molecules)

$$
\begin{aligned}
Q^1 = \ & \int \mathrm{d}^{n-1}\mathbf{r}^{(c)} \exp\left[-\beta \sum_{j=2}^{n} \frac{k}{2} \Delta r_j^{(c)2}\right] \\
& \times \frac{1}{N}\left\{\frac{1}{\Lambda^3} \int \mathrm{d}^n\mathbf{r}^{(i)} \exp\left[-\beta \sum_{j=1}^{n} \frac{k}{2} \Delta r_j^2\right]\right\}^{(N-1)},
\end{aligned}
\tag{4.20}
$$

where Λ is the de Broglie thermal wavelength, N is the number of molecules, n is the number of atoms in a molecule and Δr^2 denotes the squared displacement of an atom from its lattice position. The first integration is performed over the positions of the carrier molecule (superindex (c)). The first atom of the carrier is frozen (denoted by the superindex 1 in Q^1) and hence is excluded from the integration. The second integration is performed for the rest of the molecules that are able to move freely. Since no intermolecular potentials are present in the reference state, the integral is the same for each free molecule and therefore only has to be computed once. The factor $\frac{1}{N}$ arises from the number of possible permutations to assign a molecule (its atoms) to a lattice site.[12,28] The free energy is then
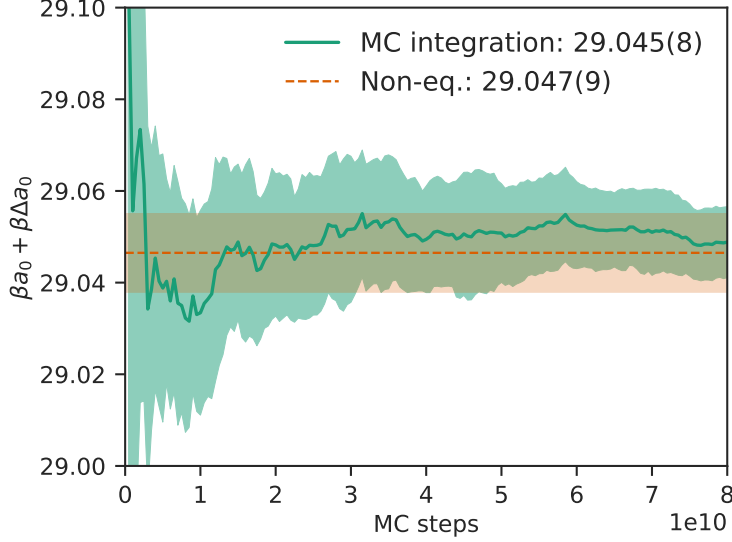
Figure 4.13: Free energy of the reference contribution of the α-solid phase of methanol as obtained from the Einstein molecule method using Monte Carlo integration. The running mean (green line) is computed from 10 separate integrations and plotted over the number of Monte Carlo steps. The orange dashed line represents results from non-equilibrium simulations with a time-step of $\Delta t = 0.1\,\text{fs}$. Statistical uncertainties across all runs are visualized as shaded region. The tether constant is $k = 3974.187\,\text{kcal}/\text{mol}\text{Å}^2$.

$$\frac{\beta A^1}{N} = \frac{1}{N} \ln N - \frac{1}{N} \ln I_1 - \frac{(N-1)}{N} \ln \frac{I_2}{\Lambda^3}, \tag{4.21}$$

where we follow the notation of REF.[28] and substitute the first and second integral of equation (4.20) as I_1 and I_2, respectively. The final expression for A_0 of the Einstein molecule reads

$$\frac{\beta A_0(\text{EM})}{N} = \frac{1}{N} \left[\ln \frac{N\Lambda^3}{V} - \ln I_1 - (N-1) \ln \frac{I_2}{\Lambda^3} \right]. \tag{4.22}$$

In comparison, the reference of the method used in this work is simply that of an *atomistic* EM[12]

$$\frac{\beta A_0}{N} = \frac{1}{N} \left[\ln \frac{N\Lambda^3}{V} - \frac{3}{2}(N-1) \ln \left(\frac{\beta k \Lambda^2}{2\pi} \right) \right]. \tag{4.23}$$

To show that $A_0 + \Delta A_0$ is equivalent to $A_0(\text{EM})$ we conducted Monte Carlo integrations of I_1 and I_2 (see figure 4.13) and compared the result with $A_0 + \Delta A_0$ from nonequilibrium simulations. The results of the simulations are presented in table 4.4.

Table 4.4: Free energy of the Einstein molecule reference state of α-solid methanol as computed from non-equilibrium simulations using a tethering constant of $k = 3974.187\,\mathrm{kcal/mol\AA^2}$ and a timestep of $\Delta t = 0.1\,\mathrm{fs}$.

time [ns]	βa_0	$\beta \Delta a_0$	$\delta \beta \Delta a_0$	$\beta a_0 + \beta \Delta a_0$	$\delta(\beta a_0 + \beta \Delta a_0)$
1.0	11.44	17.61	0.02	29.05	0.02
2.0	11.44	17.61	0.01	29.05	0.01
3.0	11.44	17.60	0.01	29.04	0.01
4.0	11.44	17.61	0.01	29.05	0.01

References

1. Panagiotopoulos, A. Z. Monte Carlo methods for phase equilibria of fluids. *J. Phys.: Condens. Matter* **12,** R25 (2000).

2. Panagiotopoulos, A. Z. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol. Phys.* **61,** 813–826 (1987).

3. Errington, J. R. Direct calculation of liquid–vapor phase equilibria from transition matrix Monte Carlo simulation. *J. Chem. Phys.* **118,** 9915–9925 (2003).

4. Weidler, D. & Gross, J. Transferable Anisotropic United-Atom Force Field Based on the Mie Potential for Phase Equilibria: Aldehydes, Ketones, and Small Cyclic Alkanes. *Ind. Eng. Chem. Res.* **55,** 12123–12132 (2016).

5. Mehta, M. & Kofke, D. A. Coexistence diagrams of mixtures by molecular simulation. *Chem. Eng. Sci.* **49,** 2633–2645 (1994).

6. Vrabec, J. & Hasse, H. Grand Equilibrium: vapour-liquid equilibria by a new molecular simulation method. *Mol. Phys.* **100,** 3375–3383 (2002).

