Modeling the formation and growth of atmospheric molecular clusters : A review

Elm, Jonas

2020-11

Elm , J , Kubecka , J , Besel , V , Jääskeläinen , M J , Halonen , R , Kurten , T & Vehkamäki , H 2020 , ' Modeling the formation and growth of atmospheric molecular clusters : A review ' , Journal of Aerosol Science , vol. 149 , 105621 . https://doi.org/10.1016/j.jaerosci.2020.105621

http://hdl.handle.net/10138/346045
https://doi.org/10.1016/j.jaerosci.2020.105621

Modelling the Formation and Growth of Atmospheric Molecular Clusters: A Review

Jonas Elm[a], Jakub Kubečka[b], Vitus Besel[b], Matias J. Jääskeläinen[b], Roope Halonen[b], Theo Kurtén[c] and Hanna Vehkamäki[b]

[a]Department of Chemistry and iClimate, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Aarhus, Denmark
[b]Institute for Atmospheric and Earth System Research, Faculty of Science, University of Helsinki, 00014 Helsinki, Finland
[c]Department of Chemistry, Faculty of Science, University of Helsinki, 00014 Helsinki, Finland

ARTICLE INFO

Keywords:
clusters
new particle formation
quantum chemistry
sampling

ABSTRACT

Molecular clusters are ubiquitous constituents of the ambient atmosphere, that can grow into larger sizes forming new aerosol particles. The formation and growth of small clusters into aerosol particles remain one of the largest uncertainties in global climate predictions. This has made the modelling of atmospheric molecular clustering into an active field of research, yielding direct molecular level information about the formation mechanism. We review the present state-of-the-art quantum chemical methods and cluster distribution dynamics models that are applied to study the formation and growth of atmospheric molecular clusters. We outline the current challenges in applying theoretical methods and the future directions to move the field forward.

Contents

*Corresponding author
jelm@chem.au.dk (J. Elm)
ORCID(s):

1. Introduction

Atmospheric vapour molecules, with strong intermolecular interactions, can accumulate in the gas phase forming molecular clusters (Kulmala et al., 2013). Provided that the interactions between the molecules are strong enough, these molecular clusters can be stable against evaporation. Stable clusters can further growth into larger sizes by uptake of additional vapour molecules and eventually form aerosol particles (here defined as 2 nm in diameter or above). If not scavenged by other sources, such as coagulation on pre-existing particles, the particles can act as cloud condensation nuclei (CCN) when they reach diameters of ∼50-100 nm. The interactions between aerosol particles and clouds remain the least understood process in global climate estimation (IPCC), and presently, up to half the number of CCN is believed to originate from new particles formed in the atmosphere from gas-phase molecules (Merikanto et al., 2009). Even minor changes in the parameterization of the early growth of particles between 1.7 nm and 3.0 nm can lead to up to a 50% increase in modelled cloud condensation nuclei (Tröstl et al., 2016). The formation of ultrafine particles presents a large problem for air quality and our health. This is a growing concern, especially in densely populated areas, such as megacities in China, where visible air pollution has become a daily plague. Long term exposure to ultrafine aerosol particles can lead to inflammation in the lungs, potentially resulting in lung cancer and cardiovascular diseases (Gan et al., 2013; Falcon-Rodriguez et al., 2017; Mei et al., 2018). The world health organisation (WHO) has estimated that ∼7 million premature deaths are annually linked to air pollution (WHO). Thus a more detailed and comprehensive understanding of the initial steps in forming new particles and their early growth is essential for improving our understanding of the global climate system and to better control the adverse health effect of ultrafine aerosol particles.

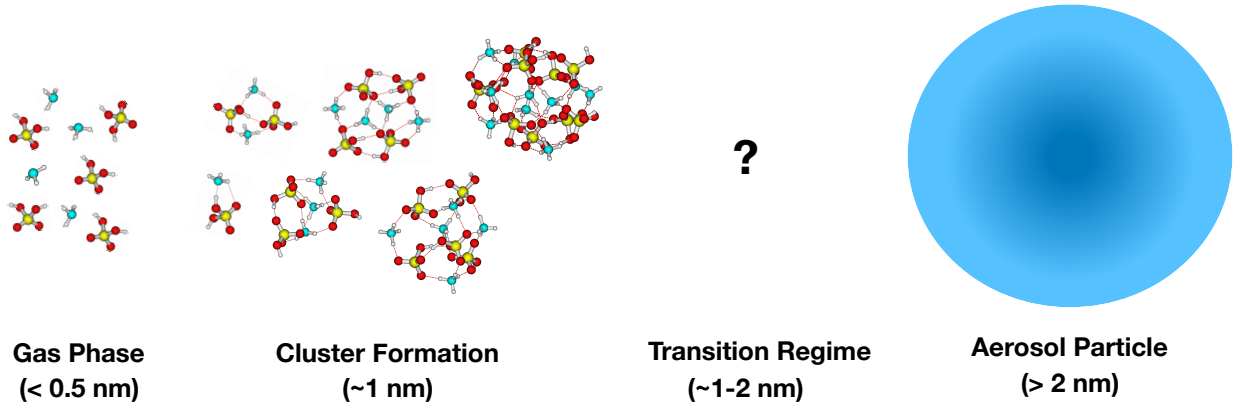

Figure 1: Schematic overview of the size range of atmospheric clusters and aerosol particles exemplified with sulfuric acid and ammonia. The critical process for cluster stabilization and growth (∼1.5 nm) is believed to occur within the transition regime between clusters and particles.

In recent years, the scientific interest in atmospheric molecular clusters has steadily increased and it presents a relatively new and rapidly expanding field of research. While many experimental techniques have been developed to study the formation and growth of atmospheric clusters into large aerosol particles, there exist no single method that can effectively uncover the process from the smallest clusters up to large particles. For instance, there is a significant gap between the largest sizes that can be measured using chemical ionization atmospheric pressure interface mass spectrometer (up to ∼1.5 nm in diameter) and the smallest size that can be measured by aerosol mass spectrometers (from ∼50 nm). Theoretical simulations can be utilized to obtain detailed knowledge about the initial steps in cluster formation and growth pathways (∼0.5-1.5 nm), which remain difficult to assess using state-of-the-art experimental techniques. Recent advances in theoretical methods have allowed the modelling of atmospheric molecular clusters to bloom and permit studies of significantly larger clusters than previously possible. However, there is still a persisting gap between the size of the currently modelled clusters and the sizes which are usually measured in the ambient atmosphere. Modelling this transition regime between 1-2 nm is of utter importance as evidence suggests that this is the region where critical processes for cluster survival occur (see Figure 1). In this review, we discuss the progress, outline some of the current challenges and future prospects in modelling the formation and growth of atmospheric molecular clusters.

2. Characterization using Experimental Techniques

2.1. Detection of Clusters and Aerosol Particles

Detection of aerosol particles can be performed using Neutral cluster and Air Ion Spectrometers (NAIS) (Kulmala et al., 2007; Manninen et al., 2009) or condensation particle counters (CPC) (McMurry, 2000) with detection thresholds down to around 2-3 nm in diameter (McMurry, 2000). State-of-the-art measurement techniques such as particle size magnifiers (PSM) (Vanhanen et al., 2011) now allows the detection of particles as small as ∼1.7 nm, but these instruments have poor counting efficiency and are not standard instrumentation in the field (Kontkanen et al., 2017). However, particle counters yield no insight into the chemical composition of the detected particles or any mechanistic understanding on how they were formed. Using offline filter techniques chemical classification of particle constituents can be performed using, for instance, gas chromatography coupled with a mass spectrometer (GC-MS) (Noziére et al., 2015). This requires a large mass of particles to be collected on the filters, and potentially important trace compounds will not be detected. Online information about the composition of aerosol particles (in the ∼50 nm to ∼1 μm size range) can be obtained using an aerosol mass spectrometer (AMS) (Davis, 1973).

To obtain online information about the composition of atmospheric molecular clusters, a chemical ionization atmospheric pressure interface mass spectrometer (CI-APi-TOF) (Jokinen et al., 2012) can be utilized. This technique relies on charging the clusters, which might change their molecular composition and lead to fragmentation inside the instrument (Zapadinsky et al., 2019; Passananti et al., 2019). This implies that the measured cluster might not be representative of the cluster that was initially formed. For instance, most water molecules that are present in the cluster will

instantly evaporate during the detection process (Ehn et al., 2010; Schobesberger et al., 2013). Various reagent ions (nitrate (Ehn et al., 2012, 2014; Hyttinen et al., 2015), iodide (Lee et al., 2014; Iyer et al., 2016), acetate/lactate (Berndt et al., 2016; Jen et al., 2016b; Hyttinen et al., 2017) and bisulfate (Sipilä et al., 2015)) are sensitive towards different molecular systems and not all clusters may be efficiently detected by a single charging technique. An improved resolution compared to the CI-APi-TOF can be achieved by applying the recently developed CI-Orbitrap technique which shows promising applications in the detection of highly oxygenated organic molecules (HOMs) (Riva et al., 2019). Cryogenic techniques can be applied to hinder the evaporation process and coupling mass spectrometry with infrared spectroscopy allows the direct detection of charged cluster compositions and makes it possible to study arrangement of molecules in the cluster (Waller et al., 2018). Using tandem MS coupled to an ion trap, the influence of adding water to the cluster structures has been directly detected (Yang et al., 2018).

2.2. Potential Compounds Involved in Particle Formation

With state-of-the-art instruments some of the compounds that might potentially be involved in new particle formation have been revealed in recent years. Over the oceans, the primary source of CCN is believed to be dimethylsulfide (DMS), which is produced by phytoplankton and emitted into the atmosphere. The emitted DMS is then oxidized by OH/Cl radicals and the mechanism is believed to involve the formation of low volatile acids such as methanesulfonic acid and sulfuric acid which drive the formation of new particles (Charlson et al., 1987). In coastal regions, iodine emissions from biogenic sources have been shown to form marine aerosols (O'Dowd et al., 2002). The mechanism is believed to be linked to the formation of iodic acid which can produce particles via sequential addition of iodic acid molecules (Sipilä et al., 2016).

