Molecular dynamics study of ion clustering in concentrated electrolyte solutions for the estimation of salt solubilities

Dominik Schaefer, Maximilian Kohns*

[a]*Laboratory of Engineering Thermodynamics (LTD), TU Kaiserslautern, Erwin-Schroedinger-Str. 44, 67663, Kaiserslautern, Germany*

Abstract

The structure and formation of ion clusters for different 1:1 electrolytes in aqueous solution were evaluated over a wide concentration range using molecular dynamics (MD) simulations. Six molecular models for alkali halides from the literature were investigated – four different models for NaCl, one model for KCl, and one model for NaI – all in combination with the SPC/E or TIP4P-Ew water models. For all model combinations, solubility data from thorough solubility calculations are available in the literature. Since such calculations are challenging and computationally very demanding, in this work we propose a simple approach for solubility estimation, which can be used on the fly or as a simple post-processing step in standard MD simulations. This approach is based on analyzing the ion clusters in solution using a clustering algorithm from the literature. In essence, we suggest that the solubility is reached at the concentration for which clusters containing six or more ions are found in solution. We compare our approach to an empirical rule for solubility estimation from the literature, which is based on the number of contact ion pairs only. For the investigated molecular models, our new approach yields results closer to the solubility data reported in the literature compared to the empirical rule from the literature. Altogether, considering its simplicity and low computational cost, it provides reasonable solubility estimates that can, e.g., also be used during the development of new models, which would be infeasible with thorough solubility calculations. Furthermore, the cluster analysis provided interesting insights into the possible formation of a hydrate in the case of the investigated NaI model.

Keywords: molecular dynamics, cluster analysis, structural properties, electrolyte solutions, solubility

1. Introduction

Molecular modeling and simulation has emerged as a versatile tool for studying the properties of fluids. The recent developments in that field with respect to electrolyte solutions have been reviewed by Panagiotopoulos [1] and by Nezbeda et al. [2]. However, irrespective of the recent progress made, with molecular simulations of electrolyte solutions there is always one potential issue: the salt solubility. Since the time span that can be covered in typical molecular dynamics (MD) simulations is far too short for establishing the stable equilibrium state between a saturated solution and a salt crystal, one may actually simulate what could be considered a metastable state. In the worst case, simulations are carried out at concentrations beyond the solubility, which can lead to invalid results of properties which are assumed to be determined for a homogeneous solution.

Knowledge of the salt solubility of the studied system would of course solve this problem. However, calculating the solubility rigorously is computationally very demanding and, hence, for most molecular models for electrolyte solutions, even popular ones, the solubility is not known. Taking a look at the few models for which solubility data are available in the literature shows a wide variety of results: as an example, for NaCl in aqueous solution at 298 K, some models vastly underestimate the experimental solubility, while others yield reasonable predictions [3, 4]. This is true for both the well-established class of non-polarizable models and for the class of polarizable models, which have attracted some interest in recent years [5–7]. The issue of salt solubility is of course closely linked to the parametrization of the investigated molecular models; however, we will not address model parametrization explicitly in the present work but rather focus on approaches for determining the solubility for a given model parametrization.

In the present work, we distinguish two types of approaches for determining the solubility. With the term *solubility calculation*, we refer to methods that rigorously calculate the solubility for a certain molecular model of an electrolyte solution. Such methods are computationally demanding (see below). With the term *solubility estimation*, we refer to methods that rather yield a rough estimate of the solubility, but which in turn work with much cheaper simulations. The main focus of the present work is to propose and assess a new scheme for solubility estimation.

*maximilian.kohns@mv.uni-kl.de

Two methods are available in the literature for rigorous solubility calculations: the *chemical potential method* [3, 5, 8–17] and the *direct coexistence method* [6, 17–22]. Calculations with the *chemical potential method* consist of two steps. First, the chemical potential of the salt in solution is determined as a function of concentration at constant temperature and pressure. Second, the chemical potential of the solid salt crystal is determined in a separate simulation at the same temperature and pressure. The solubility is then obtained as the concentration at which the chemical potential of the salt in solution equals that of the solid.

However, the chemical potential method requires knowing the structure and stoichiometry of the emerging crystal, i.e., whether a hydrate forms and, if it does, how many water molecules are present in it – only then can the chemical potential of the salt in the correct configuration be calculated. Unfortunately, it is impossible to know this before carrying out simulations, such that the only way is to use a trial-and-error approach that tests different crystal configurations and then considers the one with the lowest free energy. Furthermore, the determination of chemical potentials is not trivial and needs sophisticated evaluation schemes. The chemical potential of the salt in solution is usually obtained from a thermodynamic integration-type calculation that requires a very fine discretization along the chosen path [10]. The chemical potential of the solid salt is typically calculated with techniques based on the Einstein crystal method [23]. Nevertheless, despite the methodological difficulties and challenges, determinations of the solubility of the same model combination by different authors yield similar results with the chemical potential method [3, 9, 10, 12, 13, 17].

In the *direct coexistence method* – as the name suggests – the aqueous solution is simulated in direct coexistence with a solid salt phase and the simulation runs until equilibrium is reached. The solubility then corresponds to the salt concentration in the bulk liquid phase in equilibrium. Even though this method only needs a single simulation to determine the solubility, the time needed to reach equilibrium is on the order of microseconds [17–22]. Considering this slow equilibration process, the initial state of the system has to be chosen close to the final equilibrium state (which is of course usually not known a priori), otherwise risking prohibitively long equilibration times. Additionally, size effects of the bulk phases and the area of the solid-liquid interface are crucial for this method [18]. The first results obtained with the direct coexistence method yielded very different results by different groups of authors even for the same model combination [17–22].

The influence of size effects was investigated in recent years and, as a result, the reproducibility significantly improved [6, 18, 22], reaching mutual agreement with results obtained with the chemical potential method.

As discussed above, both methods for the rigorous calculation of the solubility of a salt in solution possess several pitfalls and are computationally very expensive. Especially during the development of new ion models, their computational cost is prohibitive. Hence, it is desirable to have empirical rules that allow for making at least a rough estimate of the solubility. Two such rules were proposed by Benavides and co-workers [4, 9]: the *chemical potential difference rule*, which is based on the difference of the chemical potentials of the salt at infinite dilution and of the crystal, and the *ion pairs rule*, which is based on the number of contact ion pairs, i.e., the number of ions in the first coordination shell around the corresponding counter ion. The *chemical potential difference rule*, however, is not readily applicable, as it still requires expensive calculations for obtaining the chemical potential of the electrolyte at infinite dilution in the solvent and that of the emerging solid salt. The use of this rule only circumvents calculating the concentration dependence of the chemical potential of the salt in solution. The *ion pairs rule*, by contrast, is a solubility estimation method: it can be applied easily and on the fly in standard simulations by sampling the cation–anion radial distribution function (RDF). The MD simulations used by Benavides and co-workers for determining that RDF used several hundred to a few thousand particles and took up to 20 ns. For this, the computational demand is several orders of magnitude lower than that needed for calculations of chemical potentials or applying the direct coexistence method. Benavides and co-workers formulated two versions of the ion pairs rule. First, they suggested that the solubility of 1:1 electrolytes with moderate solubility (i.e., lower than 10 $\mathrm{mol\,kg^{-1}}$) is at the concentration for which the number of contact ion pairs is 0.075 [9]. This did hold for the NaCl models of Joung and Cheatham [24] and Smith and Dang [25] when combined with SPC/E water [26]. In a second publication [4], they redeemed the rule as follows: the solution is most likely supersaturated if the number of contact ion pairs is greater than 0.5, i.e., the concentration range in which the salt is soluble is at a concentration for which the number of contact ion pairs less than or equal to 0.5.

The general idea of the ion pairs rule is that the solubility of a salt should in some way be related to the microscopic structure of the fluid. The formation of ion pairs is of course a logical first and important step

in the homogeneous nucleation of a salt crystal from solution. It is widely accepted that the solid phase nucleates in a two-step mechanism [27–34]: in the first step, ion clusters with a short life span and an amorphous structure form. These clusters have to reach a critical size to become stable. In the second step, ions inside of sufficiently large, stable clusters rearrange to form ordered lattice structures of the corresponding salt crystal. These ordered structures inside the cluster are surrounded by an amorphous layer of ions. In this work, we are interested in the early stages of the nucleation process, as the existence of small clusters of a certain size might indicate that, eventually, a salt crystal will form. The studies by Lanaro and Patey [31] and Zimmermann et al. [30, 35] are of particular relevance for the present work, since both groups of authors observed that the smallest clusters that can be considered to be stable over an extended period of time contain six ions. However, both Lanaro and Patey as well as Zimmermann et al. studied only aqueous NaCl solutions as described by combining the Joung-Cheatham NaCl model [24] with the SPC/E water model [26]. It is unclear whether this threshold cluster size also holds for other molecular models of NaCl or other salts.