7. Messerly, R. A., Razavi, S. M. & Shirts, M. R. Configuration-Sampling-Based Surrogate Models for Rapid Parameterization of Non-Bonded Interactions. *J. Chem. Theory Comput.* **14,** 3144–3162 (2018).

8. Vega, C., Sanz, E., Abascal, J. L. F. & Noya, E. G. Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins. *J. Phys.: Condens. Matter* **20,** 153101 (2008).

9. Polson, J. M. & Frenkel, D. Calculation of solid-fluid phase equilibria for systems of chain molecules. *J. Chem. Phys.* **109,** 318–328 (1998).

10. Frenkel, D. & Ladd, A. J. C. New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.* **81,** 3188–3193 (1984).

11. Wilding, N. B. & Sollich, P. A Monte Carlo method for chemical potential determination in single and multiple occupancy crystals. *Europhys. Lett.* **101,** 10004 (2013).

12. Vega, C. & Noya, E. G. Revisiting the Frenkel-Ladd method to compute the free energy of solids: The Einstein molecule approach. *J. Chem. Phys.* **127,** 154113 (2007).

13. Almarza, N. G. Computation of the free energy of solids. *J. Chem. Phys.* **126,** 211103 (2007).

14. Wilding, N. B. Phase Switch Monte Carlo. *AIP Conference Proceedings* **690,** 349–355 (2003).

15. Errington, J. R. Solid–liquid phase coexistence of the Lennard-Jones system through phase-switch Monte Carlo simulation. *J. Chem. Phys.* **120,** 3130–3141 (2004).

16. McNeil-Watson, G. C. & Wilding, N. B. Freezing line of the Lennard-Jones fluid: A phase switch Monte Carlo study. *J. Chem. Phys.* **124,** 064504 (2006).

17. Grochola, G. Constrained fluid λ-integration: Constructing a reversible thermodynamic path between the solid and liquid state. *J. Chem. Phys.* **120,** 2122–2126 (2004).

18. Kofke, D. A. Gibbs-Duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation. *Mol. Phys.* **78,** 1331–1336 (1993).

19. Chang, J. & Sandler, S. I. Determination of liquid–solid transition using histogram reweighting method and expanded ensemble simulations. *J. Chem. Phys.* **118,** 8390–8395 (2003).

20. Chang, J., Lenhoff, A. M. & Sandler, S. I. Determination of fluid–solid transitions in model protein solutions using the histogram reweighting method and expanded ensemble simulations. *J. Chem. Phys.* **120,** 3003–3014 (2004).

21. Eike, D. M., Brennecke, J. F. & Maginn, E. J. Toward a robust and general molecular simulation method for computing solid-liquid coexistence. *J. Chem. Phys.* **122,** 014115 (2005).

22. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129,** 124105 (2008).

23. Shirts, M. R. Reweighting from the mixture distribution as a better way to describe the Multistate Bennett Acceptance Ratio (2017).

24. Schieber, N. P., Dybeck, E. C. & Shirts, M. R. Using reweighting and free energy surface interpolation to predict solid-solid phase diagrams. *J. Chem. Phys.* **148,** 144104 (2018).

25. Eike, D. M. & Maginn, E. J. Atomistic simulation of solid-liquid coexistence for molecular systems: Application to triazole and benzene. *J. Chem. Phys.* **124,** 164503 (2006).

26. Li, L., Totton, T. & Frenkel, D. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. *J. Chem. Phys.* **146,** 214110 (2017).

27. Aragones, J. L., Valeriani, C. & Vega, C. Note: Free energy calculations for atomic solids through the Einstein crystal/molecule methodology using GROMACS and LAMMPS. *J. Chem. Phys.* **137,** 146101 (2012).

28. Aragones, J. L., Noya, E. G., Valeriani, C. & Vega, C. Free energy calculations for molecular solids using GROMACS. *J. Chem. Phys.* **139,** 034104 (2013).

29. Salgado, D. G. & Vega, C. Melting point and phase diagram of methanol as obtained from computer simulations of the OPLS model. *J. Chem. Phys.* **132,** 094505 (2010).

30. Polson, J. M., Trizac, E., Pronk, S. & Frenkel, D. Finite-size corrections to the free energies of crystalline solids. *J. Chem. Phys.* **112,** 5339–5342 (2000).

31. Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **3,** 300–313 (1935).

32. Tuckerman, M. *Statistical mechanics: theory and molecular simulation* (Oxford University Press, 2010).

33. Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 2001).

34. Naden, L. N., Pham, T. T. & Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 1. Removal of Uncharged Atomic Sites. *J. Chem. Theory Comput.* **10.** PMID: 26580188, 1128–1149 (2014).

35. Naden, L. N. & Shirts, M. R. Linear basis function approach to efficient alchemical free energy calculations. 2. Inserting and deleting particles with coulombic interactions. *J. Chem. Theory Comput.* **11,** 2536–2549 (2015).

36. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78,** 2690–2693 (14 Apr. 1997).

37. Freitas, R., Asta, M. & de Koning, M. Nonequilibrium free-energy calculation of solids using LAMMPS. *Comput. Mater. Sci.* **112,** 333–341 (2016).

38. De Koning, M. Optimizing the driving function for nonequilibrium free-energy calculations in the linear regime: A variational approach. *J. Chem. Phys.* **122,** 104106 (2005).

39. Kofke, D. A. & Cummings, P. T. Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation. *Fluid Phase Equilib.* **150,** 41–49 (1998).

40. Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **39,** 2808–2812 (1963).

41. Sellers, M. S., Lisal, M. & Brennan, J. K. Free-energy calculations using classical molecular simulation: application to the determination of the melting point and chemical potential of a flexible RDX model. *Phys. Chem. Chem. Phys.* **18,** 7841–7850 (2016).

42. Thol, M. *et al.* Equation of State for the Lennard-Jones Fluid. *J. Phys. Chem. Ref. Data* **45,** 023101 (2016).

43. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R. & van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters* **222,** 529–539 (1994).

44. Kofke, D. A. Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line. *J. Chem. Phys.* **98,** 4149–4162 (1993).

45. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117,** 1–19 (1995).