In continental regions, sulfuric acid and water have been definitively established as important components in particle formation (Sipilä et al., 2010). Nitrogen-containing compounds with high basicity such as ammonia (Kirkby et al., 2011; DePalma et al., 2012, 2014), monoamines (Kurtén et al., 2008; Almeida et al., 2013; Jen et al., 2014; Glasoe et al., 2015), diamines (Jen et al., 2016a; Elm et al., 2016a, 2017d) and guanidine (Myllys et al., 2018, 2019b; Hemmilä et al., 2020) have been shown to efficiently enhance sulfuric acid based new particle formation provided that they are present at even ppt-level mixing ratios. Carboxylic acids have also been claimed to enhance sulfuric acid based new particle formation (Zhang et al., 2004). Highly oxygenated organic molecules termed HOMs (Bianchi et al., 2019) have also been speculated to participate in sulfuric acid based new particle formation (Schobesberger et al., 2013; Riccobono et al., 2014), but most likely ionic pathways are required to facilitate the process (Kirkby et al., 2016; Rose et al., 2018). An overview of potentially important species for cluster formation and growth are shown in Figure 2.

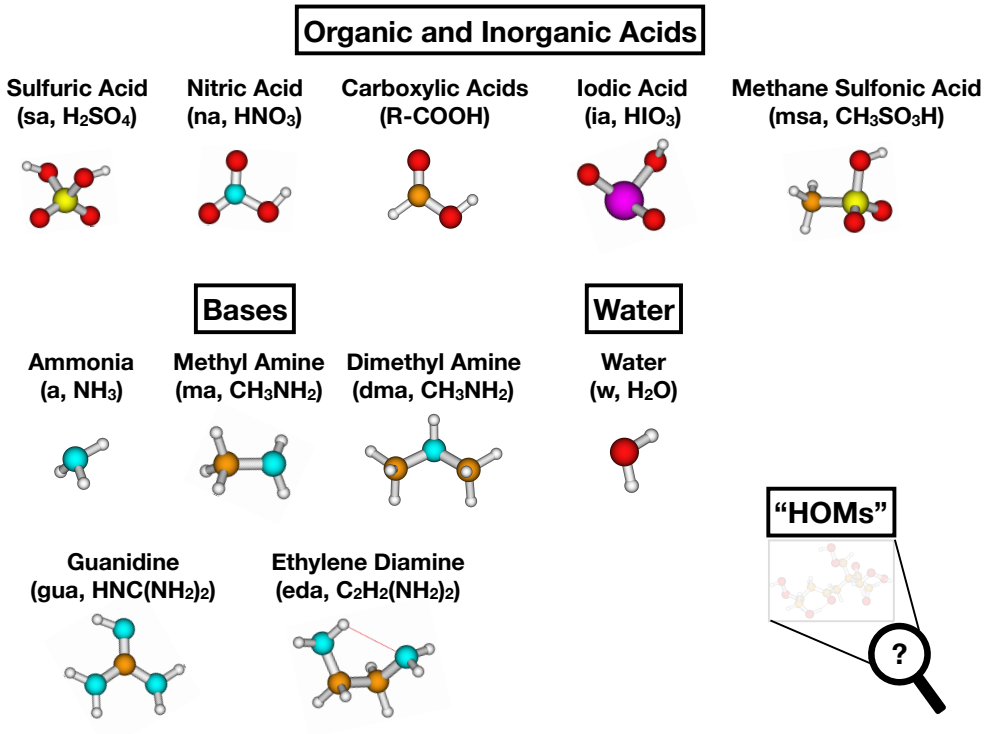

Figure 2: Potentially important chemical species involved in cluster formation and growth. The structure of potentially important highly oxygenated molecules (HOMs) remains unknown.

Ionic pathways greatly facilitate new particle formation rates when the vapour concentrations are low (Wagner et al., 2017). In general, the ion production rate in the atmosphere is too low for the ionic pathways to explain new particle formation bursts. Thus, particle formation seems to be mainly driven by neutral pathways. The overall particle formation mechanism is generally a matter of the correct timing of relevant chemical species colliding under the right atmospheric conditions (Bianchi et al., 2016). Multi-component sulfuric acid - ammonia - HOM clusters were studied in the CLOUD chamber. It was found that the components had a synergistic enhancing effect on particle formation and that covalently bound HOM dimers most likely contributed more to process compared to the monomers (Lehtipalo et al., 2018).

The total concentration of molecules in ground-level atmospheric conditions (pressure around 1 atm and temperatures on the order of 250-300 K) is on the order of 10^{19} molecules cm^{-3}. At high relative humidities, the mixing ratio of water may be around 1%, corresponding to $\sim 10^{17}$ molecules cm^{-3}. The concentration of any other condensable vapour is always many orders of magnitude lower. For example, ammonia mixing ratios can reach tens of ppb, *i.e.* 10^{11} molecules cm^{-3}, while amine concentrations are typically substantially lower. Even in very polluted environments, the most studied driver of atmospheric new-particle formation, sulfuric acid, seldom exceeds 10^{8} molecules cm^{-3}. The very reason that the key components have low gas-phase concentrations is that they efficiently transform into the condensed phase. Hence, attempting to identify atmospheric trace species in the gas phase that might be important for new particle formation presents the common conundrum of finding "a needle in a haystack".

3. Atmospheric Particle Formation

3.1. Principles and Definitions

There are multiple different ways of classifying particles in the atmosphere. For our purposes, the most important distinction is between primary and secondary particles. Primary particles, such as desert or road dust, soot, pollen etc., have been emitted into the air as pre-existing particles, though their size and other properties may subsequently change due to processing in the atmosphere. In contrast, secondary particles have been completely formed in the air by the

clustering of gas-phase molecules. Especially within aerosol physics, it is customary to separate this initial clustering process ("particle formation") from the subsequent condensation of gas molecules onto both primary and secondary particles ("particle growth"). The precise threshold between "clusters" and "particles" is ambiguous, and varies with the property of interest. Other frequently used classification categories include:

The particle origin, *e.g.* biogenic/anthropogenic: Note that this distinction is often difficult to apply, for example in the case of human-induced forest fires, or particles composed of organic nitrates formed in reactions of biogenic carbon compounds with anthropogenic nitrogen oxides.

Composition, *e.g.* organic/inorganic: In principle this classification is straightforward to apply, as "organic" can be defined as any compound containing C-H bonds. In practice, aminium bisulfate or aminium sulfate salts are often implicitly included in the "inorganic" category even though the amines contain carbon atoms, and "organic" is reserved exclusively for non-basic volatile organic compound (VOC) oxidation products.

Phase state, *e.g.* solid/liquid: The phase state of larger particles is an area of intense research, see (Virtanen et al., 2010; Shiraiwa et al., 2011; Pöschl and Shiraiwa, 2015; Shiraiwa et al., 2017; Reid et al., 2018) and will not be discussed here. For the smallest clusters, the distinction is unclear, and possibly irrelevant. It should be noted that in typical applications of quantum chemistry to atmospheric clusters, it is simultaneously assumed that clusters are "crystal-like" in the sense that their partition functions can be computed using rotational and vibrational energy levels corresponding to one or more minimum-energy structures (the global minimum and possibly a limited number of low-lying local minima), but "liquid-like" in the sense that interconversions between the different minima are rapid, compared to the timescale of evaporation reactions. This is based on the assumptions that there are no significant free energy barriers between the different minima. If the latter condition does not hold, then the assumption that clusters quickly re-organize to the global minimum structure after each collision would not be valid. Under the assumption that the rapid interconversion between minima holds, the probability that a given cluster is found in some specific configuration (k) is proportional to $\exp(-\Delta G_{\mathrm{bind},k}/\mathrm{RT})$, where $\Delta G_{\mathrm{bind},k}$ is the binding energy of the configuration (see section 4.4), R is the molar gas constant and T is the temperature.

Particle formation is often loosely called nucleation. Strictly speaking, if the evaporation probability of even a dimer containing only two molecules is lower than the probability that the clusters grow, particle formation is kinetically controlled, and should not be called nucleation, and classical nucleation theory (equation (1)) cannot be applied. If clusters have to grow up to a threshold size (the critical cluster size) before the growth probability exceeds the evaporation probability; the process is genuine nucleation. Since the evaporation rate does not depend on the vapour concentrations (it is an inherent property of the cluster), while the growth rate is directly proportional to the vapour concentration, the very same particle formation mechanism may or may not involve nucleation, depending on the vapour concentration. In the case of nucleation, the formation free energy, related to the binding free energy and definition in section 4.5, exhibits a barrier. Figure 3 sketches three different types of commonly encountered formation free energy surfaces.

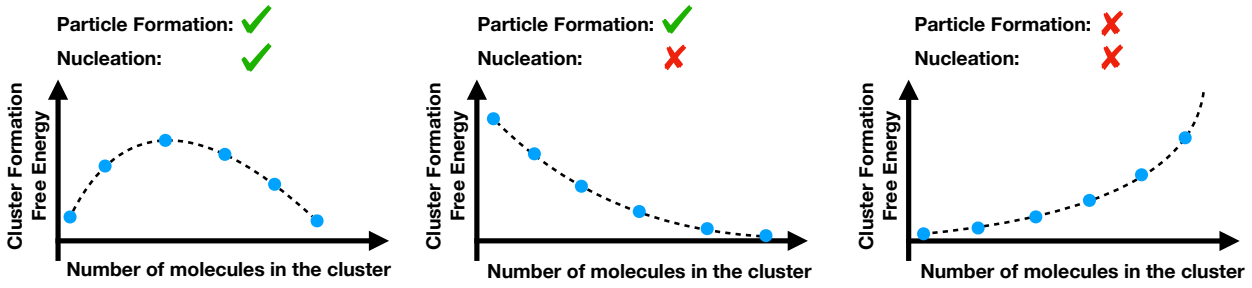

Figure 3: Sketch of different types of cluster formation free energy surfaces and whether they involve nucleation/particle formation.