Following the aforementioned studies on crystal nucleation and inspired by the ion pairs rule, in the present work, we systematically investigate the structure of aqueous electrolyte solutions at concentrations below, at, and beyond the solubility. We investigated cation–anion cluster formation with a neighbor search algorithm [36] to identify the coordination of anions and cations from configurational data, which are readily available from MD simulations. With an observable derived from the cluster size distribution, a new solubility estimation method was developed and compared to the ion pairs rule by Benavides et al. [4, 9]. To anticipate some of the findings from the present work, we formulate this new *cluster rule* as follows: when running MD simulations with different salt concentrations at the same temperature and pressure, the estimated salt solubility is the lowest initial composition of the solution for which clusters containing six or more ions are observed.

2. Molecular Modeling and Simulation

In the present work, we carry out MD simulations of aqueous solutions of different alkali halide salts as described by different molecular models from the literature. The models and some technical simulation details are explained in Sections 2.1 and 2.2, respectively. The obtained simulation re-

sults are evaluated focusing on two aspects: the number of contact ion pairs needed for using the ion pairs rule by Benavides et al., and the distribution of ion cluster sizes for the development and assessment of the cluster rule, see Sections 2.3 and 2.4, respectively.

2.1. Molecular Models

All models for the ions and water used in this work consist of one Lennard-Jones (LJ) interaction site and up to three point charge interaction sites (PC). The potential energy u_{ij} of the interaction of two particles i and j is given by

$$\begin{aligned} u_{ij} &= u_{ij}^{\mathrm{LJ}} + u_{ij}^{\mathrm{PC}} \\ &= \sum_{a=1}^{M_i^{\mathrm{LJ}}} \sum_{b=1}^{M_j^{\mathrm{LJ}}} 4\varepsilon_{ijab} \left[\left(\frac{\sigma_{ijab}}{r_{ijab}} \right)^{12} - \left(\frac{\sigma_{ijab}}{r_{ijab}} \right)^{6} \right] \\ &\quad + \sum_{c=1}^{M_i^{\mathrm{PC}}} \sum_{d=1}^{M_j^{\mathrm{PC}}} \frac{1}{4\pi\epsilon_0} \frac{q_{ic} q_{jd}}{r_{ijcd}} \,. \end{aligned} \tag{1}$$

Here, *a*–*d* are indices for the interaction sites, M_i is the total number of interaction sites of a particle i, σ_{ijab} and ε_{ijab} are the LJ size and energy parameters corresponding to the LJ interaction site a of particle i and site b of particle j, while q_{ic} and q_{jd} denote the magnitudes of the point charges in an analogous manner. Furthermore, r is the distance between interaction sites and ϵ_0 is the vacuum permittivity. Interactions between unlike LJ interaction sites are described using the Lorentz-Berthelot combining rules [37, 38]

$$\sigma_{ijab} = \frac{\sigma_{ia} + \sigma_{jb}}{2} \tag{2}$$

$$\varepsilon_{ijab} = \sqrt{\varepsilon_{ia} \varepsilon_{jb}} \,. \tag{3}$$

All molecular models for ions studied in this work consist of a single LJ interaction site with a superimposed point charge of $+1$ e (for the Na^+ and K^+ cations) or -1 e (for the Cl^- and I^- anions). Several ion models from the literature are considered, which only differ in the choice of the two LJ parameters σ and ε. We investigate the NaCl models by Smith and Dang (SD) [25], by Joung and Cheatham (JC) [24], and by Reiser et al.

(RDVH) [39, 40]. Additionally, the KCl model by Joung and Cheatham [24] and the NaI model by Reiser et al. [39, 40] are investigated. An overview of the parameters of these ion models is given in Table 1. Throughout the present work, we use two well-established water models: the SPC/E model [26] (for all electrolyte solutions but one) and the TIP4P-Ew model [41] (in combination with the corresponding JC ions for NaCl). This choice of ion models in combination with the water models is motivated by the fact that data on the salt solubility in these systems are available in the literature, which are compiled in Table 2.

Table 1: Parameters of the molecular models of ions studied in this work. All ion models consist of a single LJ site with a superimposed point charge. All models are used together with the SPC/E water model, except for the JC-TIP4P-Ew models, which are used together with the TIP4P-Ew water model.

species	model name	σ Å	ε/k_B K	q e	ref.
Na^+	SD	2.35	65.419	1	[25]
	JC	2.16	177.456	1	[24]
	JC-TIP4P-Ew	2.1845	84.769	1	[24]
	RDVH	1.89	200	1	[39, 40]
K^+	JC	2.84	216.236	1	[24]
Cl^-	SD	4.4	50.322	−1	[25]
	JC	4.83	6.434	−1	[24]
	JC-TIP4P-Ew	4.9178	5.869	−1	[24]
	RDVH	4.41	200	−1	[39, 40]
I^-	RDVH	4.78	200	−1	[39, 40]

Table 2: Solubility of the investigated alkali halide models in water at 298 K. The water model is SPC/E throughout, except for JC-TIP4P-Ew-NaCl, for which it is the TIP4P-Ew model. The asterisk (*) indicates the values used as a benchmark in the figures of this work. Statistical uncertainties are given in parentheses.

salt	model	$x_{\mathrm{cat}}^{\mathrm{sol}}$ / mol mol^{-1}	ref.
NaCl	SD	0.011	[13]
		0.011(2)	[10]
		0.011(2)	[3] *
		0.012(1)	[9]
		0.016(7)	[17]
NaCl	JC	0.057(1)	[10]
		0.058(4)	[13]
		0.059(1)	[3] *
		0.059(4)	[10]
		0.063(4)	[12]
		0.074(5)	[17]
NaCl	JC-TIP4P-Ew	0.063(5)	[21] *
NaCl	RDVH	0.085(1)	[3] *
KCl	JC	0.044(1)	[3] *
NaI	RDVH	0.166(4)	[3] *

2.2. Simulation Details

All simulations in this work are molecular dynamics simulations (MD) carried out with the open source program *ms*2 [42]. All simulations are initialized by placing the particles randomly on a cubic grid. Most simulations of the present work are carried out with 4 000 particles, but to study the influence of system size, several simulations are repeated using 20 000 particles instead. The simulation box is cubic with periodic boundary conditions. A fifth-order Gear predictor-corrector scheme [43, 44] is used for integrating the equations of motion. The simulation time step is 1.214 fs. For all simulations, the cut-off radius is 15 Å and long-range interactions are handled with Ewald summation [45] for the electrostatics and the standard tail corrections for the LJ interactions [43].

All simulations are carried out in two steps: First, the density is determined with an NpT simulation. In that simulation, the system is first equilibrated for 50 000 time steps in the NVT ensemble and then for 500 000 time steps in the NpT ensemble, before starting the production run of 1 000 000 time steps. The pressure is set using the Andersen barostat [46] and the temperature is set using the velocity scaling thermostat [43]. Second, an NVT simulation is carried out at the density obtained in the NpT run and at the same temperature and composition. That NVT run consists of 500 000 time steps of equilibration and a production phase of 2 000 000 time steps. To study the influence of the duration of the production phase, some simulations are carried out for 20 000 000 time steps instead. In the NVT run, we sample the radial distribution function (cf. Section 2.3) and perform a cluster analysis (cf. Section 2.4). The instantaneous particle positions used for the cluster analysis are saved every 100 000 time steps. For sampling the RDF, the range up to the cut-off radius is discretized into 500 equally sized bins for which particle positions were recorded every 10 000 time steps.

All literature solubility data used as a baseline were reported for 298 K and 1 bar. Hence, all molecular simulations in this work were carried out at that temperature and pressure. We refrain from repeating this throughout the manuscript. For each salt, several simulations were carried out at different concentrations. Throughout this work, we use the true mole fraction of the cation to denote composition. It is defined as

$$x_{\mathrm{cat}} = \frac{n_{\mathrm{cat}}}{n_{\mathrm{cat}} + n_{\mathrm{an}} + n_{\mathrm{solv}}} \,, \tag{4}$$

in which n_i is the true mole number of species i (cation 'cat', anion 'an', and solvent 'solv').

2.3. Number of Contact Ion Pairs

As a simple measure for the arrangement of ions in solution, the radial distribution function (RDF) between the cation 'cat' and the anion 'an' is defined as [43]

$$g_{\text{cat–an}}(r) = \frac{1}{\bar{\rho}_{\text{an}}} \frac{\mathrm{d}N_{\text{an}}}{4\pi r^2 \mathrm{d}r} \ , \tag{5}$$

in which $\bar{\rho}_{\text{an}}$ is the total number density of anions in the simulation box, N_{an} is the number of anions found in a shell with a thickness of $\mathrm{d}r$ around the central cation, and r is the distance between the central ion and the current shell. With the cation–anion RDF, the number of anions in the first coordination shell around the central cation, $n_{\text{cat–an}}$, is determined by

$$n_{\text{cat–an}} = 4\pi\bar{\rho}_{\text{an}} \int_0^{r_{\text{min},1}} g_{\text{cat–an}}(r)\, r^2 \mathrm{d}r \ . \tag{6}$$

The upper integration limit $r_{\text{min},1}$ is the position of the first minimum in the cation–anion RDF. For the 1:1 electrolytes studied in this work, $n_{\text{cat–an}}$ corresponds to the number of contact ion pairs [4, 9, 47], which will be denoted as n_{CIP} in the following. With n_{CIP} defined by Eq. (6), evaluating the ion pairs rule by Benavides et al. [4, 9] is straightforward: if $n_{\text{CIP}} > 0.5$, the solution is supersaturated with the salt, i.e., the composition at which $n_{\text{CIP}} = 0.5$ can be used as an estimate for the salt solubility. We will denote that estimate with the symbol $x_{\text{cat}}^{\text{sol,CIP}}$ throughout this work.