46. Mastny, E. A. & de Pablo, J. J. Melting line of the Lennard-Jones system, infinite size, and full potential. *J. Chem. Phys.* **127,** 104504 (2007).

47. Hockney, R. W. & Eastwood, J. W. *Computer simulation using particles* (Adam Hilger, 1988).

48. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

49. Abascal, J. L. F. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123,** 234505 (2005).

50. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17,** 1302–1322 (3 Feb. 1978).

51. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).

52. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101,** 4177–4189 (1994).

53. Mecklenfeld, A. & Raabe, G. Efficient solvation free energy simulations: impact of soft-core potential and a new adaptive λ-spacing method. *Mol. Phys.* **115,** 1322–1334 (2017).

54. Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **12,** 1799–1805 (2016).

55. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7,** 1–26 (Jan. 1979).

56. Grossfield, A. *et al.* Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living J. Comp. Mol. Sci.* **1,** 5067– (Oct. 27, 2018).

57. Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29,** 397–411 (May 2015).

58. Gonzalez-Salgado, D. & Vega, C. A new intermolecular potential for simulations of methanol: The OPLS/2016 model. *J. Chem. Phys.* **145,** 034508 (2016).

59. Conde, M. M., Gonzalez, M. A., Abascal, J. L. F. & Vega, C. Determining the phase diagram of water from direct coexistence simulations: The phase diagram of the TIP4P/2005 model revisited. *J. Chem. Phys.* **139,** 154505 (2013).

60. Moustafa, S. G., Schultz, A. J. & Kofke, D. A. A comparative study of methods to compute the free energy of an ordered assembly by molecular simulation. *J. Chem. Phys.* **139,** 084105 (2013).

61. Tan, T. B., Schultz, A. J. & Kofke, D. A. Suitability of umbrella- and overlap-sampling methods for calculation of solid-phase free energies by molecular simulation. *J. Chem. Phys.* **132,** 214103 (2010).

Chapter 5

Conclusion

In this work I studied the potential of mean force (PMF) between pairs of gold nano crystals with alkyl thiol ligands in vacuum by means of molecular dynamics simulations and developed a correction term for three-body contributions. The PMF between nano crystals was determined from a series of simulations in the canonical ensemble by integration of the forces acting on the nano crystals' centers of mass. The procedure was applied to nano crystals with varying core sizes, ligand lengths as well as temperatures. I showed that temperature dependence of the two-body PMF can be modeled applying a thermodynamic perturbation theory. The model can be efficiently parameterized using two simulated PMF and it can subsequently be extrapolated to a wide range of temperatures which drastically reduces the amount of simulations needed. In contrast, the simulated three-body correction does – to good approximation – not depend on temperature. I used this finding to develop an empirical model based on the temperature independent repulsive contribution of the pair PMF.

These findings shed light on some aspects of effective interactions between nano crystals and they may constitute a small stepping stone towards understanding and modeling of nano crystalline super structures. Effective interactions, however, merely mark the beginning of an elaborate process to develop a bottom-up approach to model the formation of super structures where the next steps would consist of estimating stable candidate crystal structures as well as actual simulation of said structures to estimate phase transitions (including multi-body potentials) – both of which are vibrant fields of research on their own.

Although the initial objective was on determining phase equilibria of nano crystals based on the effective (temperature dependent) interaction potentials, I realized the need for research even for simple fluids, which motivates the second part of this study. In the second part I studied solid-liquid phase equilibria utilizing free energies as obtained from molecular dynamics simulations. This study proposes combining the "extended Einstein crystal" approach and the "Einstein molecule" method to determine the absolute free

energy of the solid phase: the extended Einstein crystal method supplied me with the free energy pathway while I utilized the Einstein molecule method to restrain center of mass movement of the crystal. In combination with the force constant as obtained from unrestrained equilibrium simulations, this constitutes a numerically stable and efficient way to compute the absolute free energy of a solid. In addition, I studied different free energy methods, i.e. a method based on overlap sampling, thermodynamic integration as well as a nonequilibrium method, and assessed their ease of applicability, convergence and statistical uncertainty. Phase coexistence lines were constructed in two steps. First, I reused simulation data conducted to determine an initial melting point to make an estimate of additional melting conditions via reweighting. And second, I conducted additional, short simulations in the isothermal-isobaric ensemble at the estimated phase equilibrium states in order to refine the initial estimate through simulations conducted in parallel. For the systems studied in this work, I found that a single set of additional simulations was sufficient to trace the coexistence line for a wide range of pressures (or temperatures). This work included absolute free energies and phase coexistence lines for three systems of different complexity – argon, methanol and water. For all systems my results showed good agreement with literature data. The method worked very well for rigid molecules and while I did not test it for fully flexible molecules it should be directly applicable as the original extended Einstein crystal study was presented for fully flexible molecules.

Independent of the actual choice of method (which includes the actual simulation method and code i.e. the used algorithms as well as parameters and implementation) it is crucial to make reproduction of the reported results as easy as possible. With ever increasing complexity and quantity of methods it is getting more and more difficult and (timely) demanding to gain traction in this exciting field as a new practitioner and it should be in the interest of the scientific community – more than is currently the case – to lower this barrier as much as possible.

Appendix A

Appendix

A.1 Argon

A.1.1 FCC solid phase details

FIG. A.1 shows the probability distribution of atomic displacements used to determine the tethering constant for the solid phase. The tethering constant is tested by performing a simulation in which all atoms are tethered while no intermolecular interactions are present. The resulting probability distribution of atomic displacements can be compared to unrestrained simulation results. The exact value of the tethering constant is not important; it is sufficient to estimate the order of magnitude since results should not depend on the tethering constant that is used. We use the nearest integer value for convenience.
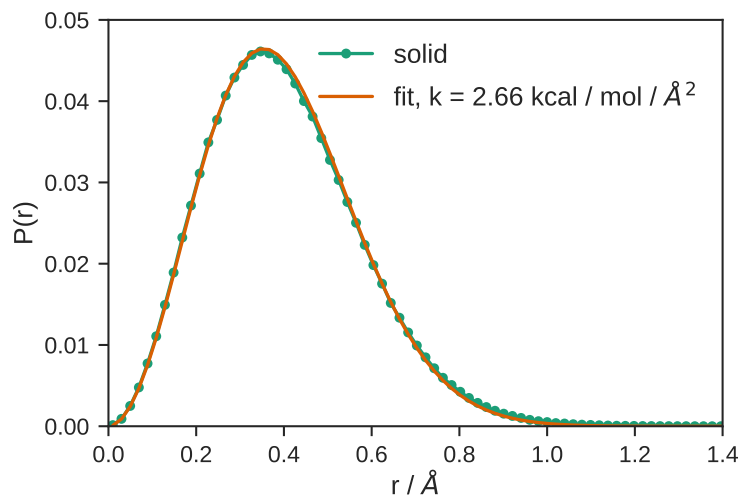