Another important distinction to make is that between (thermal) evaporation, and various types of scattering or non-accommodation. While both refer to the process of a molecule escaping from a cluster, the timescales of the processes are very different at least for condensing vapors and conditions relevant to the lower atmosphere. Cluster-molecule collisions occur on timescales of picoseconds (the time it takes a molecule traveling at typical thermal velocities of hundreds of meters per second to cross a typical molecular/cluster with a diameter of a nanometer or less). Collisions may then lead to immediate scattering (on the timescale of picoseconds), or to the uptake of the molecule onto the cluster surface for at least one vibrational period (typically also on the order of picoseconds for intermolecular vibrations). For example, any excess energy from the collision event will be removed by thermalizing collisions (with N_2 and O_2 gas molecules) occurring at a rate of around 10^{10} s^{-1} at atmospheric pressure. Thus, loss due to energy non-accommodation occurs within a few nanoseconds. The net growth probability (or growth rate) can be defined as the product of the collision probability (or rate) and the uptake or accommodation coefficient. This coefficient is often assumed to be unity, and despite considerable research, no universal consensus has emerged as to the correct value of the coefficient even for simple systems such as pure H_2O surfaces. In contrast to collisions and non-accomodation processes, evaporation of fully accommodated molecules occurs at substantially longer timescales; for example, for sulfuric acid - base clusters relevant to new-particle formation, the evaporation rates at 298 K are usually well below 10^6 s^{-1} (Olenius et al., 2013; Elm, 2017). For example, for a typical H_2SO_4 concentration of around 10^7 molecules cm^{-3} and a molecule-cluster collision rate around cm^3 $molecule^{-1}$ s^{-1}, the threshold for "stability" is an H_2SO_4 evaporation rate below 10^{-3} s^{-1}.

To simplify the discussion above, we have referred exclusively to the collision and evaporation of single molecules. However, for sufficiently strongly bonded multicomponent systems (*e.g.* sulfuric acid and many amines), the concentration of small clusters may be comparable to that of the free monomers, and cluster-cluster collisions, as well as the reverse process of cluster fragmentation to two smaller clusters, must be taken into account. Even in the much weaker bound H_2SO_4-H_2O system, the very large atmospheric concentration of H_2O leads to a large fraction of the H_2SO_4 monomers existing as hydrates, and this must be accounted for even though most other cluster-cluster collisions or fragmentation reactions can be ignored.

3.2. Birth-death Equations

The time evolution of the cluster distribution is governed by a set of birth-death equations:

$$\frac{dc_i}{dt} = \sum_{j<i} \beta_{j,(i-j)} c_j c_{(i-j)} - \sum_j \beta_{i,j} c_i c_j + \sum_j \gamma_{(i+j),j} c_{(i+j)} - \sum_{j<i} \gamma_{i,j} c_i + \text{sources} + \text{sinks} \tag{1}$$

where c_i is the concentration of clusters of type i. In one-component systems, i is simply the number of molecules in the cluster, and in multicomponent systems, it is a list of numbers of molecules of each type in the cluster. $\beta_{i,j}$ is the growth rate (collision rate multiplied by the accommodation coefficient) resulting from collisions of clusters i and j, and $\gamma_{(i,j)}$ is the thermal evaporation/fragmentation rate of the cluster i yielding clusters j and $i - j$. The first sum term represents the birth of clusters of type i by collisions of smaller clusters, the second sum term the growth of clusters i to larger sizes. The third sum term represents production of i clusters by evaporation/fragmentation of larger cluster and the last sum term the removal of i clusters as they evaporate/fragment. Sources are often active only for a few cluster types, for example, monomers produced by chemical reactions. Sinks include for example attachment of molecules and clusters to larger particles and experimental chamber walls and usually affect all cluster sizes. In the case when only collisions where at least one of the parties is a monomer and only evaporation of monomers are allowed, they reduce to Becker-Döring equations (Becker and Döring, 1935) also known as Szilard-Farkas equations (Farkas, 1927) and if evaporation/fragmentation is ignored, they reduce to Smoluchowski coagulation equations (Smoluchowski, 1916). Figure 4 shows a simplified schematic representation of the time evolution of a cluster population.

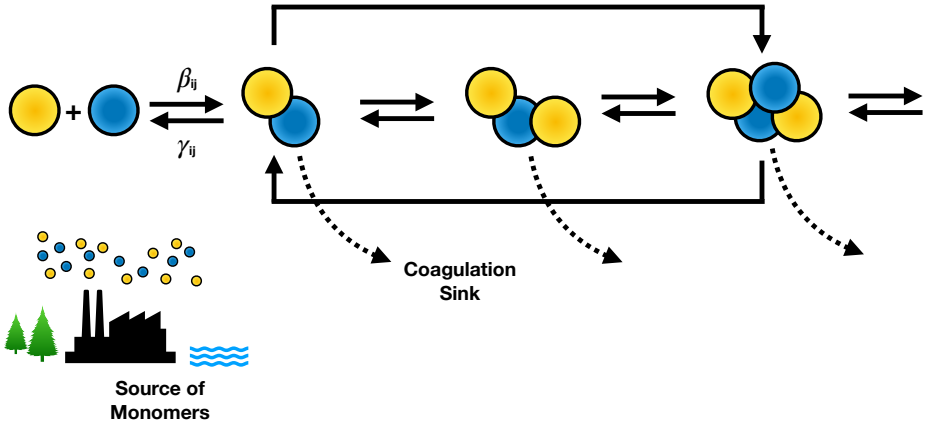

Figure 4: Simplified schematic representation of the birth-death equations.

The birth-death equations themselves contain only a few implicit assumptions about the clustering system. First, a single set of birth-death equations describes clustering at one point in space. If the concentrations, or conditions in general, vary strongly in space, the equations should be solved separately for different regions. Second, the use of thermal evaporation rates assumes that the system, including the clusters, has a well-defined temperature, which is not affected by the clustering process. This implies first of all that the clusters are only a small part of the total system. This certainly holds in the atmosphere, as the vast majority of gas molecules are N_2 and O_2. Second, it also implies that the nascent clusters formed by collisions have a sufficient number of internal degrees of freedom, permitting them to live long enough to be thermalized by collisions with a carrier gas (*e.g.* N_2 and O_2). Due to the relatively large number of atoms in the key molecules likely responsible for atmospheric particle formation, the assumption of thermalization at pressures encountered in the lower atmosphere (0.1 to 1 atm, or roughly 10 000 to 100 000 Pa) is unlikely to be a major source of error. The probability of fragmentation or evaporation due to energy non-accommodation (*i.e.* imperfect thermalization) decreases rapidly with the number of vibrational modes in the cluster, which is equal to $3N - 6$, where N is the number of atoms in the cluster. A common misconception, possibly originating from simulations of clustering of atoms, *e.g.* metals or noble gases, is that N would equal the number of molecules. For example, in a sulfuric acid dimer ($N = 14$), there are already 36 vibrational modes. Even if only a fraction of these are accessible for accommodating excess energy, thermalization at 1 atm is likely to be essentially complete - especially after accounting

for the almost inevitable presence of one or more H_2O molecules, which provide both additional vibrational modes, and a "sacrificial cooling" - type mechanism in which water molecules are lost instead of H_2SO_4 (Kurtén et al., 2010). It should be mentioned that if the colliding molecules are more complex, the amount of excess energy released after collision is also increased which would promote non-thermal evaporation. We note that the assumption of complete thermalization also implies a lack of dependence of the clustering process on the total pressure (see section 4.5 for a discussion of the reference pressure, which often causes confusion). While the assumption of thermalization is likely valid throughout the lower atmosphere, there are situations in which energy non-accommodation and non-thermal fragmentation need to be accounted for, such as inside mass of spectrometer instruments (Hogan. Jr and Fernandez. de. la. Mora, 2010; Thomas et al., 2016; Zapadinsky et al., 2019). This can conceptually be thought of as adding an energy axis to the birth-death equations, with especially fragmentation rates being summed over energies in addition to daughter cluster types.

There are two main approaches for solving the birth-death equations. The historical approach has been to use the liquid-drop model based on classical bulk thermodynamics to predict cluster properties, and a further set of simplifying assumptions allowing the birth-death equations to be solved analytically. This approach is known as ***classical nucleation theory (CNT)***. More recently, computational chemistry tools have permitted the calculation of cluster properties from molecular interactions, and these can be combined with numerical solutions of the birth-death equations for a limited set of cluster sizes. We refer to the latter approach as ***cluster distribution dynamics***.

Regardless of the approach chosen, one additional approximation is almost inevitably done when solving the birth-death equations: evaporation rates ($\gamma_{(i,j)}$) are never calculated or simulated directly for atmospherically relevant systems, but instead indirectly deduced using statistical mechanics tools. The reason for this is simple: the timescales for the key evaporation rates of the relevant molecules are far too long for direct simulation of the dynamic process, even with the cheapest possible computational models (classical force fields). By far the most common approach is to first calculate the collision rates for example from kinetic gas theory, and then derive evaporation rates by applying detailed balance to an equilibrium situation. The use of kinetic gas theory assumes that the clusters are in the free molecular regime, as their size is much smaller than their mean free path. For electrically neutral molecules, collision rates can then be calculated using simple analytic formula, with the following assumptions:

1. Clusters and molecules are spherical, with radii deduced from their bulk densities.
2. There are no long-range interactions between the molecules.
3. The velocities follows the Maxwell-Bolzmann distribution (this is inherent in the thermalization assumption).