2.4. Cluster Analysis

Instantaneous configurations, i.e., snapshots of the simulated system as a function of simulation time, are analyzed with a cluster analysis tool provided by the visualization tool *Ovito* [36]. The employed cluster search algorithm is shown schematically in Figure 1. The cluster algorithm searches for neighboring ions of opposite charge in a radius r^{C} around the central ion. In the present work, we use the value of the first minimum position in the cation-anion RDF as that radius r^{C}. All particles that are within the search radius are considered to be part of the same cluster. Single ions are counted as a cluster of size 1.

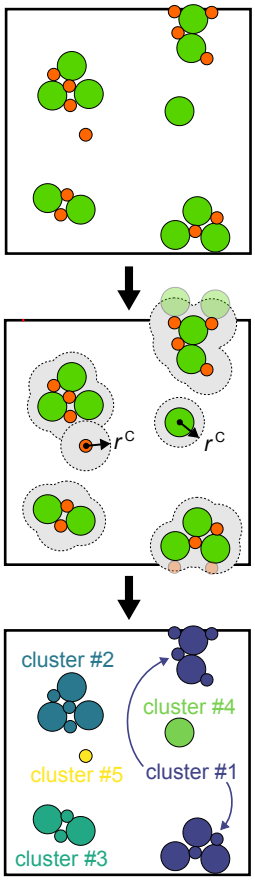


Figure 1: Schematic of the cluster algorithm [36]. Top: a configuration of the ions from a simulation; anions are shown in green, cations in orange. Water molecules are not shown to improve clarity. Middle: neighbor search in the radius r^{C} around each ion. Around each ion, only the corresponding counter ion is being searched for in the radius r^{C} and identified neighbors are grouped into a common cluster. The transparent ions represent neighboring counter ions detected on the other side of the periodic boundary conditions. Bottom: categorization of the ions into clusters of the same color.

In this work, the information obtained by the cluster analysis is condensed into a single, intensive observable: the fraction of the number of ions in clusters that contain j or more ions, denoted as $\bar{X}^{\text{C-}j}$ in the following. It is defined as

$$\bar{X}^{\text{C-}j} = \frac{\left\langle \sum_{i=1}^{M^{\text{C-}j}} N_i^{\text{C}} \right\rangle}{N_{\text{cat}} + N_{\text{an}}} \,, \tag{7}$$

where the summation runs over all ion clusters in the solution, $M^{\text{C}-j}$ is the number of clusters with j or more ions, N_i^{C} is the number of ions in the cluster i, and N_{cat} and N_{an} are the total numbers of anions and cations in the simulation box, respectively. The angular brackets denote the ensemble average, which in this case corresponds to the time average of the investigated length of the production run. For a given state point of an electrolyte solution, the property $\bar{X}^{\text{C-}j}$ depends on three variables: the minimum cluster size j considered in the evaluation, the neighbor search radius r^{C} (which is equivalent to the first minimum position in the cation–anion RDF), and the length of the production run τ. Furthermore, by virtue of the molecular simulations as such, the cluster distribution might be influenced by the system size. The potential influences of all these parameters on the outcome of the cluster analysis are investigated in Section 3.1.

3. Results and Discussion

3.1. Influence Parameters on the Property $\bar{X}^{C\text{-}j}$

3.1.1. Influence of the Minimal Cluster Size j and the System Size

Figure 2 shows the cluster size distribution of RDVH-NaCl in SPC/E water for three different salt concentrations and two different system sizes: 4 000 particles (as used in the remainder of this work) and 20 000 particles.

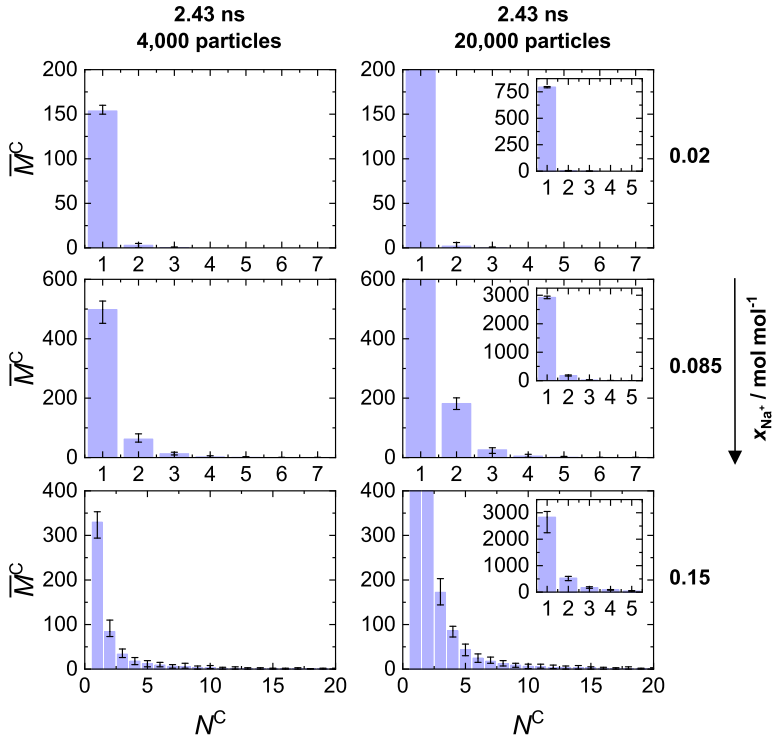


Figure 2: The mean number of clusters $\bar{M}^{\mathrm{C}}$ as a function of the number of ions per cluster N^{C} at 298 K and 1 bar for simulations for different cation mole fractions of the RDVH-NaCl model in SPC/E water. Simulations in the left column had the same system size and the same length of the production run as used in the rest of this work. The right column shows simulations with exactly the same simulation settings, only the number of particles is increased five-fold. Plots that are in the same row have the same axis limits for $\bar{M}^{\mathrm{C}}$ to enable a straightforward comparison between the different system sizes, while the insets in the right column focus on the smallest cluster sizes with the highest number of clusters. The concentration shown in the middle row (i.e., 0.085 mol mol^{-1}) corresponds to the salt solubility (cf. Table 2). The error bars contain 98 % of the data points of the respective cluster size obtained over the course of the production run of each simulation. The values of the cluster radius r^{C} are in the range of 3.58 Å – 3.83 Å (cf. Figure 4).

Note that simulations were carried out at many more concentrations, but only those three are shown here to keep the presentation as simple as possible. Both system sizes show that at the lowest concentration ($x_{\mathrm{Na^+}} = 0.02\ \mathrm{mol\,mol^{-1}}$), only single ions, ion pairs, and very few clusters containing three ions are found. At the concentration that is close to the salt solubility ($x_{\mathrm{Na^+}} = 0.085\ \mathrm{mol\,mol^{-1}}$), again single ions are predominant, the extent of ion pairing becomes quite noticeable, and clusters of up to six ions are observed. At the highest concentration ($x_{\mathrm{Na^+}} = 0.15\ \mathrm{mol\,mol^{-1}}$), corresponding to a highly supersaturated solution, clusters of up to 20 ions are found. In other simulations at similar concentrations, occasionally also somewhat larger clusters were observed. These clusters had an elongated, filamentous, and amorphous shape. One example of such a large cluster is discussed in the Appendix A. Similarly shaped clusters have been reported by Lanaro and Patey [31]. In general, for all concentrations, the cluster distribution peaks at 1, which means that most ions are present as single ions, even at high concentrations. Moreover, with increasing cluster size, the cluster size distribution rapidly decreases and approaches zero, resembling an exponential decay.

Qualitatively, irrespective of system size, the cluster distributions are similar if the same salt concentration is investigated. At the same salt concentration, the only difference between the smaller and larger systems is the actual number of clusters of a certain size, which is of course larger for the larger systems. Quite interestingly, at the lowest concentration ($x_{\mathrm{Na^+}} = 0.02\ \mathrm{mol\,mol^{-1}}$), the smaller system size tends to favor ion pairing over the larger one, but at and above the solubility, that trend reverses. However, there is considerable uncertainty in this interpretation due to the fluctuations in the cluster size distributions indicated by the error bars, especially at low concentrations. Notwithstanding and most importantly for the remainder of the present work, clusters containing six or more ions are only found at or above the salt solubility independent of system size.

Figure 3 shows the lowest mole fraction $x^*_{\mathrm{Na^+}}$ at which the property $\bar{X}^{\mathrm{C}\text{-}j}$ (cf. Eq. (7)) is greater than zero as a function of j for a system size of 4 000 particles. In other words, $x^*_{\mathrm{Na^+}}$ is the lowest mole fraction at which clusters of size j or higher are observed at least once during the simulation.