Figure A.1: Probability distribution of atomic displacement of the argon fcc phase as determined from unrestrained NVT simulations (green symbols, $T = 83.88 \, \text{K}$, $\rho = 0.02512 \, \text{Å}^{-3}$). The orange line represents a fit of the tethering constant to match the probability of simulation data.

Table A.1: Contributions to the absolute free energy of the fcc solid phase as obtained from nonequilibrium simulations for different switching times ($T = 83.88\,\mathrm{K}$, $\rho = 0.02512\,\mathring{\mathrm{A}}^{-3}$, $k = 3.0\,\mathrm{kcal/mol\mathring{A}^2}$). For an atomistic solid there is no orientational contribution and hence $\beta\Delta a_0 = 0.0$. Confidence intervals are determined from 10 independent simulations. The reference, βa_0, was calculated using the Lennard-Jones diameter as de Broglie wavelength, $\Lambda = \sigma$.

time [ns]	βa_0	$\beta\Delta a_1$	$\delta\beta\Delta a_1$	$\beta\Delta a_{\mathrm{sol}}$	$\delta\beta\Delta a_{\mathrm{sol}}$
0.5	5.2533	-11.4832	0.0009	-6.2299	0.0009
1.0	5.2533	-11.4837	0.0008	-6.2304	0.0008
2.0	5.2533	-11.4831	0.0005	-6.2298	0.0005
3.0	5.2533	-11.4834	0.0004	-6.2301	0.0004
4.0	5.2533	-11.4832	0.0003	-6.2299	0.0003
5.0	5.2533	-11.4832	0.0003	-6.2299	0.0003
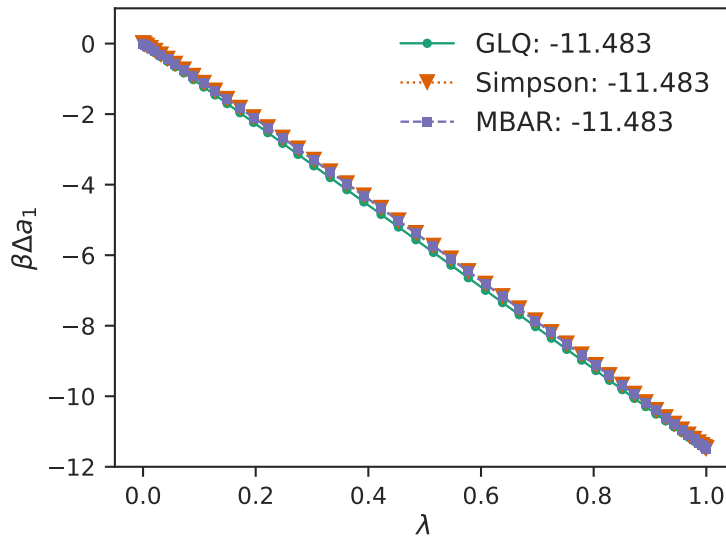
Figure A.2: $\beta\Delta a_1$ vs. λ of the argon fcc phase from equilibrium simulations at $T = 83.88\,\mathrm{K}$ and $\rho = 0.02512\,\mathring{\mathrm{A}}^{-3}$ using a tether constant of $k = 3.0\,\mathrm{kcal/mol\mathring{A}^2}$.

FIG. A.2 presents the free energy path of $\beta\Delta a_1$ as a function of the coupling parameter λ. Shown here are results for 52 simulations. We used the GLQ scheme to distribute 50 λ-values and added 2 additional values for the end states, $\lambda = 0$ and $\lambda = 1$, for Simpson and MBAR. Using a GLQ scheme, typically a smaller number of λ-values of about 25 values is sufficient (for the systems studied here). As shown in FIG. A.2 the free energy path shows almost linear behavior when plotted versus λ which is advantageous for TI. To achieve overlap between states, especially when approaching $\lambda \to 1$ (no tethers), we found that additional λ-points are necessary for MBAR. However, rather short simulation times of 1 ns can be used and for the same total simulation time (not necessarily the same number of λ-values) GLQ and MBAR yield similar statistical uncertainties.

A.1.2   Center-of-mass movement of the Einstein Molecule

During computation of the absolute free energy of the solid phase, βa_{sol}, we observed that – when turning off the tethers to calculate $\beta \Delta a_1$ – the center-of-mass of the whole crystal moves considerably at the end point $\lambda = 1$ (tethers are coupled via $(1-\lambda)u_{\mathrm{tether}}$). While it is relatively consistent over the course of a simulation, this translational movement of the whole crystal leads to large displacements of atoms with respect to their equilibrium lattice positions. This effect is presented in FIG. A.3 for the argon fcc-phase. Using the regular Einstein molecule (EM) method leads to behavior presented as green line. Because no tethers are active, it is generally possible for atoms to switch lattice positions which would lead to large displacements. However, this is not the case here as swaps of positions would not influence the center-of-mass of the crystal which is what we observed in our simulations. Freezing additional atoms decreases this effect as the orange (2 fixed atoms) and purple (3 fixed atoms) lines show. This effect is present also when unconstrained Monte-Carlo simulations are used (not shown here). To our knowledge, this effect was not observed or discussed in literature. Often, a GLQ scheme is used to perform integration of $\beta \Delta a_1$ where simulations in the end points still have tethers active. We found that only when going to comparably large numbers of GLQ nodes (> 30 for argon) tethers in the end states are small enough to observe this effect. The center-of-mass movement is less pronounced for molecular crystals, but to ensure convergence of $\beta \Delta a_1$ we recommend to increase simulation time at states where tethers are very weak to make sure that enough uncorrelated samples can be extracted and also, when using MBAR, to make sure that there is sufficient overlap in the end states. Our results suggest that – when sufficiently sampled – GLQ and MBAR can be reliably used to compute the free energy differences albeit the discussed difficulties.