For ion-molecule and anion-cation collisions, long-range interactions cannot be ignored, and at least in cluster distribution dynamics these collision rates are typically calculated based on semi-empirical parametrizations (Su and Bowers, 1973; Su and Chesnavich, 1982). Similar approaches could and should also be applied in CNT simulations involving charged clusters. Note that anion-anion and cation-cation collisions do not occur because of electrostatic repulsion.

The key computational concept involved in deriving evaporation rates from collision rates within both CNT and cluster distribution dynamics is the cluster binding free energy ΔG_{bind} or closely related ΔG_{form} (see sections 4.4 and 4.5): the free energy change in going from the isolated monomers to the cluster. In quantum chemistry based studies, this is usually specified as the Gibbs binding free energy, but we note that given the assumption that the clusters are only a vanishingly small part of the total system, the Gibbs free energy, Helmholz free energy and grand potential are actually identical (Vehkamäki, 2006).

Once the collision rate $\beta_{i,j}$ is known, the evaporation rate $\gamma_{(i+j),j}$ can be obtained through detailed balance, *i.e.* by assuming that in equilibrium, the flow of clusters from one type to another due to a single growth process is exactly cancelled by the corresponding evaporation process. Another way to formulate this is simply to note that the ratio of the forward and backward rates of any reaction must equal the equilibrium constant. Expressing the equilibrium cluster distribution (or the equilibrium constant) in terms of the cluster binding free energies (see section 4.5), results in the following form for the evaporation rate:

$$\gamma_{(i+j),j} = \frac{p_{\text{ref}}\beta_{i,j}}{k_{\text{B}}T} \exp\left(\frac{\Delta G_{\text{add},(i+j)}}{RT}\right). \tag{2}$$

This is under the assumption that there are no rearrangement barriers. Here $\Delta G_{\text{add},(i+j)}$ is the free energy of the reaction, where cluster j is added to cluster i and it can be expressed in terms of the cluster binding free energies as

$\Delta G_{\mathrm{bind},(i+j)} - \Delta G_{\mathrm{bind},j} - \Delta G_{\mathrm{bind},i}$. p_{ref} is the reference pressure at which the free energies are calculated. Its numerical value does not affect the evaporation rates, as explained in section 4.5.

3.3. Classical Nucleation Theory

In addition to the common approximations described above, CNT makes the following assumptions:

1. All cluster concentrations are constant in time.
2. There are no sink terms, and the source term is included only implicitly by keeping the monomer concentrations constant.
3. Only monomer-cluster collisions and monomer evaporations are considered. Also, the growth rate (in practice, monomer-cluster collision rate) of the cluster is approximated to be independent of the cluster size or type, for clusters close in size to the critical cluster.
4. Cluster formation free energies can be calculated by the liquid drop model, based on pure-component saturation vapor pressures, bulk liquid activity coefficients (describing the intermolecular interactions in multicomponent systems), flat-surface surface tensions, and bulk liquid densities. This takes into account the size dependence of the evaporation rate using the Kelvin equation: small clusters evaporate faster.
5. There exists one dominant cluster growth pathway (*i.e.* the lowest free energy path from monomers to particles), and this pathway contains precisely one free energy barrier, which is relatively high. In other words, the barriers along any other pathways are assumed to be much higher. We note that the mathematical approximations required by multicomponent CNT rely on the cluster composition (ratio of different constituent monomers to each other) remaining roughly similar along the dominant growth pathway.

After these approximations, the multicomponent CNT nucleation rate can be expressed as:

$$J = R_{\mathrm{avg}} c_{\mathrm{all\ clusters}} Z \exp\left(-\frac{\Delta G^*_{\mathrm{form,\ liquid\ drop}}}{RT}\right). \tag{3}$$

Where R_{avg} is the monomeric growth rate averaged over the different components, $c_{\mathrm{all\ clusters}}$ is a normalization coefficient for the cluster distribution (usually taken to be equal to the sum of the monomer concentrations, assuming that the cluster concentrations are negligible compared to monomer concentrations). $\Delta G^*_{\mathrm{form,\ liquid\ drop}}$ is the formation free energy of the critical cluster calculated using the liquid drop model corresponding to the highest free energy along the minimum free energy path from monomers to particles. In the capillarity approximation the free energy barrier is given by:

$$\Delta G_{\mathrm{form,\ liquid\ drop}} = \frac{4}{3}\pi\sigma r^{*2} \tag{4}$$

where r^* is the radius of the critical cluster:

$$r^* = \frac{2\sigma v_n}{k_{\mathrm{B}} T \ln\left(p_{1,n}/p_{\mathrm{equil},n}\right)} \tag{5}$$

Here σ is the surface tension of the liquid, $p_{1,n}$ is the partial pressure of monomer of type n, and $p_{\mathrm{equil},n}$ is the equilibrium vapour pressure of component n over a flat surface of liquid which has the same composition as the critical cluster. v_n is the partial molecular volume of component n in the liquid. The mole fractions of the critical cluster solved by requiring that $r\ *$ in equation (5) has the same value independent of which component n is used on the right hand side of this equation. In semi-classical fashion, the formation free energy used in equation (3) can be derived from the quantum chemical calculations (see section 4.5). The Zeldovich factor Z accounts for two separate things: the concentration of the critical clusters in the nucleating supersaturated vapour differs (by a factor of roughly 0.5) from that in a hypothetical supersaturated equilibrium vapour, and some clusters that have crossed the free energy barrier evaporate back to pre-critical sizes. Readers familiar with chemical kinetics will recognize many analogies between the CNT expression and various formulations of transition state theory. The discussion above refers to classical nucleation theory for neutral systems. In applications of CNT containing ions, the contribution to the formation free energy from the ion-molecule interactions is treated using classical Thomson theory (Yue and Chan, 1979).

One of the central problems in using classical nucleation theory to predict atmospheric particle formation is the treatment of proton transfer. Proton transfer reactions are essential for forming stable clusters, and the extent of proton

transfer depends not only on the relative amounts of, for example, acid and base molecules in the cluster, but also the actual numbers of molecules, in other words the cluster size. The liquid drop model assumes that clusters of all sizes exhibit the bulk value for the extent of proton transfer, which depends on composition, but not size. This introduces a significant error in the formation free energies of the smallest clusters. For example, for a cluster with one ammonia and one sulfuric acid molecule, the liquid drop model predicts a complete proton transfer, but the more realistic quantum chemical methods indicate that no proton transfer will occur. The liquid drop model thus overestimates the stability of these cluster drastically.

3.4. Cluster Distribution Dynamics

Cluster distribution dynamics are based on first obtaining collision (or growth) and evaporation rates for some predefined set of clusters (using for example free energies from quantum chemical calculations), and then numerically solving the birth-death equations up to some maximum cluster size (for example, 4 H_2SO_4 and 4 NH_3 molecules). Chemical intuition can often be used to reduce the number of clusters that need to be explicitly modelled; for example a $(NH_3)_4$ cluster will have a very high evaporation rate, and such clusters can thus be omitted from the set of simulated clusters (defined here as the set of clusters for which the birth-death equations are explicitly solved). An example of a typical acid - base simulation cluster set is shown in Figure 5 represented by 4×4 grid.

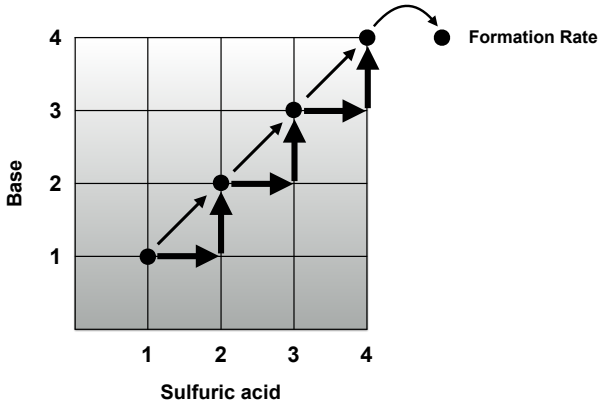

Figure 5: Schematic representation of a typical 4×4 simulation grid of sulfuric acid-base clusters. Clusters leaving the simulation set are defined as "new particles" and allowed to contribute to the formation rate $\boldsymbol{J}$. The size of the arrows correspond to the major/minor growth routes.

In addition to not relying on bulk thermodynamic models, the advantage of cluster distribution dynamic simulations is that they are not restricted by the rather stringent approximations inherent in CNT. For example, time-dependence, cluster-cluster collisions or cluster fragmentations to two smaller clusters can be trivially included. The most obvious error sources in cluster distribution dynamic simulations are the sink processes and the accuracy of the free energies used in determining the evaporation rates from equation (2). Section 5 discusses the accuracies and error sources related to the free energies. A less obvious, but equally important, potential source of error is related to the boundary conditions used in the simulation, including for example the limit of the set of simulated clusters, the treatment of clusters outside this set, and the choice of parameters kept constant (*e.g.* monomer concentration, total concentration of molecules of a certain type in the system, or monomer production rate). See the next section for a discussion of the importance of these settings.

Most applications of cluster distribution dynamics to date have used molecule-cluster and cluster-cluster collision rates directly from kinetic gas theory for neutral molecules and clusters, and standard parametrizations for collisions involving ions. However, some studies have applied “sticking factors” (equivalent to accommodation coefficients in the discussion above) below unity (Almeida et al., 2013; Lu et al., 2020), leading to reduced growth rates - and by the logic of detailed balance, also reduced evaporation rates. Such sticking factors may be motivated simply by improving the agreement with experimental results, or by qualitative physical arguments such as sulfuric acid - amine clusters needing to collide with each other in a particular orientation for sticking to occur as *e.g.* the non-polar alkyl groups of the amines

are unlikely to bind to each other sufficiently strongly. However, molecular dynamics simulations of collisions between sulfuric acid and dimethyl amine suggests that the intermolecular interactions between the molecules are strong enough to alleviate initial collision orientations that are not optimal, leading to a sticking factor of one (Loukonen et al., 2014). Very recently, molecular dynamics simulations have also allowed the direct simulations of collision rates accounting explicitly for both the non-spherical shape of molecules, as well as attractive dipole-dipole interactions. For the case of H_2SO_4-H_2SO_4 collisions, this has been found to result in an enhancement of the collision rate by about a factor of 2.2 (Halonen et al., 2019); even larger enhancements are likely for *e.g.* H_2SO_4-amine clusters as their dipole moment are larger. It is noteworthy that a similar enhancement factor has already been suggested for H_2SO_4-base clustering based on comparison to experimental studies (Kürten et al., 2014). We note that while these enhancements (or sticking) factors may seem large, they are still much smaller than the typical uncertainties in the evaporation rates calculated based on quantum chemical data due to the exponential dependence on the formation free energies in Equation (2).