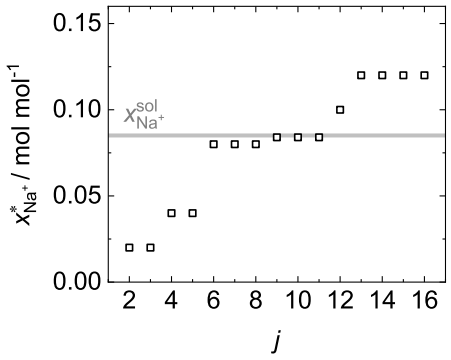


Figure 3: The lowest mole fraction $x^*_{\mathrm{Na^+}}$ at which $\bar{X}^{\mathrm{C}\text{-}j}$ is greater than zero (i.e., clusters containing j or more ions are found) as a function of j at 298 K and 1 bar for the RDVH-NaCl model in SPC/E water. The investigated systems contained 4 000 particles. The gray shaded area depicts the solubility $x^{\mathrm{sol}}_{\mathrm{Na^+}}$ by Mester and Panagiotopoulos [3] including uncertainties. The values of the cluster radius r^{C} are in the range of 3.58 Å − 3.83 Å (cf. Figure 4).

Interestingly, for $6 \leq j \leq 11$, the mole fraction $x^*_{\mathrm{Na^+}}$ forms a plateau, the value of which agrees well with the solubility $x^{\mathrm{sol}}_{\mathrm{Na^+}}$ from the literature [3]. It is gratifying that the value $j = 6$ is in accord with the smallest stable cluster size reported in the literature [30, 31, 35], but for a different NaCl model than the one studied here (JC instead of RDVH) and for much longer simulations. Based on these findings, we formulate the empirical *cluster rule* as follows: we hypothesize that the lowest cation mole fraction $x^{\mathrm{sol,C}\text{-}6}_{\mathrm{cat}}$ for which $\bar{X}^{\mathrm{C}\text{-}6}$ is greater than zero, i.e., the concentration for which the first clusters containing six or more ions are observed, can serve as a rough estimate for the salt solubility. We will assess the cluster rule from Section 3.2 onwards.

3.1.2. Influence of the Ion Mole Fraction on the RDF and the First Minimum

The first minimum position of the cation–anion RDF was chosen as the cluster search radius r^{C}. This choice obviously has some influence on the obtained cluster size distribution and, in turn, also on $\bar{X}^{\mathrm{C}\text{-}6}$. Therefore, we investigated the influence of the salt concentration on the RDF, with a special attention on the position of the first minimum. Figure 4 shows the cation–anion RDFs and the corresponding positions of the first minimum from simulations of RDVH-NaCl in SPC/E water at different salt concentrations.

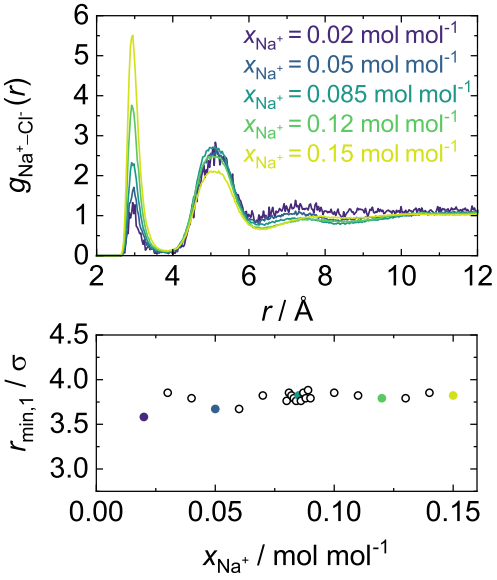


Figure 4: The cation–anion radial distribution functions (top panel) and the corresponding position of the first minimum $r_{\mathrm{min},1}$ (bottom panel) as a function of the mole fraction of the cation at 298 K and 1 bar for the RDVH-NaCl model in SPC/E water. The colors of the filled circles in the bottom panel correspond to the radial distribution function shown in the same color in the top panel. $r_{\mathrm{min},1}$ is used as the neighbor search radius r^{C} for the cluster algorithm (cf. Figure 1).

As expected, with increasing concentration the height of the first peak in the cation–anion RDF also increases, which corresponds to an increased number of ion pairs. This furthermore indicates that clustering might be more prevalent at higher concentrations (cf. Figures 2 and 3). The height of the first minimum is roughly the same for all concentrations, while the height of the second maximum decreases with increasing concentration. The positions of the first and second maxima, and, most importantly for the present work, the position of the first minimum, are practically independent of composition, even when considering highly supersaturated solutions. Some fluctuations in the first minimum position with increasing salt concentration are evident at low concentrations, but for solutions in the vicinity of the solubility limit or beyond it, practically the same value is obtained for the first minimum position. Hence, the variation of $r_{\mathrm{min},1}$ with concentration is small, so that its value – which is used as the neighbor search radius r^{C} – has no noticeable impact on the outcome of the cluster algorithm.

3.1.3. Influence of the Length of the Production Run

Figure 5 shows the running average of the property $\bar{X}^{\text{C-6}}$ as a function of simulation time of RDVH-NaCl in SPC/E water for three different salt concentrations. The simulations shown there were ten times longer than the other simulations shown in the rest of this work.

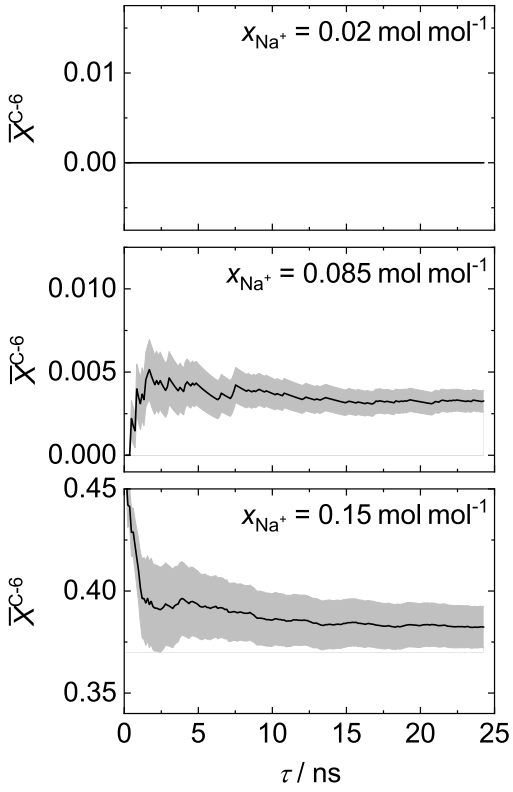


Figure 5: The running average of the property $\bar{X}^{\text{C-6}}$ as a function of production time τ at 298 K and 1 bar for the RDVH-NaCl model in SPC/E water for different salt concentrations. The concentration shown in the middle panel corresponds to the salt solubility (cf. Table 2). The gray shaded area denotes the standard deviation of the data up to the simulation time τ. The values of the cluster radius r^{C} are in the range of 3.58 Å – 3.83 Å (cf. Figure 4).

At the lowest concentration, clusters of size six are never found, even if simulations are run for 24.3 ns. At the salt concentration that is close to the salt solubility ($x_{\mathrm{Na^+}} = 0.085$ mol mol^{-1}), the first clusters containing more than six ions emerge after about 0.4 ns, and $\bar{X}^{\text{C-6}}$ stays greater than zero and slightly increases with simulation time. At the highest concentration, interestingly, $\bar{X}^{\text{C-6}}$ starts from a very high value and then slightly decreases with simulation time, but stays well above zero for all times. Hence, once clusters with six or more ions form, they tend to be stable in solution, and/or grow in size with simulation time. Since the definition of the property $\bar{X}^{\text{C-6}}$ includes all ions in clusters of size six or larger, these growing clusters are explicitly considered in the analysis. Moreover, it seems that a production run of about 2.4 ns is sufficient to determine if clusters with size six or larger will form during a simulation or not.

3.2. Results for the NaCl Models with SPC/E water

Figure 6 illustrates the use and the results of the ion pairs rule of Benavides et al. for the three investigated NaCl models with SPC/E water: it shows the number of contact ion pairs as a function of salt concentration.