A.1.3   States for initial phase coexistence and tracing of the co-existence line

The argon system in this study consists of 5324 particles. The larger the size of a system the narrower the energy distributions will be and – as a consequence – to achieve overlap between adjacent states, more thermodynamic states are necessary to cover a certain range of temperatures and pressures than for smaller systems. To construct the initial phase co-existence point at $p = 420.0\,\mathrm{bar}$ we use temperature steps of $\Delta T = 2\,\mathrm{K}$ between $70\,\mathrm{K}$ and $100\,\mathrm{K}$ leading to a total of 16 simulations for each phase. Additional 50 simulations in both phases were then added to trace the coexistence line between $420\,\mathrm{bar}$ and $7000\,\mathrm{bar}$ (a list of all temperatures and pressures is given in the *simulation_materials/argon_npt_states.dat* file of the supporting information).
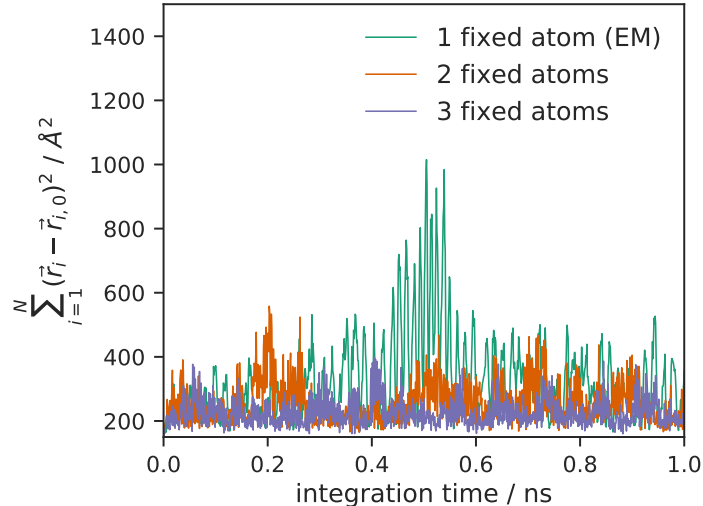
Figure A.3: Sum of squared atom displacements with respect to their lattice positions, $(\vec{r}_i - \vec{r}_{i,0})^2$, over simulation time with different numbers of frozen "carrier" atoms. No tethers are active in this state, i.e. it represents the end state $\lambda = 1$ for $\beta \Delta a_1$. Freezing multiple atoms reduces the translation of the crystal as a whole.

A.1.4 Coexistence temperatures and pressures as compared to literature

In TABLE A.2 results for fcc-liquid coexistence temperatures and pressures (in Lennard-Jones units) from this work are compared to literature data.[1] Note that uncertainties are presented as 67% confidence intervals as reported in the reference.

Table A.2: Comparison of argon fcc melting temperatures in Lennard-Jones units with literature.[1] Uncertainties in this table are given as 67% CI as reported in the reference.

P	T_{m}	$T_{\mathrm{m}}^{\mathrm{lit}}$
1.0	0.7820(2)	0.7793(4)
5.0	1.0788(2)	1.078(3)
10.0	1.4012(4)	1.399(4)
15.0	1.6916(7)	1.691(2)
20.0	1.972(3)	1.967(2)

A.2 OPLS-UA methanol

A.2.1 α-solid phase details

For methanol, two free energy differences have to be computed. $\beta \Delta a_0$, which is the path from tethered central atoms to restrained (orientational) but non-interacting molecules,

and $\beta\Delta a_0$, which is the path from tethered, non-interacting molecules to the solid of interest. FIG.A.4 visualizes the first contribution while FIG.A.5 depicts the second contribution in terms of the free energy versus the coupling parameter λ. Statistical uncertainties for the different contributions are reported in the main manuscript. While the second contribution, $\beta\Delta a_1$, shows almost linear behavior just like the argon case, $\beta\Delta a_0$ shows a steep slope for small values of λ. A distribution of λ values according to the GLQ scheme for $\beta\Delta a_1$ with a total of 25 points is sufficient whereas for $\beta\Delta a_0$ more intermediate values are needed to achieve convergence. Interestingly, a distribution of λ-values using the switching function from the non-equilibrium method showed the best results for MBAR and Simpson.[2]

The behavior of the free energy versus the coupling value shows that the first contribution, restraining the orientation of molecules, also results in larger (absolute) statistical uncertainties as compared to the second contribution. More simulation time could be invested to lower the statistical uncertainty of $\beta\Delta a_0$ since it is very cheap to compute because no intermolecular interactions are present in this stage. In this study we did not follow such an optimization approach and use the same total simulation time for both contributions which leads to different simulation times for the equilibrium simulations depending on the number of λ-values that were used.
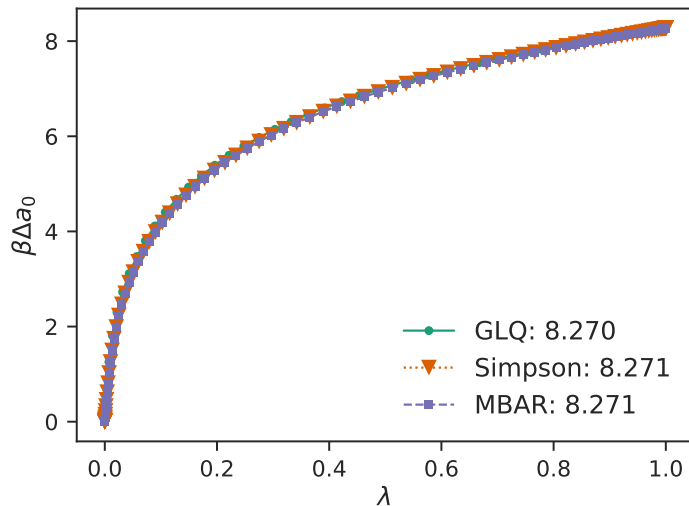

Figure A.4: Free energy path for $\beta\Delta a_0$ over the coupling parameter λ for the α-phase of OPLS-UA methanol at $T = 150\,\mathrm{K}$, $\rho = 0.01878\,\text{Å}^{-3}$ and $k = 8.0\,\mathrm{kcal/molÅ^2}$. As limiting case for convergence of the free energy, 100 λ-values were used. $\lambda = 0$ represents the state of tethered "central atoms" whereas for $\lambda = 1$ orientations of molecules are restrained.