3.5. Models and Common Misrepresentations

There exist a variety of different cluster distribution dynamics models with varying complexity. Early studies include the dynamics of pure water cluster (Schenter et al., 1999; Kathmann et al., 1999), the two component sulfuric acid - water system (Wyslouzil and Wilemski, 1995) and the three component sulfuric acid - ammonia - water system (Yu, 2006a). Yu (Yu, 2006b; Yu et al., 2018) further extended the two and three component models to also consider the effect of ionic species by considering the enhanced growth of charged clusters via dipole-charge interactions. The models by Yu also explicitly considers the formation of neutral species via ion-ion recombination.

Kürten *et al.* (Kürten et al., 2018; **?**) has developed models for comparison with data from the CLOUD experiments. This has led to a model of the CLOUD sulfuric acid - dimethylamine - water system, where the evaporation of the clusters where set to zero (Kürten et al., 2018). Kürten further extended the model to allow evaporation to simulate sulfuric acid - ammonia cluster formation with thermochemical parameters derived from CLOUD experiments (Kürten, 2019). Carlson and Zeuch (Carlsson and Zeuch, 2018) recently developed a gas phase and sectional particle dynamics model that allows the simulation of the shape and the modes of experimentally observed particle size distributions. Kulmala *et al.* (Kulmala, 2010) developed a Dynamical Atmospheric Cluster Model (DACM) where the birth-death equations were explicitly solved. McGrath *et al.* (McGrath et al., 2012) further improved this approach by introducing the Atmospheric Cluster Dynamics Code (ACDC). The ACDC method presents a flexible tool for studying cluster distribution dynamics, by generating the birth-death equations by analysis of all possible cluster combinations to find which evaporations and collisions that can create/destroy a given cluster.

Cluster distribution dynamics models present a powerful toolkit to understand particle formation pathways, which gives direct information on the formation mechanism and to interpret experimental results. However, caution with these models should be advised as they are very sensitive to the thermochemical input data. A common misconception is that the modelled results can be taken as the quantitative "true" atmospheric value. Discrepancies between model and experiment can result from:

a) The proposed mechanism is incorrect.
b) Inaccuracies in the model.
c) Inaccuracies in the experiment.
d) All of the above.

Another common issue leading to erroneous conclusions is that the choice of concentrations, temperature and losses is not reasonable for the environment being simulated. For example, quite often measurements of ground level concentrations are directly extrapolated to higher altitudes where the temperature is significantly lower. As the saturation vapour pressure is temperature dependent, it is not possible to have very high concentrations at low temperature. On a similar note, a very high concentration of vapours in the polluted atmosphere, will also be associated with a large loss of clusters due to condensation on pre-existing particles. These effects should be taken into consideration when setting up the dynamics model.

Another very important thing to consider is how to model the monomer concentration of a given species. Should the concentration be held constant or be allowed to change over the course of the simulation? A constant concentration corresponds either to continuous production of the monomer or a compound which is massively in excess, *e.g.*, its losses are negligible compared to the total concentration. Alternatively, the monomer concentration can be simulated using a fixed amount of formed monomers, which eventually will lead to depletion of the monomer in the gas phase

as the compound partitions to the cluster phase. The effect of monomer depletion was explicitly demonstrated by Liu *et al.* (Liu et al., 2019) who studied the chemical reaction between methanol and SO_3 forming the less-volatile methyl hydrogen sulfate compound. One would assume that the formation of a less volatile species would lead to an increased particle formation potential, but on the contrary, the formation of methyl hydrogen sulfate consumes an appreciable amount of atmospheric SO_3. This implies that less SO_3 is available to form sulfuric acid by reacting with water and thus limiting particle formation. The complete opposite trend was observed in the simulations if the concentration of SO_3 was kept constant instead of being allowed to be consumed by the reactions with methanol.

Overall, we suggest that researchers consult the following checklist when setting up cluster dynamics simulations:

- Check that the combinations of conditions (temperature and concentrations) is realistic, and corresponds to something actually found in the atmosphere (or in an experimental setting). For example, ground-level concentrations of condensing species cannot be directly used in simulations of low-temperature, high-altitude conditions.
- Vapour concentration: Should it be constant or allowed to be consumed?
- Check that the losses are realistic. For instance, very high concentration of vapours in the polluted atmosphere, will also be associated with a large condensation sink.
- Competing pathways. If a new process catalyses a studied new reaction, it should be checked if it also catalyse the baseline reaction.
- Check the absolute J-value when considering the enhancing effects of compounds. If the calculated J-value is essentially zero a few orders of magnitude enhancement will not change the fact that no particles are formed.
- Check the total elemental composition. If alternative reactions of e.g SO_3 (not leading to sulfuric acid) are included in the calculation then less SO_3 is available to form sulfuric acid and this needs to be accounted for. Back-check that the required total amount of SO_2 is realistic.

To illustrate the sensitivity of cluster distribution dynamics models to the input parameters we performed ACDC simulations on the sulfuric acid - dimethylamine system. In Figure 6 the particle formation rate J as a function of sulfuric acid concentration is shown. The formation rate is simulated both with and without losses (constant coagulation sink of 2.6×10^{-3} s^{-1}) and with either a fixed sulfuric acid free monomer concentration or fixed total sulfuric acid concentration as sources. The effect of modifying all cluster formation free energies by ± 1 kcal/mol is illustrated by shading.

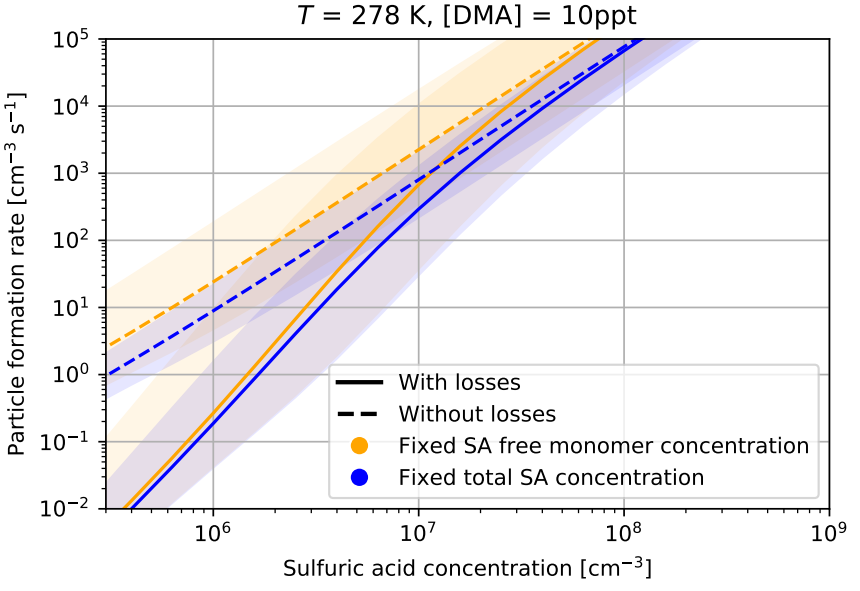

Figure 6: The simulated formation rate as a function of sulfuric acid concentration at $T = 278$ K and 10 ppt dimethylamine.

Losses have a large effect (up to 2-3 orders of magnitude) on the particle formation rate when the concentration of sulfuric acid is low, which is common in many atmospheric environments. The effect from how to constrain the sulfuric acid concentration (whether the sulfuric acid monomers are kept as a fixed free monomer concentration or a fixed total sulfuric acid concentration) is smaller, but can still lead to at least a factor of two differences in the formation rates. The ±1 kcal/mol change to the free energies corresponds to the best attainable accuracy of quantum chemical methods such as CCSD(T) calculations and is seen to manifest as 1-2 orders of magnitude changes in the formation rate. This clearly shows, in accordance with the above checklist, that it is very important to correctly assign the input parameters to match the simulated conditions as even minor wrong assignments can lead to several orders of magnitude difference in the simulated particle formation rates. To better quantify the effect of using different constraints on the sulfuric acid concentration, Figure 7 shows the enhancement factor with respect to the formation rate J as a function of dimethylamine concentration.

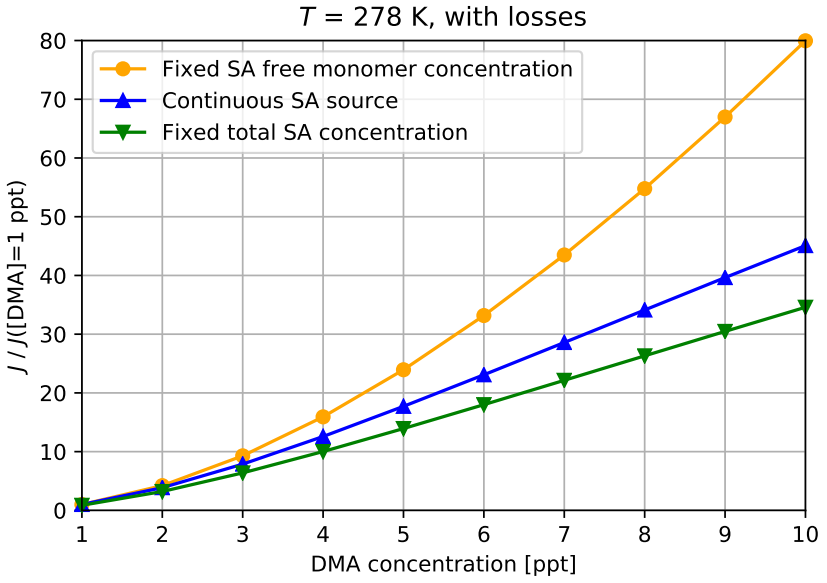

Figure 7: The enhancement in particle formation rate J using different constraints for the sulfuric acid concentration. Fixed constant sulfuric acid monomer concentration at 10^7 molecules cm^{-3} (yellow), fixed total sulfuric acid concentration at 10^7 molecules cm^{-3} (green) and a continuous source of sulfuric acid monomers (blue).