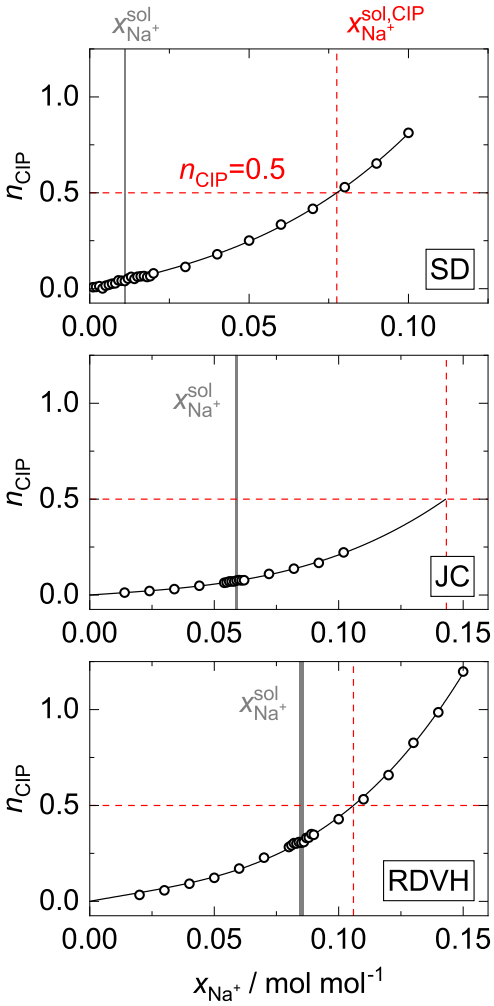


Figure 6: The number of contact ion pairs n_{CIP} as a function of the mole fraction of Na^+ at 298 K and 1 bar of all investigated NaCl models (SD, JC, RDVH) in SPC/E water. The vertical red dashed line corresponds to the solubility $x^{\mathrm{sol,CIP}}_{\mathrm{Na^+}}$ determined with the ion pairs rule [4, 9]. The gray shaded area depicts the solubility $x^{\mathrm{sol}}_{\mathrm{Na^+}}$ reported by Mester and Panagiotopoulos [3] including uncertainties. The solid black line corresponds to a third order polynomial used to determine $n_{\mathrm{CIP}} = 0.5$ and hence $x^{\mathrm{sol,CIP}}_{\mathrm{Na^+}}$.

In order to evaluate $x^{\mathrm{sol,CIP}}_{\mathrm{Na^+}}$ from the ion pairs rule, a third order polynomial was fitted to the simulation results and evaluated for the ion mole fraction for which $n_{\mathrm{CIP}} = 0.5$. In the case of the JC model, even at very high concentrations corresponding to highly supersaturated solutions, no values of n_{CIP} in the proximity of 0.5 were obtained, so that the polynomial had to

be extrapolated quite considerably. However, for the present discussion it suffices to know that the concentration for which $n_{\mathrm{CIP}} = 0.5$ is much larger than the actual solubility, at least by a factor of two. In principle, in all three cases the general statement of the ion pairs rule holds: the number of contact ion pairs at the solubility reported in the literature is lower than 0.5. In the case of the RDVH model, $x_{\mathrm{Na^+}}^{\mathrm{sol,CIP}}$ is actually quite close to the literature solubility value, such that the ion pairs rule provides a useful estimate of the true solubility. However, for the other two models, $x_{\mathrm{Na^+}}^{\mathrm{sol,CIP}}$ is much larger than the solubility reported in the literature: for the JC model, $x_{\mathrm{Na^+}}^{\mathrm{sol,CIP}}$ is three times greater and for the SD model, it is seven times greater. As a result, since $x_{\mathrm{Na^+}}^{\mathrm{sol,CIP}}$ is the estimated upper limit of the solubility, the ion pairs rule leaves a wide margin for the concentration range in which one may assume that the salt is completely soluble when in fact it is not.

For comparison, Figure 7 shows the results of the evaluation of the cluster analysis algorithm. There, the property $\bar{X}^{\text{C-6}}$ is depicted as a function of salt concentration for the three investigated NaCl models. Also, the results of the solubility determined with the ion pairs rule is shown (cf. Figure 6).

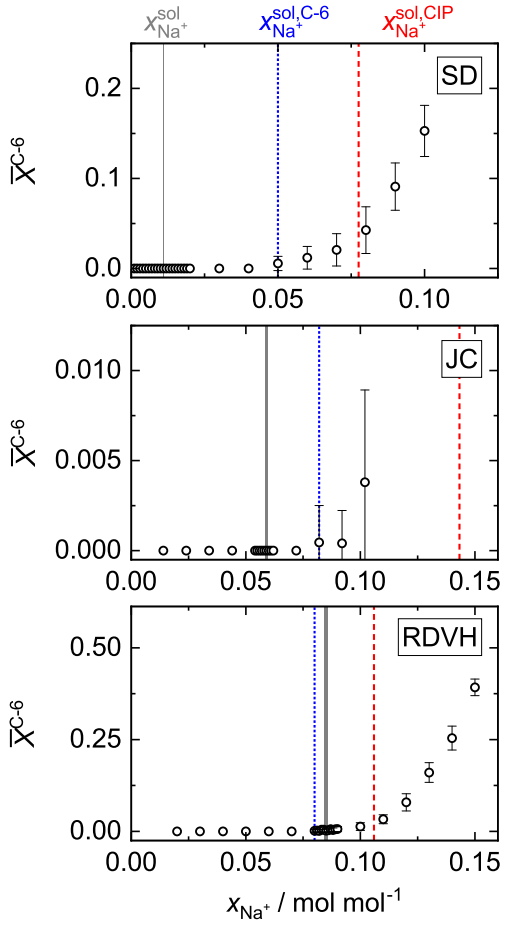


Figure 7: The property $\bar{X}^{\text{C-6}}$ as a function of the mole fraction of Na^+ at 298 K and 1 bar of all investigated NaCl models (SD, JC, RDVH) in SPC/E water. The vertical blue dotted line corresponds to the lowest mole fraction at which clusters with six or more ions occur. The vertical red dashed line corresponds to the solubility $x^{\text{sol,CIP}}_{\text{Na}^+}$ determined with the ion pairs rule [4, 9] (cf. Figure 6). The gray shaded area depicts the solubility $x^{\text{sol}}_{\text{Na}^+}$ reported by Mester and Panagiotopoulos [3] including uncertainties.

It is found that for all three NaCl models investigated here, the concentration at which clusters of six or more ions are found is closer to the solubility reported in the literature than the solubility estimated by the ion pairs rule. In the case of the SD and JC models, the cluster rule overestimates the solubility from the literature, while for the RDVH model, the solubility is slightly underestimated. However, slightly underestimating the solubility is actually preferable compared to overestimating it, since

that prevents studying state points for which the solution is supersaturated. Furthermore, for the SD model both schemes highly overestimate the solubility reported in the literature. However, the SD model for NaCl is quite different from the other two in that it quite drastically underestimates the solubility of real NaCl.

3.3. Results for the JC-TIP4P-Ew-NaCl Model

Figure 8 shows the properties n_{CIP} and $\bar{X}^{\mathrm{C\text{-}6}}$ as a function of the salt concentration for the JC-TIP4P-Ew-NaCl model.

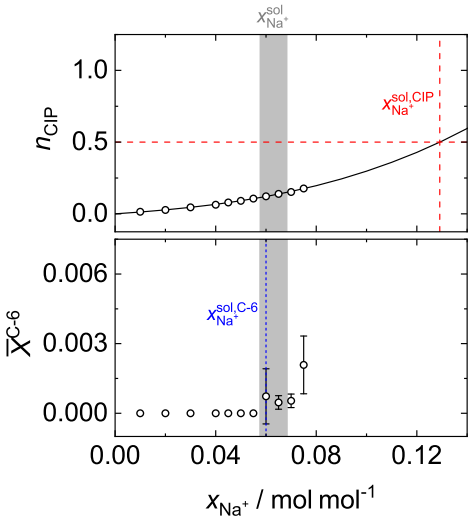


Figure 8: The number of contact ion pairs n_{CIP} (top panel) and the property $\bar{X}^{\mathrm{C\text{-}6}}$ (bottom panel) as functions of the salt concentration at 298 K and 1 bar for the JC-TIP4P-Ew-NaCl model. The vertical blue dotted line corresponds to the lowest mole fraction at which clusters with six or more ions occur. The vertical red dashed line corresponds to the solubility $x^{\mathrm{sol,CIP}}_{\mathrm{Na^+}}$ determined with the ion pairs rule [4, 9]. The gray shaded area depicts the solubility $x^{\mathrm{sol}}_{\mathrm{Na^+}}$ reported by Manzanilla-Granados [21] including uncertainties.

As for the JC-NaCl model with SPC/E, the number of contact ion pairs had to be extrapolated significantly to reach a concentration for which n_{CIP} reaches 0.5. Thus, the ion pairs rule overestimates the solubility of JC-NaCl also in TIP4P-Ew water. By contrast, the cluster rule matches the solubility data reported by Manzanilla-Granados [21] well, showing its usefulness irrespective of the used water model.

3.4. Results for the JC-KCl Model

Figure 9 shows the properties n_{CIP} and $\bar{X}^{\mathrm{C\text{-}6}}$ as a function of the salt concentration for solutions of the JC-KCl model in SPC/E water.