FIG.A.6 shows a comparison between the absolute free energy as obtained via non-equilibrium simulations and as obtained via MBAR.
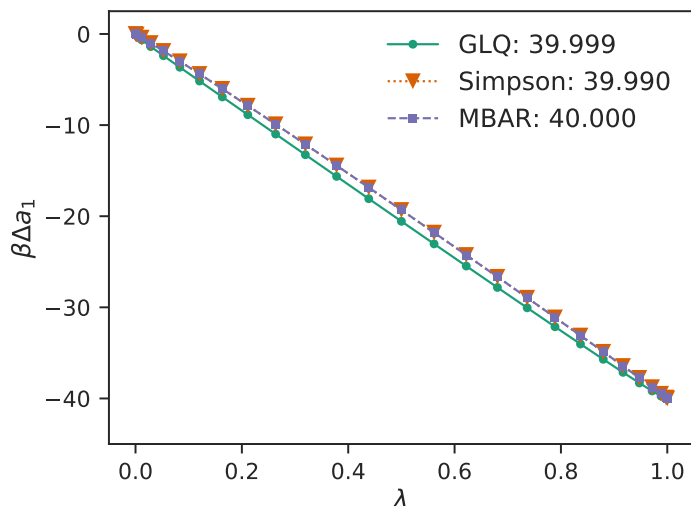
Figure A.5: Free energy path for $\beta\Delta a_1$ over the coupling parameter λ for the α-phase of OPLS-UA methanol at $T = 150\,\mathrm{K}$, $\rho = 0.01878\,\mathring{\mathrm{A}}^{-3}$ and $k = 8.0\,\mathrm{kcal/molÅ^2}$. 25 λ-values were used.

A.2.2 Comparison between CH3 and oxygen as central atom

For methanol it is not clear which interaction site should be used as "central atom". To investigate possible differences in the resulting absolute free energies we conducted simulations using CH3 and oxygen sites as "central atoms". Results are shown in FIG.A.7 with numerical values given in TABLE A.3, suggesting that for sufficiently long simulation times the choice of the "central atoms" does not matter for OPLS-UA methanol.

A.2.3 States for initial phase coexistence and tracing of the coexistence line

For methanol, we use smaller systems as compared to the study of argon. As a consequence, larger intervals for temperatures and pressures can be chosen when conducting the simulations in the isobaric-isothermal ensemble. FIG.A.8 depicts the construction of the initial coexistence point utilizing simulations of the α-solid phase, the β-solid phase and the liquid phase. These simulations were conducted at $p = 1\,\mathrm{bar}$ and temperatures between $150\,\mathrm{K}$ and $300\,\mathrm{K}$, where temperature intervals of $\Delta T = 10\,\mathrm{K}$ were sufficient to achieve phase space overlap.

To compute the phase diagram (only α-liquid transition) up to $7000\,\mathrm{bar}$ we carried out 9 additional simulations in both phases. A list of all temperatures and pressures is given in the *simulation_materials/methanol_npt_states.dat* file of the supporting information.
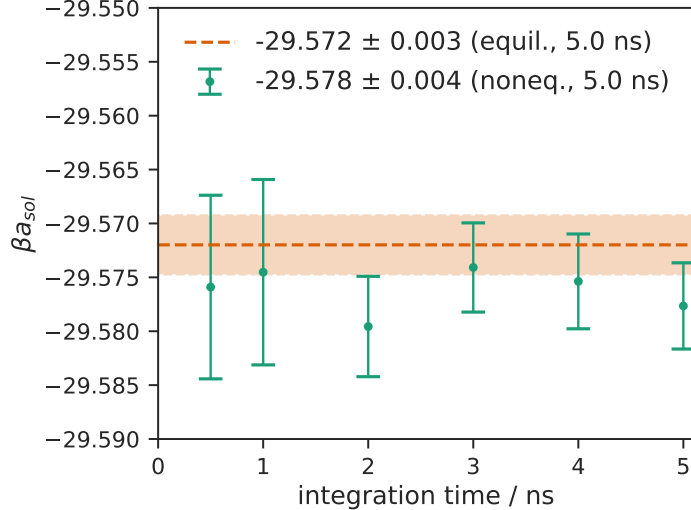
Figure A.6: Absolute free energy of the α-solid phase of OPLS-UA methanol from non-equilibrium simulations (symbols with error bars) over switching time compared to equilibrium simulations (MBAR, dashed line with uncertainty as shaded region) at $T = 150\,\mathrm{K}$, $\rho = 0.01878\,\mathrm{\mathring{A}}^{-3}$ and $k = 8.0\,\mathrm{kcal/mol\mathring{A}^2}$. The de Broglie wavelength is set to $\Lambda = 1.0\,\mathrm{\mathring{A}}$.

A.3 TIP4P/2005 water

A.3.1 Solid phase details

Analogously to OPLS-UA methanol, FIG.A.9 and FIG.A.10 depict the two contributions of the absolute free energy of TIP4P/2005 ice Ih. As before, computation of $\beta\Delta a_0$ requires more intermediate λ-values than computation of $\beta\Delta a_1$. Interestingly, the slope in the vicinity of $\lambda = 0$ is less steep than for methanol which could stem from the fact that water is a symmetric molecule and the central tether is applied almost at the center-of-mass position of the molecule. Missing uncertainties in the plots are reported in the main manuscript.

FIG.A.11 depicts a comparison between the absolute free energy as obtained via non-equilibrium simulations and as obtained from MBAR.