Here the enhancement factor is defined as $J_{\text{new}}/J_{\text{base}}$, where J_{base} is given by:

$$J_{\text{base}} = J\left([\text{DMA}]_{\text{constant = 1 ppt}}, [\text{H}_2\text{SO}_4]_{\text{constant = } 10^7 \text{molecules cm}^{-3}}\right) \tag{6}$$

There are large differences in the enhancing potential of the formation rate depending on the constraint put on the sulfuric acid concentration. Using a fixed free sulfuric acid monomer concentration, will allow a higher enhancement at increasing dimethylamine concentration, as with increasing DMA concentration the total concentration of sulfuric acid molecules, including those bound to clusters, increases. This is essentially an unphysical behaviour if the sulfuric acid concentration is not in large excess. As illustrated by these sensitivity tests it is of utter importance that the above mentioned checklist is followed when applying cluster distribution dynamics models. The potentially largest error involved in particle formation studies using cluster distribution dynamics models still lies in the applied thermochemistry and especially whether the correct global minimum structure has been found. The next sections will present how to obtain as accurate as possible thermochemical parameters for the models using state-of-the-art quantum chemical methods.

4. Thermochemistry from Quantum Chemistry

4.1. Fundamental Thermochemistry Calculations

This section presents how thermodynamical functions are calculated using statistical mechanics and the input from quantum chemical calculations. The following equations closely follow the computational chemistry book by Frank Jensen (Jensen, 2006) and how thermochemistry is calculated in the Gaussian program (Ochterski, 2000). The fundamental key quantity in statistical mechanics is the total molecular partition function q_{tot}, which allows for the calculation of all macroscopic functions:

$$q_{\text{tot}} = \sum_{i}^{\infty} g_i e^{(-\varepsilon_i/k_{\text{B}}T)} \tag{7}$$

Here, ε_i are the energy levels and g_i is the degeneracy of each energy level. The sum runs over all possible quantum states in the molecule. From q_{tot} important thermodynamic functions such as the enthalpy (H) and entropy (S) can be calculated:

$$H = k_{\text{B}}T^2 \left(\frac{\partial \ln(q_{\text{tot}})}{\partial T}\right)_V + k_{\text{B}}TV \left(\frac{\partial \ln(q_{\text{tot}})}{\partial V}\right)_T \tag{8}$$

$$S = k_{\text{B}}T \left(\frac{\partial \ln(q_{\text{tot}})}{\partial T}\right)_V + k_{\text{B}} \ln(q_{\text{tot}}) \tag{9}$$

From these thermodynamic functions the Gibbs free energy (G) can be obtained at any given atmospheric temperature:

$$G = H - TS = k_{\text{B}}TV \left(\frac{\partial \ln(q_{\text{tot}})}{\partial V}\right)_T - k_{\text{B}}T \ln(q_{\text{tot}}) \tag{10}$$

It is standard practice to calculate the thermochemical quantities using quantum chemical methods at 298.15 K and at reference pressure of 1 atm. The Gibbs free energy can easily be re-calculated at any atmospheric temperature using equation (10) under the assumption that H and S do not change considerably with temperature. It should be mentioned that the temperature effect is included in an *ad hoc* fashion using statistical mechanics, and the cluster structures remain 0 K minimum energy structures. To correctly account for temperature effects, long molecular dynamics or Monte Carlo simulations need to be performed. However, for atmospheric molecular clusters, this is not trivial as commonly used molecular mechanics force fields are not able to correctly describe proton transfer reactions, which are prevalent during the cluster formation process. A few specialized force fields for cluster formation of sulfuric acid and water exists such as the classical potential by Ding *et al.* (Ding et al., 2003a) and reactive force field by Stinson *et al.* (Stinson et al., 2016), which is based on a self-consistent iterative two-state empirical valence bond (EVB) method. Development of new force fields is important as *ab initio* molecular dynamics with current implementations is far too time consuming.

Essentially, the partition function q_{tot} depends on all possible quantum states of the molecular system, and in practice is impossible to obtain. For an isolated molecular system, the total energy, enthalpy and entropy (E_{tot}, H_{tot} and S_{tot}) can be approximated by the individual terms involving translational, rotational, vibrational and electronic states:

$$E_{\text{tot}} = E_{\text{trans}} + E_{\text{rot}} + E_{\text{vib}} + E_{\text{elec}} \tag{11}$$

$$H_{\text{tot}} = H_{\text{trans}} + H_{\text{rot}} + H_{\text{vib}} + H_{\text{elec}} \tag{12}$$

$$S_{\text{tot}} = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} + S_{\text{elec}} \tag{13}$$

This approximation of the energy implies that the total partition function can be written as a product of the individual contributions:

$$q_{\text{tot}} = q_{\text{trans}} q_{\text{rot}} q_{\text{vib}} q_{\text{elec}} \tag{14}$$

The translational degrees of freedom of a molecule/cluster with mass m is given by:

$$q_{\text{trans}} = \left(\frac{2\pi m k_{\text{B}}T}{h^2}\right)^{3/2} V = \left(\frac{2\pi m k_{\text{B}}T}{h^2}\right)^{3/2} \frac{k_{\text{B}}T}{P} \tag{15}$$

Here V is the volume, which depends on the number of particles. Using the ideal gas law the volume can be replaced with the partial pressure P of molecules/clusters. For calculating the rotational degrees of freedom, it is assumed that the rotation of the molecule is independent from the rotational and vibrational quantum numbers. This is known as the ***rigid-rotor*** approximation and the rotational partition function is calculated as:

$$q_{\text{rot}} = \frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 k_{\text{B}}T}{h^2}\right)^{3/2}\sqrt{I_1 I_2 I_3} \tag{16}$$

Here I_i are the principal moments of inertia and σ is the rotational symmetry factor. The symmetry factor depends on the molecular point group and is for instance 2 for a water molecule (C_{2v}), 3 for ammonia (C_{2v}) and 12 for ammonium (T_d). For most cluster systems there is no symmetry and the symmetry factor is simply unity as they belong to the C_1 point group. This is not necessary the case of the reacting monomers, where it is important to correctly assign the symmetry factor (Kubečka et al., 2019). As noted from equation (15) and (16), the translational and rotational partition functions only depend on the molecular geometry via the moments of inertia and the atomic masses.

The electronic partition function is calculated as a sum over all electronic quantum states in the system:

$$q_{\text{elec}} = \sum_i^{\infty} g_i e^{(-\epsilon_i/k_{\text{B}}T)} \tag{17}$$

For closed shell molecules of atmospheric interest the separation between the ground state and electronic excited states is significantly larger than $k_{\text{B}}T$ implying that only the ground state is important in the summation yielding $q_{\text{elec}} = 1$. This is not necessarily the case for open-shell systems. For instance, due to spin-orbit coupling the OH radical has two low lying states with a vertical excitation energy between the ground state ($^2\Pi_{3/2}$) and first electronic state ($^2\Pi_{1/2}$) of 140 cm^{-1} (Li et al., 2006). This yields a $q_{\text{elec}} = 2 + 2e^{140/k_{\text{B}}T} = 3.019$ at 298.15 K.

The vibrational degrees of freedom are described as ***harmonic oscillators*** and then the partition function can be written as:

$$q_{\text{vib}} = \prod_i^{3N-6} \frac{e^{-h\nu_i/2k_{\text{B}}T}}{1 - e^{-h\nu_i/k_{\text{B}}T}} \tag{18}$$

Here ν_i are the vibrational frequency calculated using quantum chemical methods and the sum runs over all $3N$ degrees of freedom in a molecule, minus 3 from translation and 3 from rotation. It should be noted that for linear molecules there are only 2 degrees of freedom from rotation and the sum runs over $3N - 5$ vibrational modes. From equation (15)-(18) the total partition function can be calculated and inserted in (8) to obtain the enthalpy H:

$$H_{\text{trans}} = \frac{5}{2}RT \tag{19}$$

$$H_{\text{rot}} = \frac{3}{2}RT \tag{20}$$

$$H_{\text{vib}} = R\sum_i^{3N-6}\left(\frac{h\nu_i}{2k_{\text{B}}} + \frac{h\nu_i}{k_{\text{B}}}\frac{1}{e^{h\nu_i/k_{\text{B}}T} - 1}\right) \tag{21}$$

$$H_{\text{elec}} = 0 \tag{22}$$

In a similar manner the entropy S can be obtained from (9):

$$S_{\text{trans}} = \frac{5}{2}R + R\ln\left[\left(\frac{2\pi m k_{\text{B}}T}{h^2}\right)^{3/2}\frac{k_{\text{B}}T}{P}\right] \tag{23}$$

$$S_{\text{rot}} = \frac{3}{2}R + R\ln\left[\frac{\sqrt{\pi}}{\sigma}\left(\frac{8\pi^2 k_{\text{B}}T}{h^2}\right)^{3/2}\sqrt{I_1 I_2 I_3}\right] \tag{24}$$

$$S_{\text{vib}} = R\sum_i^{3N-6}\left(\frac{h\nu_i}{k_{\text{B}}T(e^{h\nu_i/k_{\text{B}}T} - 1)} - \ln\left[1 - e^{-h\nu_i/k_{\text{B}}T}\right]\right) \tag{25}$$

$$S_{\text{elec}} = R\ln(g_0) \tag{26}$$

These equations allow the calculations of the Gibbs free energy G for a cluster of arbitrary composition.