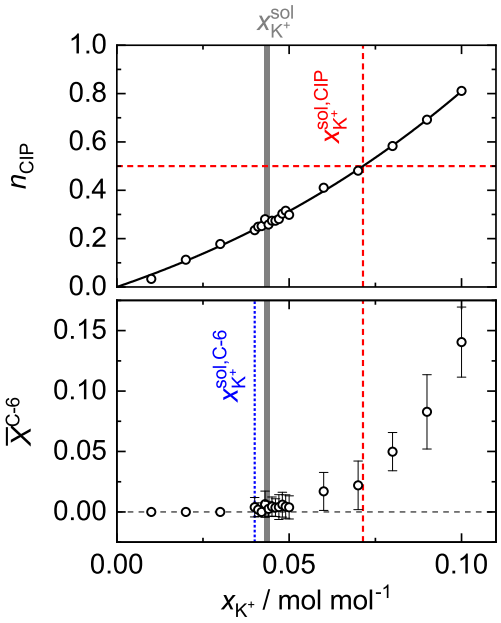


Figure 9: The number of contact ion pairs n_{CIP} (top panel) and the property $\bar{X}^{\mathrm{C\text{-}6}}$ (bottom panel) as functions of the salt concentration at 298 K and 1 bar for solutions of JC-KCl in SPC/E water. The vertical blue dotted line corresponds to the lowest mole fraction at which clusters with six or more ions occur. The vertical red dashed line corresponds to the solubility $x_{\mathrm{K^+}}^{\mathrm{sol,CIP}}$ determined with the ion pairs rule [4, 9]. The gray shaded area depicts the solubility $x_{\mathrm{K^+}}^{\mathrm{sol}}$ reported by Mester and Panagiotopoulos [3] including uncertainties.

The findings are qualitatively similar to the results obtained for the NaCl models. The ion pairs rule yields a broad concentration range in which the salt is estimated to be soluble and overestimates the solubility reported in the literature. The cluster analysis underestimates the solubility slightly, but as mentioned before this is actually the preferred case from a practical standpoint.

3.5. Results for the RVDH-NaI Model

Figure 10 shows the properties n_{CIP} and $\bar{X}^{\mathrm{C\text{-}6}}$ as a function of the salt concentration for solutions of the RDVH-NaI model in SPC/E water.

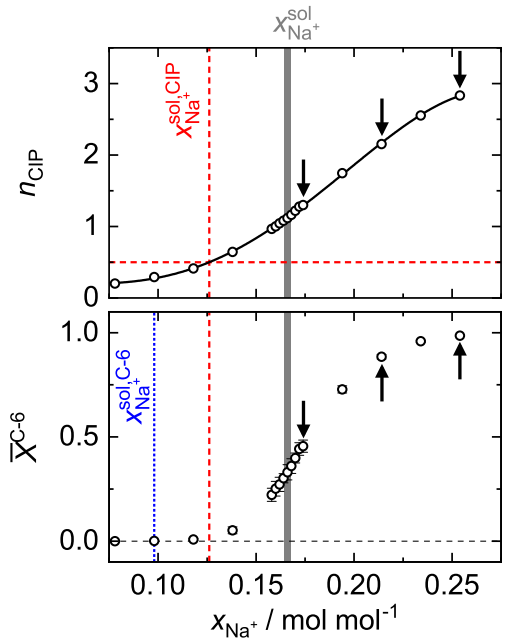


Figure 10: The number of contact ion pairs n_{CIP} (top panel) and the property $\bar{X}^{C\text{-}6}$ (bottom panel) as functions of the mole fraction of Na^+ at 298 K and 1 bar for solutions of RDVH-NaI in SPC/E water. The vertical blue dotted line corresponds to the lowest mole fraction at which clusters with six or more ions occur. The vertical red dashed line corresponds to the solubility $x^{sol,CIP}_{Na^+}$ determined with the ion pairs rule [4, 9]. The gray shaded area depicts the range of the solubility $x^{sol}_{Na^+}$ from the literature [3]. The black arrows indicate the simulations from which the snapshots shown in Figure 11 were captured.

In contrast to the examples discussed before, both rules underestimate the solubility reported in the literature. In agreement with the results reported by Benavides et al. [4] for this model combination, the value of n_{CIP} is about 1.2 to 1.3 at the solubility. In the study of Benavides et al., RDVH-NaI was the only model for which n_{CIP} at the solubility reported in the literature was much larger than 0.5. This finding was attributed to the size difference of the ions, which leads to a lower chemical potential of the NaI crystal.

The values of both n_{CIP} and the property $\bar{X}^{C\text{-}6}$ reach significantly higher values compared to the NaCl and KCl simulations (cf. Figures 6 and 7), which means anions and cations coordinate closely and large clusters are observed. Beyond the solubility, the slopes of both n_{CIP} and $\bar{X}^{C\text{-}6}$ decrease with increasing salt concentration, in contrast to what is observed for the

NaCl and KCl models. The property $\bar{X}^{\text{C-6}}$ approaches a value of 1 at the highest cation mole fraction, which means nearly all ions coordinate in clusters of six or more ions.

To elaborate more on this interesting finding, Figure 11 shows snapshots at the end of three simulations with different salt concentrations. The corresponding cation mole fractions are marked with arrows in Figure 10, i.e., $x_{\text{Na}^+} = 0.174,\ 0.214,\ 0.254\ \text{mol}\,\text{mol}^{-1}$.

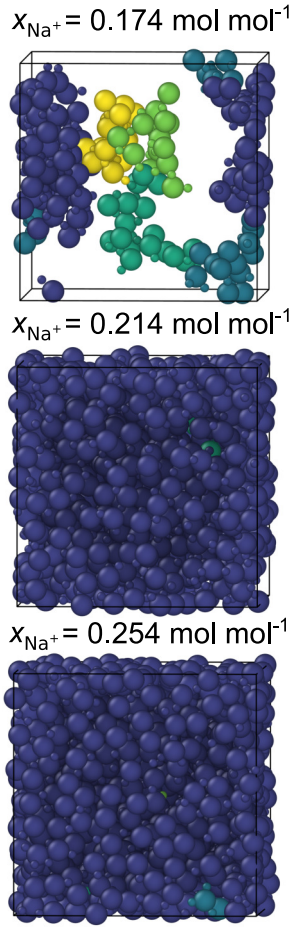


Figure 11: Snapshots taken at the end of the simulation of RDVH-NaI in SPC/E water for three mole fractions beyond the solubility reported in the literature [3] at 298 K and 1 bar (cf. black arrows in Figure 10). Only the five largest ion clusters are shown and categorized by color (dark blue: largest cluster, turquoise: 2$^{\text{nd}}$ largest cluster, dark green: 3$^{\text{rd}}$ largest cluster, light green: 4$^{\text{th}}$ largest cluster, yellow: 5$^{\text{th}}$ largest cluster); water is not shown to improve clarity.

With increasing concentration, more ions coordinate in increasingly bigger clusters, which eventually fill the entire simulation box, as can be seen at the two highest cation mole fractions. At the highest cation mole fraction, at which $\bar{X}^{\text{C-6}}$ approaches 1, the ratio of NaI to water is 1:2.

While this investigation gives no strict proof for it, these findings support the hypothesis that for the combination of the RDVH-NaI model and SPC/E water, the salt precipitates as the dihydrate $NaI \cdot 2\ H_2O$ rather than the pure NaI crystal. It is quite intriguing that also real NaI precipitates as the dihydrate at 298 K and 1 bar [48]. However, the solubility reported by Mester and Panagiotopoulos [3] was determined under the assumption that the pure NaI crystal forms and is therefore not applicable to the dihydrate case. Both methods to calculate the solubility exactly – the direct coexistence method and the chemical potential method – rely on a priori knowledge of the solid phase. The results of our study indicate that investigating the structure of the solution closely, especially with the help of a cluster algorithm, allows for detecting situations in which the formation of hydrates might occur.

This is an important aspect because in reality, the stable form of the precipitating crystal – i.e., whether a hydrate forms and if it does, with how many water molecules – strongly depends on temperature for many electrolytes in aqueous solution. In the case of NaI investigated here, the dihydrate forms at ambient temperature, but below -12.3 °C the pentahydrate $NaI \cdot 5\ H_2O$ forms [48]. Taking NaCl as a further example, at room temperature the pure NaCl crystal forms, but at temperatures below 0 °C, the dihydrate $NaCl \cdot 2\ H_2O$ is the stable solid form [49]. Many alkali and alkaline earth metal halide salts behave similarly, i.e., they form different hydrates depending on temperature [49].

However, it is not guaranteed that molecular models correctly reproduce this behavior. Molecular models might form different hydrates or non-hydrates that might not be expected when simply looking at what the real electrolyte does in experiments. To the best of our knowledge, the only case in which this topic has been thoroughly investigated is the $CaCl_2$ study of Moučka et al. [14]. It is known from experiments that this electrolyte precipitates as the hexahydrate at ambient temperature, but the the dihydrate forms at temperatures above 318 K. However, Moučka et al. found that for the investigated molecular model, only the dihydrate is stable at both temperatures and, hence, the model most likely forms the dihydrate at both temperatures. Investigating the structure of the simulated system

closely, e.g., with the help of a cluster algorithm such as the one used in the present work, is one way to at least keep this important aspect in mind. In general, hydrate-forming systems are interesting candidates to investigate nucleation kinetics in future works, since nucleation events of hydrates are more probable compared to pure ion crystals because water is part of the forming solid and must not be absent from the nucleation site, which is the case for non-hydrates.