A.3.2 Liquid phase details

For the free energy of the liquid phase we traverse a path from a Lennard-Jones fluid (using the full TIP5P/2005 Lennard-Jones parameters for oxygen) to the TIP5P/2005 potential by turning on coulombic interactions introducing a linear coupling parameter λ. The reference state is calculated from the equation of state of Thol et al.[3] For the ideal gas contribution we use $\Lambda = 1\,\mathrm{\mathring{A}}$ for the de Broglie wavelength. A visualization of the free energy pathway is presented in FIG.A.12 and an overview of all contributions to the absolute free energy is given in TABLE A.4.
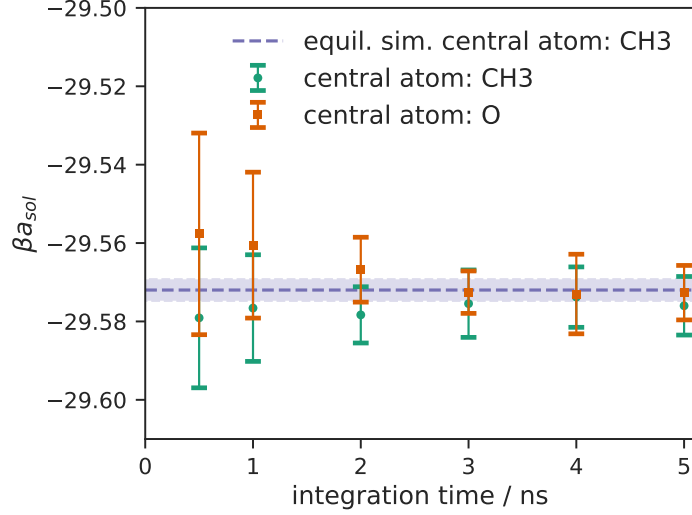
Figure A.7: Absolute free energy of the α-solid phase of OPLS-UA methanol from non-equilibrium simulations for CH3 and oxygen (O) as "central atoms" (symbols) compared to equilibrium results via MBAR at $T = 150\,\mathrm{K}$, $\rho = 0.01878\,\mathrm{\AA}^{-3}$ and $k = 8.0\,\mathrm{kcal/molÅ^2}$. Numerical data of the results are presented in TABLE.A.3.

A.3.3 States for initial phase coexistence and tracing of the co-existence line

For water it is sufficient to use larger intervals for temperatures and pressures when conducting the simulations in the isobaric-isothermal ensemble. For simulations at $p = 1\,\mathrm{bar}$ we use temperature intervals of $\Delta T = 10\,\mathrm{K}$ between $200\,\mathrm{K}$ and $300\,\mathrm{K}$. 5 additional simulations for both phases are conducted to compute the phase diagram up to a pressure of $6000\,\mathrm{bar}$.

Table A.3: Absolute free energy of the α-solid of OPLS-UA methanol from non-equilibrium simulations for CH3 and oxygen as "central atoms" for different switching times. Statistical uncertainties are determined from 5 independent runs. Raw simulation data as well as a post-processing script to generate the results presented here (jupyter notebook) are given in the *simulation_material/meoh_alpha_non_equilibrium_evaluation* folder.

time [ns]	βa_0	$\beta\Delta a_0$	$\delta\beta\Delta a_0$	$\beta\Delta a_1$	$\delta\beta\Delta a_1$	$\beta\Delta a_{\text{sol}}$	$\delta\beta\Delta a_{\text{sol}}$
CH3							
0.5	2.157	8.275	0.016	-40.012	0.008	-29.579	0.018
1.0	2.157	8.272	0.013	-40.006	0.005	-29.577	0.014
2.0	2.157	8.272	0.006	-40.008	0.005	-29.578	0.007
3.0	2.157	8.275	0.008	-40.008	0.004	-29.575	0.009
4.0	2.157	8.272	0.007	-40.003	0.003	-29.574	0.008
5.0	2.157	8.274	0.007	-40.008	0.002	-29.576	0.007
Oxygen							
0.5	2.157	8.286	0.025	-40.001	0.007	-29.558	0.026
1.0	2.157	8.283	0.017	-40.001	0.007	-29.561	0.019
2.0	2.157	8.276	0.008	-40.001	0.001	-29.567	0.008
3.0	2.157	8.270	0.005	-40.000	0.001	-29.573	0.005
4.0	2.157	8.272	0.010	-40.002	0.002	-29.573	0.010
5.0	2.157	8.271	0.007	-40.001	0.002	-29.573	0.007

Table A.4: Contributions to the absolute free energy of TIP4P/2005 liquid as computed via MBAR ($\beta\Delta a$, turning on electrostatic interactions) and the equation of state of Thol et al. at $T = 298\,\text{K}$ and $\rho = 0.03332\text{Å}^{-3}$. For the free energy of the ideal gas, $\ln(\rho\Lambda^3) - 1$, a value for the de Broglie wavelength of $\Lambda = 1\,\text{Å}$ was used.

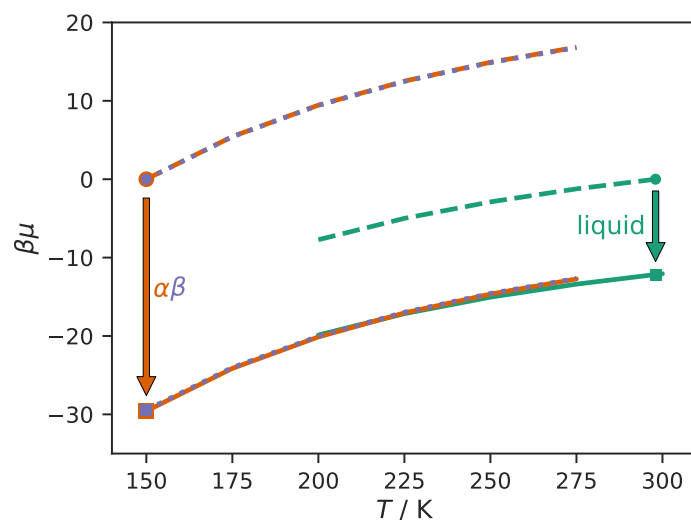| βa_{ig}	βa_{res}	$\beta\Delta a_{\text{liq}}$	βa_{liq}
-4.4016	2.1198	-13.3191(13)	-15.6000(13)