4.2. Treating Low Vibrational Frequencies

The rigid-rotor and harmonic oscillator approximations are employed in the calculations for obtaining the free energy via equation (10) and (19)-(26). However, the vibrational entropy contribution S_{vib} should be handled with care. For instance, the existence of hindered internal rotations can lead to low vibrational frequency modes that gives rise to large errors in the correspondingly calculated harmonic oscillator partition function. Several approaches to correct for hindered rotation exists, (see for example the approach by Truhlar (Truhlar, 1991) and Pfaendtner *et al.* (Pfaendtner et al., 2007) for a general review) but these are approaches are limited to isolated molecules, and not straightforward to apply to molecular clusters.

When calculating the vibrational frequencies of clusters, it is common that several low-lying frequencies with less than 100 wave numbers emerge. Essentially, these low frequency vibrations are not strictly vibrations as such, but corresponds to a collective pivotal movement of the cluster. The second term in the vibrational contribution to the total entropy of a harmonic oscillator, given by (25), goes to infinity as the vibrational frequency goes towards zero, and thus the harmonic oscillator approximation becomes a large source of errors when low vibrational frequencies are present in the cluster. One way to correct this deficiency is to employ a quasi-harmonic approximation. Ribeiro *et al.* (Ribeiro et al., 2011) employed a simple approach where all vibrational frequencies below 100 cm^{-1} were raised to up 100 cm^{-1}. Stefan Grimme (Grimme, 2012) proposed an alternative approach where the contribution of low frequencies to the entropy is replaced by a corresponding rotational entropy. For each normal mode ν_i, below a certain cut-off value ν_0, the moment of inertia is given by a free-rotor:

$$I_{\text{free}} = \frac{\hbar}{4\pi\nu_i} \tag{27}$$

To restrict the effective moment of inertia (I') to reasonable values, an average molecular moment of inertia B_{av} is introduced:

$$I' = \frac{I_{\text{free}} B_{\text{av}}}{I_{\text{free}} + B_{\text{av}}} \tag{28}$$

Here B_{av} represents a limiting value for small values of ν_i. Usually, a value of $B_{\text{av}} = 10^{-44}$ kg m^2 is applied. In the Grimme-type quasi-harmonic approximation the entropy of a low vibrational mode is given by:

$$S_{\text{rot,qh}} = \frac{1}{2}R + R\ln\left[\left(\frac{8\pi^3 I' k_{\text{B}} T}{h^2}\right)^{1/2}\right] \tag{29}$$

This effectively replace all the vibrational frequencies below the cut-off value ν_0 (usually 50-200 cm^{-1}) with the entropy of a free-rotor. The effect of the choice of cut-off, in the range 100-200 cm^{-1}, on the reaction free energies for sulfuric acid - dimethylamine and sulfuric acid - putrescine clusters has been shown to be negligible (Elm et al., 2017d). To smoothly connect between the two approximations the $S_{\text{rot,qh}}$ and S_{vib} values are connected by the damping function given by Head-Gordon (Chai and Head-Gordon, 2008):

$$S_{\text{rot,qh - vib}} = \left(\frac{1}{1+(\nu_0/\nu_i)^4}\right) S_{\text{vib}} + \left(1 - \frac{1}{1+(\nu_0/\nu_i)^4}\right) S_{\text{rot,qh}} \tag{30}$$

The quasi-harmonic approximations substantially reduce the error introduced by low vibrational frequencies. It is computationally inexpensive to apply and can efficiently be applied on, for instance, Gaussian output files by using the Goodvibes.py python script developed by Paton and co-workers (Funes-Ardois and Paton, 2016). Thus we recommend that the quasi-harmonic approximation is routinely applied to studies of atmospheric molecular clusters, especially when large clusters are studied as more low vibrational frequencies emerge when the cluster size increases and this will result in an increased error.

4.3. Anharmonicity

In general, anharmonicity is sometimes defined as any effect that makes the partition function deviate from the single-minimum rigid-rotor harmonic oscillator partition function. Depending on the community, anharmonicity may refer to:

1) The presence of higher energy conformers on the thermodynamic properties.
2) Anharmonicity present in the vibrational frequencies of the individual molecules.

Here we will refer to case 2) as anharmonicity. The former will be referred to as the effect of conformers on the calculation of the free energies and will be discussed in section 6.3.

Vibrational anharmonicity can be calculated by including higher order terms in the Taylor expansion of the energy as a function of the nuclear coordinates. Several different models exist such as the vibrational self-consistent method (VSCF) (Bowman, 1978), second-order vibrational perturbation theory (VPT2) (Willetts et al., 1990) and vibrational coupled–cluster theory (VCC) (Christiansen, 2004a,b). These approaches are quite computationally demanding and are usually neglected in atmospheric cluster calculations.

Partanen *et al.* (Partanen et al., 2012, 2016a) studied the potential energy surfaces of the sulfuric acid monohydrate in order to identify the relative free energy of different conformers, the effect of anharmonicity and the effect of conformers. This approach is extremely computationally demanding as it required the full potential energy surface for each cluster configuration. Due to the difficulties in calculating the anharmonic vibrational frequencies the authors concluded that the effect of multiple conformers should be taken into consideration before correcting for anharmonicity.

Anharmonic calculations can be applied to small benchmark systems with the goal of deriving universal scaling factors between the harmonic and anharmonic vibrational frequencies. This process has been applied to many single molecules and has provided an *ad hoc* approach for approximating anharmonic frequencies (Pople et al., 1981; III). However, this is not a well established procedure for molecular clusters and has only been applied to a few systems. It should be noted that scaling factors correct for both the deficiency in the applied *ab initio* method, as well as anharmonicity within a single factor.

Temelso *et al.* applied scaling factors on pure $(H_2O)_{1-10}$ water clusters (Temelso et al., 2011) and $(H_2SO_4)_{1-2}(H_2O)_{0-6}$ (Temelso et al., 2012a,b) clusters and found that scaled anharmonicity had a significant effect on the calculated Gibbs free energy of the clusters. Myllys *et al.* (Myllys et al., 2016b) reported scaling factors for three popular DFT functionals (PW91, M06-2X and ωB97X-D) based on a test set of six small cluster formation reactions. It was found that the stretching motion along a hydrogen bond towards the acceptor atoms with frequencies in the range 2000-3000 cm^{-1} showed large discrepancies compared to the other points, indicating that there is a large nonsystematic effect from including anharmonicity. However, significantly more work is needed to obtain larger benchmark sets in order to derive scaling factors to account for anharmonicity that can be routinely applied to atmospheric cluster formation studies.

4.4. Cluster Reactions and Thermochemistry

Being able to obtain the Gibbs free energy ($G[i]$) of each molecule/cluster species i we can now proceed into the calculation of the cluster reaction and binding free energies. For a given chemical reaction:

$$\mathrm{A} + \mathrm{B} + \mathrm{C} + \cdots \leftrightharpoons \mathrm{D} + \mathrm{E} + \mathrm{F} + \cdots \tag{31}$$

The reaction Gibbs free energy ($\Delta G_{\mathrm{react}}$) is given by:

$$\Delta G_{\mathrm{react}} = \big(G[\mathrm{D}] + G[\mathrm{E}] + G[\mathrm{F}] + \dots\big) - \big(G[\mathrm{A}] + G[\mathrm{B}] + G[\mathrm{C}] + \dots\big) \tag{32a}$$

Or more generally:

$$\Delta G_{\mathrm{react}} = \sum G_{\mathrm{products}} - \sum G_{\mathrm{reactants}} \tag{32b}$$

Analogically, the ***binding free energy*** (ΔG_{bind}) for forming a cluster $\mathrm{A} + \mathrm{B} + \mathrm{C} + \cdots \leftrightharpoons (\mathrm{ABC}\cdots)$ can be calculated as follows:

$$\Delta G_{\mathrm{bind}} = G[(\mathrm{ABC}\cdots)] - \big(G[\mathrm{A}] + G[\mathrm{B}] + G[\mathrm{C}] + \cdots\big) \tag{33a}$$

$$\Delta G_{\mathrm{bind}} = G_{\mathrm{cluster}} - \sum G_{\mathrm{monomers}} \tag{33b}$$

The binding free energy is a measure of the total free energy gain in assembling the cluster from infinitely separated monomers. Note that the binding free energy is also commonly termed the formation free energy, but confusingly it differs slightly by definition from the quantity called formation free energy in the field of nucleation, see section 4.5. More appropriately, a given cluster $(\mathrm{ABC}\cdots)$ can be formed in stepwise reactions:

$$\mathrm{A} + \mathrm{B} \leftrightharpoons (\mathrm{AB}) \tag{34}$$

$$\begin{array}{rcl} (\mathrm{AB}) + \mathrm{C} & \leftrightharpoons & (\mathrm{ABC}) \\ & \vdots & \end{array} \tag{35}$$

These types of reactions will be denoted as ***addition reactions*** (ΔG_{add}). The reverse reactions are also possible which are denoted ***fragmentation reactions*** (ΔG_{frag}). It should be mentioned that fragmentation is occasionally also termed dissociation or fission and in general the fragmentation of monomers is called evaporation. In the case of reaction (35), the corresponding reaction free energies are given by:

$$\Delta G_{\mathrm{add}} = G[(\mathrm{ABC})] - \big(G[\mathrm{AB}] + G[\mathrm{C}]\big) \tag{36}$$

$$\Delta G_{\mathrm{frag}} = \big(G[\mathrm{AB}] + G[\mathrm{C}]\big) - G[(\mathrm{ABC})] = -\Delta G_{\mathrm{add}} \tag{37}$$

The (ABC) cluster can also be formed *via* the reaction A + (BC) $\leftrightharpoons$ (ABC) and as the clusters grow larger, there exist several different formation pathways. These can also include the collision between smaller clusters such as: (AB)+(CD) $\leftrightharpoons$ (ABCD). This implies that for a large cluster system, an array of different formation and fragmentation pathways exist. It should be mentioned that the above equations are also valid for other thermochemical functions, such as the enthalpy and entropy, by simply exchanging G with either H or S.