4. Conclusions

The structure of electrolyte solutions was investigated with molecular dynamics simulations, using a neighbor search algorithm to identify ion clusters [36]. The information of the cluster size distribution was condensed into a single, intensive property: the fraction $\bar{X}^{\text{C-6}}$ of ions in clusters containing six or more ions. It was found that the lowest concentration at which the first clusters with six or more ions occur, i.e., for which $\bar{X}^{\text{C-6}}$ is greater than zero, is a good indicator to estimate the salt solubility. Together with the ion pairs rule by Benavides et al. [4], this cluster rule was assessed using six molecular models of alkali halides in water for which solubility data are available in the literature. Overall, reasonable agreement was obtained between these reference data and the two empirical rules, given their simplicity, with the cluster rule slightly outperforming the ion pairs rule. For the case of aqueous NaI, the cluster analysis further revealed that the emerging salt crystal might actually be the dihydrate, as is the case for real NaI in aqueous solution and in contrast to what has been assumed in the literature thus far. Altogether, the insights of the present work provide a point of departure for further investigations on how structuring and clustering of ions in electrolyte solutions relate to the salt solubility. Key advantages of the employed cluster distribution analysis include that it is simple to use, e.g., in post-processing, and does not need any knowledge of the emerging crystal, sophisticated evaluation schemes, or elongated simulation times.

Acknowledgments
The present work was conducted under the auspices of the Boltzmann-Zuse Society of Computational Molecular Engineering (BZS) and the simulations were carried out on the Regional University Computing Center Kaiserslautern (RHRK) under the grant TUK-MTD, the High Performance Computing Center Stuttgart (HLRS) under the grant MMHBF2, as well as the Leibniz Supercomputing Centre (LRZ) under the grant $(AMSEL)^2$ (pn56mo). We thank Yunus Sari for his assistance in carrying out molecular simulations and Yannik Sinnwell for his assistance in data assessment.

Appendix A. Structure of Large Clusters Observed at High Salt Concentrations

As stated in Section 3.1.1, among all simulations carried out in the present work, clusters containing more than 20 ions were rare and clusters of this size were not long-lived. Figure 12 shows the biggest cluster observed in this work (excluding the RDVH-NaI simulations), which was found at the highest investigated concentration, $x_{\mathrm{Na^+}} = 0.15\ \mathrm{mol\,mol^{-1}}$, of RDVH-NaCl in SPC/E water.

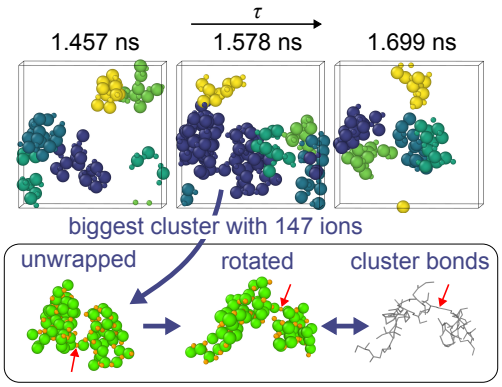


Figure 12: Snapshots from the simulation at the highest concentration of RDVH-NaCl in SPC/E water ($x_{\mathrm{Na^+}} = 0.15\ \mathrm{mol\,mol^{-1}}$). Top panels: snapshots showing the five biggest clusters in three consecutive frames with coloring as in Figure 11 of the main text; the frame that contains the biggest overall cluster with 147 ions is shown in the middle panel, while snapshots of the previous and subsequent configurations are shown to its left and right, respectively. Bottom panels, from left to right: the biggest cluster containing 147 ions unwrapped (i.e., depicted as a coherent object without visual disconnections arising from the periodic boundary conditions), the same unwrapped cluster rotated for improved visualization, and the corresponding cluster bounds used by the cluster algorithm. The red arrows indicate the single cation that connects the two sub-clusters, which build up the large cluster.

This cluster contains 147 ions out of the total of 1 200 for that simulation and has an elongated, filamentous, and amorphous shape. Upon closer investigation, the cluster consists of at least two sub-clusters connected by a single cation (cf. Figure 12, bottom panel). The left sub-cluster consists of 85 ions and the right sub-cluster consists of 62 ions. The merging of several sub-clusters by a chain of single ions is caused by the neighbor-based cluster algorithm used in this work. Other cluster approaches, such as centroid-based clustering (e.g., k-means algorithm [50]) or density-based clustering

(e.g., DBSCAN [51]), may have identified the two individual sub-clusters instead. Notwithstanding, for the topic investigated in the present work, this makes essentially no difference since both cases still yield the same value for the property $\bar{X}^{\text{C-6}}$.

Nomenclature

Abbreviations

an	Anion
cat	Cation
CIP	Contact ion pairs
JC	Models by Joung and Cheatham
LJ	Lennard-Jones interaction site
PC	Point charge interaction site
RDF	Radial distribution function
RDVH	Models by Reiser et al.
SD	Models by Smith and Dang
solv	Solvent

Symbols and Indices

a–d	Indices for interaction sites
ε	LJ energy parameter
ϵ_0	Vacuum permittivity
$g_{\text{cat–an}}$	Cation–anion RDF
i, j	Indices for particles
j	Minimal cluster size
$M_{\text{cat–an}}$	Molar mass of a salt
M_i	Total number of interaction sites of particle i
$\bar{M}^{\text{C}}$	Number of clusters of size N^{C} in a single frame
$n_{\text{cat–an}}$	Number of anions in the first coordination shell around the central cation
n_{CIP}	Number of contact ion pairs (equivalent to $n_{\text{cat–an}}$)
N_i	Number of particles of species i
N^{C}	Number of ions in a cluster
p	Pressure
q	Magnitude of point charges
r	Distance between interaction sites
$\bar{\rho}_i$	Total number density of species i
$r_{\text{min,1}}$	Position of the first minimum of the RDF
σ	LJ size parameter
T	Temperature
V	Volume
$\bar{X}^{\text{C-6}}$	Number of ions in clusters containing six or more ions divided by the total number of ions (cf. text)
$x_{\text{cat}}^{\text{sol}}$	Solubility limit
x_{cat}^{*}	Lowest mole fraction at which the first clusters with j or more ions are observed
$x_{\text{cat}}^{\text{sol,CIP}}$	Solubility estimated with the ion pair rule
$x_{\text{cat}}^{\text{sol,C-6}}$	Solubility estimated with the cluster rule

References

[1] A. Z. Panagiotopoulos, Simulations of activities, solubilities, transport properties, and nucleation rates for aqueous electrolyte solutions, The Journal of Chemical Physics 153 (1) (2020) 010903. doi:10.1063/5.0012102.

[2] I. Nezbeda, F. Moučka, W. R. Smith, Recent progress in molecular simulation of aqueous electrolytes: force fields, chemical potentials and solubility, Molecular Physics 114 (11) (2016) 1665–1690. doi:10.1080/00268976.2016.1165296.

[3] Z. Mester, A. Z. Panagiotopoulos, Temperature-dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations, The Journal of Chemical Physics 143 (4) (2015) 044505. doi:10.1063/1.4926840.

[4] A. Benavides, M. Portillo, J. Abascal, C. Vega, Estimating the solubility of 1:1 electrolyte aqueous solutions: the chemical potential difference rule, Molecular Physics 115 (9-12) (2017) 1301–1308. doi:10.1080/00268976.2017.1288939.

[5] F. Moučka, I. Nezbeda, W. R. Smith, Chemical potentials, activity coefficients, and solubility in aqueous NaCl solutions: Prediction by polarizable force fields, Journal of Chemical Theory and Computation 11 (4) (2015) 1756–1764. doi:10.1021/acs.jctc.5b00018.

[6] J. Kolafa, Solubility of NaCl in water and its melting point by molecular dynamics in the slab geometry and a new BK3-compatible force field, The Journal of Chemical Physics 145 (20) (2016) 204509. doi:10.1063/1.4968045.

[7] J. Dočkal, M. Lísal, F. Moučka, Molecular force field development for aqueous electrolytes: 2. polarizable models incorporating crystalline chemical potential and their accurate simulations of halite, hydrohalite, aqueous solutions of NaCl, and solubility, Journal of Chemical Theory and Computation 16 (6) (2020) 3677–3688. doi:10.1021/acs.jctc.0c00161.

[8] E. Sanz, C. Vega, Solubility of KF and NaCl in water by molecular simulation, The Journal of Chemical Physics 126 (1) (2007) 014507. doi:10.1063/1.2397683.

[9] A. L. Benavides, J. L. Aragones, C. Vega, Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route, The Journal of Chemical Physics 144 (12) (2016) 124504. doi:10.1063/1.4943780.

[10] Z. Mester, A. Z. Panagiotopoulos, Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations, The Journal of Chemical Physics 142 (4) (2015) 044507. doi:10.1063/1.4906320.

[11] S. H. Saravi, A. Z. Panagiotopoulos, Activity coefficients and solubilities of NaCl in water–methanol solutions from molecular dynamics simulations, The Journal of Physical Chemistry B 126 (15) (2022) 2891–2898. doi:10.1021/acs.jpcb.2c00813.