Figure A.8: Construction of the initial coexistence point of OPLS-UA methanol via NpT simulations at $p = 1\,\text{bar}$ from the liquid (green), α-solid and β-phase.
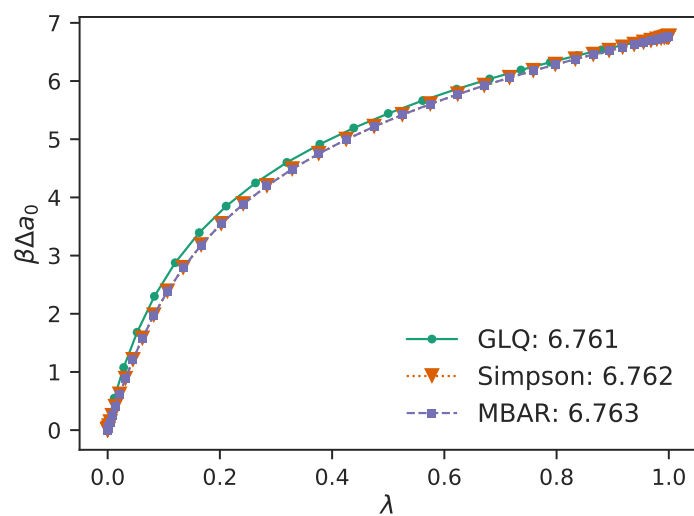
Figure A.9: Free energy path for $\beta \Delta a_0$ over the coupling parameter λ for the Ih-phase of TIP4P/2005 water.
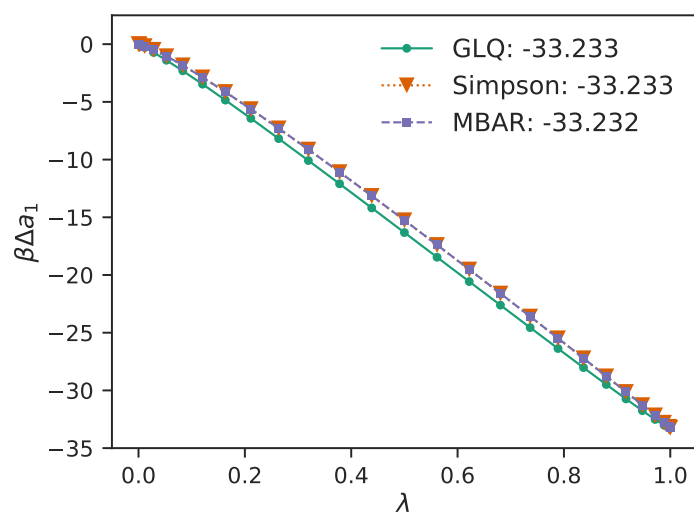

Figure A.10: Free energy path for $\beta \Delta a_1$ versus the coupling parameter λ for the Ih-phase of TIP4P/2005 water. 25 λ-values were used.
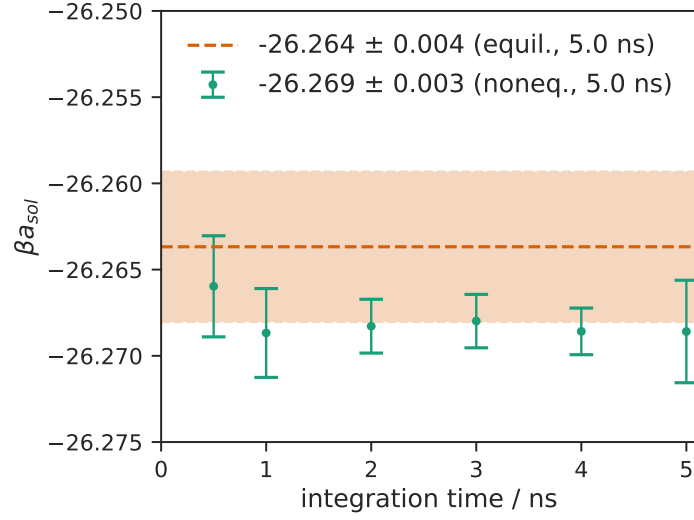
Figure A.11: Absolute free energy of the Ih-phase of TIP4P/2005 water from non-equilibrium simulations (symbols with error bars) over switching time as compared to equilibrium simulations (MBAR, dashed line). The de Broglie wavelength is set to $\Lambda = 1.0\,\text{Å}$.
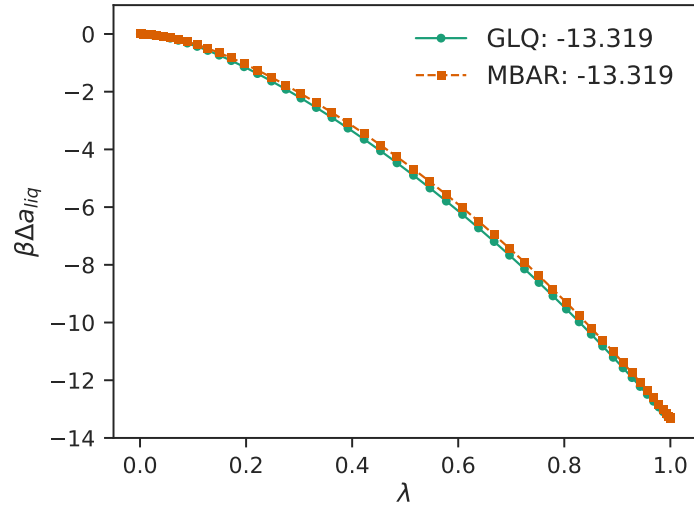

Figure A.12: Free energy difference between a Lennard-Jones fluid and TIP4P2005 water at $T = 298\,\text{K}$ and $\rho = 0.03332\text{Å}^{-3}$. The coupling parameter λ turns on coulombic interactions.

References

1. Mastny, E. A. & de Pablo, J. J. Melting line of the Lennard-Jones system, infinite size, and full potential. *J. Chem. Phys.* **127,** 104504 (2007).

2. Freitas, R., Asta, M. & de Koning, M. Nonequilibrium free-energy calculation of solids using LAMMPS. *Comput. Mater. Sci.* **112,** 333–341 (2016).

3. Thol, M. *et al.* Equation of State for the Lennard-Jones Fluid. *J. Phys. Chem. Ref. Data* **45,** 023101 (2016).