[12] F. Moučka, M. Lísal, J. Škvor, J. Jirsák, I. Nezbeda, W. R. Smith, Molecular simulation of aqueous electrolyte solubility. 2. osmotic ensemble Monte Carlo methodology for free energy and solubility calculations and application to NaCl, The Journal of Physical Chemistry B 115 (24) (2011) 7849–7861. doi:10.1021/jp202054d.

[13] F. Moučka, I. Nezbeda, W. R. Smith, Molecular simulation of aqueous electrolytes: Water chemical potential results and Gibbs-Duhem equation consistency tests, The Journal of Chemical Physics 139 (12) (2013) 124505. doi:10.1063/1.4821153.

[14] F. Moučka, J. Kolafa, M. Lísal, W. R. Smith, Chemical potentials of alkaline earth metal halide aqueous electrolytes and solubility of their hydrates by molecular sim-

ulation: Application to $CaCl_2$, antarcticite, and sinjarite, The Journal of Chemical Physics 148 (22) (2018) 222832. doi:10.1063/1.5024212.

[15] A. S. Paluch, S. Jayaraman, J. K. Shah, E. J. Maginn, A method for computing the solubility limit of solids: Application to sodium chloride in water and alcohols, The Journal of Chemical Physics 133 (12) (2010) 124504. doi:10.1063/1.3478539.

[16] L. Li, T. Totton, D. Frenkel, Computational methodology for solubility prediction: Application to sparingly soluble organic/inorganic materials, The Journal of Chemical Physics 149 (5) (2018) 054102. doi:10.1063/1.5040366.

[17] J. L. Aragones, E. Sanz, C. Vega, Solubility of NaCl in water by molecular simulation revisited, The Journal of Chemical Physics 136 (24) (2012) 244508. doi:10.1063/1.4728163.

[18] J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti, A. Z. Panagiotopoulos, On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures, The Journal of Chemical Physics 145 (15) (2016) 154111. doi:10.1063/1.4964725.

[19] I. S. Joung, T. E. Cheatham, Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters, The Journal of Physical Chemistry B 113 (40) (2009) 13279–13290. doi:10.1021/jp902584c.

[20] K. Kobayashi, Y. Liang, T. Sakka, T. Matsuoka, Molecular dynamics study of salt–solution interface: Solubility and surface charge of salt in water, The Journal of Chemical Physics 140 (14) (2014) 144705. doi:10.1063/1.4870417.

[21] H. M. Manzanilla-Granados, H. Saint-Martín, R. Fuentes-Azcatl, J. Alejandre, Direct coexistence methods to determine the solubility of salts in water from numerical simulations. Test case NaCl, The Journal of Physical Chemistry B 119 (26) (2015) 8389–8396. doi:10.1021/acs.jpcb.5b00740.

[22] T. Yagasaki, M. Matsumoto, H. Tanaka, Lennard-Jones parameters determined to reproduce the solubility of NaCl and KCl in SPC/e, TIP3p, and TIP4p/2005 water, Journal of Chemical Theory and Computation 16 (4) (2020) 2460–2473. doi:10.1021/acs.jctc.9b00941.

[23] D. Frenkel, A. J. C. Ladd, New Monte Carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres, The Journal of Chemical Physics 81 (7) (1984) 3188–3193. doi:10.1063/1.448024.

[24] I. S. Joung, T. E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, The Journal of Physical Chemistry B 112 (30) (2008) 9020–9041. doi:10.1021/jp8001614.

[25] D. E. Smith, L. X. Dang, Computer simulations of NaCl association in polarizable water, The Journal of Chemical Physics 100 (5) (1994) 3757–3766. doi:10.1063/1.466363.

[26] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials, The Journal of Physical Chemistry 91 (24) (1987) 6269–6271. doi:10.1021/j100308a038.

[27] P. R. ten Wolde, D. Frenkel, Enhancement of protein crystal nucleation by critical density fluctuations, Science 277 (5334) (1997) 1975–1978. doi:10.1126/science.277.5334.1975.

[28] V. J. Anderson, H. N. W. Lekkerkerker, Insights into phase transition kinetics from colloid science, Nature 416 (6883) (2002) 811–815. doi:10.1038/416811a.
[29] D. Erdemir, A. Y. Lee, A. S. Myerson, Nucleation of crystals from solution: Classical and two-step models, Accounts of Chemical Research 42 (5) (2009) 621–629. doi:10.1021/ar800217x.
[30] N. E. R. Zimmermann, B. Vorselaars, D. Quigley, B. Peters, Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates, Journal of the American Chemical Society 137 (41) (2015) 13352–13361. doi:10.1021/jacs.5b08098.
[31] G. Lanaro, G. N. Patey, Birth of NaCl crystals: Insights from molecular simulations, The Journal of Physical Chemistry B 120 (34) (2016) 9076–9087. doi:10.1021/acs.jpcb.6b05291.
[32] H. Jiang, A. Haji-Akbari, P. G. Debenedetti, A. Z. Panagiotopoulos, Forward flux sampling calculation of homogeneous nucleation rates from aqueous NaCl solutions, The Journal of Chemical Physics 148 (4) (2018) 044505. doi:10.1063/1.5016554.
[33] H. Jiang, P. G. Debenedetti, A. Z. Panagiotopoulos, Nucleation in aqueous NaCl solutions shifts from 1-step to 2-step mechanism on crossing the spinodal, The Journal of Chemical Physics 150 (12) (2019) 124502. doi:10.1063/1.5084248.
[34] D. James, S. Beairsto, C. Hartt, O. Zavalov, I. Saika-Voivod, R. K. Bowles, P. H. Poole, Phase transitions in fluctuations and their role in two-step nucleation, The Journal of Chemical Physics 150 (7) (2019) 074501. doi:10.1063/1.5057429.
[35] N. E. R. Zimmermann, B. Vorselaars, J. R. Espinosa, D. Quigley, W. R. Smith, E. Sanz, C. Vega, B. Peters, NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates, The Journal of Chemical Physics 148 (22) (2018) 222838. doi:10.1063/1.5024009.
[36] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO – the open visualization tool, Modelling and Simulation in Materials Science and Engineering 18 (1) (2010) 015012. doi:10.1088/0965-0393/18/1/015012.
[37] H. A. Lorentz, Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase, Annalen der Physik 248 (1) (1881) 127–136. doi:10.1002/andp.18812480110.
[38] D. Berthelot, Sur le mélange des gaz, Compt. Rendus 126 (1898) 1703–1706.
[39] S. Deublein, J. Vrabec, H. Hasse, A set of molecular models for alkali and halide ions in aqueous solution, The Journal of Chemical Physics 136 (8) (2012) 084501. doi:10.1063/1.3687238.
[40] S. Reiser, S. Deublein, J. Vrabec, H. Hasse, Molecular dispersion energy parameters for alkali and halide ions in aqueous solution, The Journal of Chemical Physics 140 (4) (2014) 044504. doi:10.1063/1.4858392.
[41] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, T. Head-Gordon, Development of an improved four-site water model for biomolecular simulations: Tip4p-ew, Journal of Chemical Physics 120 (20) (2004) 9665 – 9678. doi:10.1063/1.1683075.
[42] R. Fingerhut, G. Guevara-Carrion, I. Nitzke, D. Saric, J. Marx, K. Langenbach, S. Prokopev, D. Celný, M. Bernreuther, S. Stephan, M. Kohns, H. Hasse, J. Vrabec, ms2: A molecular simulation tool for thermodynamic properties, release 4.0, Computer Physics Communications 262 (2021) 107860. doi:10.1016/j.cpc.2021.107860.
[43] M. P. Allen, D. J. Tildesley, Computer simulation of liquids, Oxford university press,

2017. doi:10.1093/oso/9780198803195.001.0001.
[44] C. W. Gear, The numerical integration of ordinary differential equations of various orders, Tech. rep., Argonne National Lab., Ill. (1966).
[45] P. P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, Annalen der Physik 369 (3) (1921) 253–287. doi:10.1002/andp.19213690304.
[46] H. C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, The Journal of Chemical Physics 72 (4) (1980) 2384–2393. doi:10.1063/1.439486.
[47] D. Saric, M. Kohns, J. Vrabec, Dielectric constant and density of aqueous alkali halide solutions by molecular dynamics: A force field assessment, The Journal of Chemical Physics 152 (16) (2020) 164502. doi:10.1063/1.5144991.
[48] T. R. Briggs, W. F. Geigle, Note on the system sodium iodide–water., The Journal of Physical Chemistry 44 (3) (1940) 373–377. doi:10.1021/j150399a009.
[49] R. Cohen-Adad, J. W. Lorimer, Alkali Metal and Ammonium Chlorides in Water and Heavy Water (Binary Systems) (IUPAC Solubility Data Series), Pergamon, 1991.
[50] S. Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory 28 (2) (1982) 129–137. doi:10.1109/tit.1982.1056489.
[51] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD) 96 (34) (1996) 226–231.