Figure 8 shows an example of a cluster formation reaction and a cluster fragmentation reaction involving sulfuric acid and ammonia.

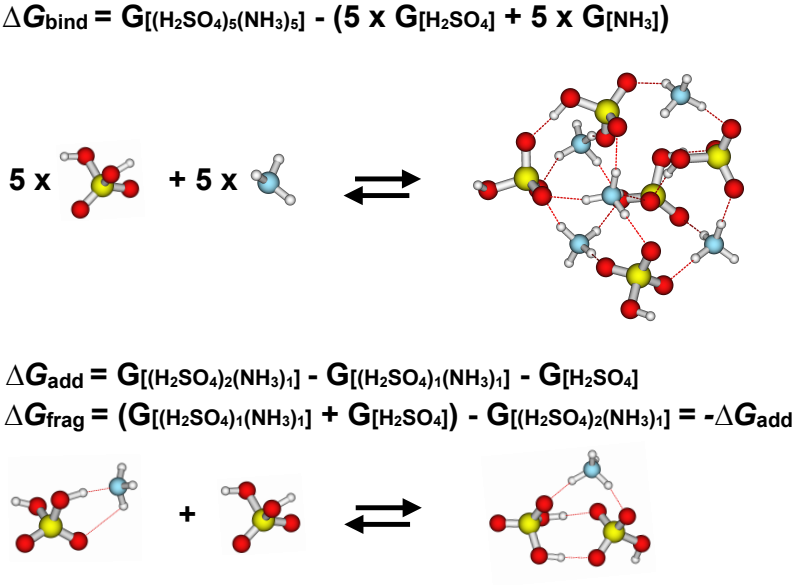

Figure 8: Illustration of a cluster formation reaction (top) and an addition/fragmentation reaction (bottom).

As also shown by the illustration, the binding free energy value ΔG_{bind} is a theoretical quantity that corresponds to the non-realistic process of the simultaneous collision of five sulfuric acid and five ammonia molecules to form the cluster. The cluster monomer addition or fragmentation free energy relates to a concrete reaction, and can be used directly to calculate the evaporation rate using Equation (2).

4.5. Role of the Reference Pressure and Relation between the Binding and Formation Free Energies

As the entropy depends on the pressure entering Equation (23) for both the reactant(s) and the product(s), the numerical values for the binding or reaction free energies typically also depend on this pressure. The exception to this are reactions where the number of reactants and product entities is equal, in which case the pressure-dependent terms cancel out. As the used reference pressure (or reference concentration) is an arbitrary number, this may seem counterintuitive and can lead to misunderstandings especially when the sign of ΔG_{add} is assumed to carry a special meaning, for example, with a negative ΔG_{add} assumed to imply "spontaneity". This implies that the threshold free energy for a spontaneous reaction is different in gas-phase chemistry, where a reference pressure of 1 atm = 101325 Pa

is commonly used, and in liquid-phase chemistry, where the reference concentration is usually 1 mol/L. In reality, the reference pressure cancels out in the calculation of dimensionless equilibrium constants or evaporation rates through detailed balance.

While the value of the reference pressure is thus irrelevant and is cancelled out, for example, in the calculation of evaporation rates, the actual concentrations of the molecular species participating in clustering processes are naturally of importance for understanding the process. Using the binding free energy that we have introduced in Section 4.4, calculated at the reference pressure p_{ref}, the equilibrium partial pressure for a multicomponent cluster with i_n molecules of type $n = 1, 2, ..., n$, is:

$$\frac{p_{\text{cluster}}}{p_{\text{ref}}} = \left(\frac{p_{1,1}}{p_{\text{ref}}}\right)^{i_1} \left(\frac{p_{1,2}}{p_{\text{ref}}}\right)^{i_2} ... \left(\frac{p_{1,n}}{p_{\text{ref}}}\right)^{i_n} ... \exp\left(\frac{-\Delta G_{\text{bind}}}{RT}\right) \tag{38}$$

Here, the partial pressure of monomers of type n is given by $p_{1,n}$ (first subscript 1 refers to monomer). Here, the reference pressure p_{ref} is related to the translational partition function (q_{trans}) and hence the system volume V via equation (15). This relation can be converted to:

$$p_{\text{cluster}} = p_{\text{ref}} \exp\left(-\frac{\Delta G_{\text{bind}} - RT \sum_n i_n \ln\left(\frac{p_{1,n}}{p_{\text{ref}}}\right)}{RT}\right) \tag{39}$$

where the index n goes over all the types of molecules. Within the realm of the nucleation theory, the partial pressure of a clusters in equilibrium vapour is written as (Courtney, 1961):

$$p_{\text{cluster}} = p_{\text{all clusters}} \exp\left(-\frac{\Delta G_{\text{form}}}{RT}\right) \tag{40}$$

where ΔG_{form} is the cluster formation free energy as defined in statistical physics-based nucleation theory, and $p_{\text{all clusters}}$ is strictly speaking the sum of partial pressures of monomers and all other clusters is the system. Often it is approximated to be equal to the partial pressure of monomers assuming that they dominate the size distribution, $p_{\text{all clusters}} \approx \sum_n p_{1,n}$, as already pointed out for cluster concentration in the context of equation (3). To directly compare equation (39) with (40), we can rewrite (39) as the following:

$$p_{\text{cluster}} = p_{\text{all clusters}} \exp\left(-\frac{\Delta G_{\text{bind}} - RT \sum_n i_n \ln\left(\frac{p_{1,n}}{p_{\text{ref}}}\right) + RT \ln\left(\frac{p_{\text{all clusters}}}{p_{\text{ref}}}\right)}{RT}\right) \tag{41}$$

By comparing equation (39) with equation (41) we notice than the nucleation style formation free energy is related to the binding free energy as:

$$\Delta G_{\text{form}} = \Delta G_{\text{bind}} - RT \sum_n i_n \ln\left(\frac{p_{1,n}}{p_{\text{ref}}}\right) + RT \ln\left(\frac{p_{\text{all clusters}}}{p_{\text{ref}}}\right) \tag{42}$$

In traditional nucleation studies, ΔG_{form} curves as a function of the number of molecules i, or more generally surfaces as a function of several i_n, are computed not only at equilibrium conditions but also in the supersaturated nucleating vapour, as we have done in Figure 3. The critical cluster is identified as a maximum of such a one-component curve, or a saddle point on a multicomponent surface, as mentioned after equation (3). The numerical values of the ΔG_{form} values on these surfaces are typically much higher (less negative/more positive) than on the standard surface ΔG_{bind}. The last term $RT \ln\left(\frac{p_{\text{all clusters}}}{p_{\text{ref}}}\right)$ in equation (42) is constant with respect to cluster size and composition. It only corresponds to a shift in the entire free energy surface and leaves conclusions about barriers, valleys, saddle points etc intact and this term is often neglected. Free energy surfaces plotted in this fashion are sometimes denoted "actual" free energy surfaces as opposed to reference or standard free energy surfaces, where all partial pressures are set to 1 atm:

$$\Delta G_{\text{actual}} = \Delta G_{\text{bind}} - RT \sum_n i_n \ln\left(\frac{p_{1,n}}{p_{\text{ref}}}\right) \tag{43}$$

Equations (42) and (43) indicate that the monomer pressures are taking the role of the reference pressure when converting from binding free energies to the nucleation style formation free energies. In a one-component case where monomers are dominating the size distribution (i.e $p_{\text{all clusters}} \approx p_1$), equation (42) simplifies to:

$$\Delta G_{\text{form}} = \Delta G_{\text{bind}} - RT(i-1)\ln\left(\frac{p_1}{p_{\text{ref}}}\right) \tag{44}$$

If monomers (of all types) dominate, the multicomponent formula in equation (42) can be written as:

$$\Delta G_{\text{form}} = \Delta G_{\text{bind}} - RT\sum_{n\geq 1} i_n \ln\left(\frac{p_{1,n}}{p_{\text{ref}}}\right) + RT\ln\left(\frac{\sum_n p_{1,n}}{p_{\text{ref}}}\right) \tag{45}$$

This can be further approximated if only one type of monomer (for instance type 1) dominates even all the other monomers and we get:

$$\Delta G_{\text{form}} = \Delta G_{\text{bind}} - RT(i_1-1)\ln\left(\frac{p_{1,1}}{p_{\text{ref}}}\right) - RT\sum_{n\geq 2} i_n \ln\left(\frac{p_{1,n}}{p_{\text{ref}}}\right) \tag{46}$$

This is, for instance, the case with water vapour in the atmosphere. With or without the last term in equation (42), these equations lead to monomers having nonzero ΔG_{form} which originate from the definition of (40) and is not an unphysical effect as the exponent function calculated for a monomer should give the fraction of monomers from all clusters and not unity (unless monomers of one type are practically 100% of the clusters). Alternatively, the self-consistent distribution function by Wilemski and Wyslouzil (Wilemski and Wyslouzil, 1995) can be used which intrinsically sets the monomer free energies to zero.

A frequent misconception concerning the ΔG_{form} surfaces is that the product pressure is neglected, as it is kept at 1 atm. The explanation for this seeming discrepancy is that the formation free energy in nucleation theory is used to solve for the product cluster (for example, in the case of CNT the critical cluster) concentration. This contrasts with the typical calculations done in, for example, chemical thermodynamics, where both the reactant and the product concentrations are used to deduce whether a reaction will proceed in the forward or reverse direction.

Another common misunderstanding is that the pressure dependence of quantum chemically computed free energies implies that the effect of bath gas molecules (in practice e.g. N_2) is taken into account in the calculations. This is incorrect - modelling the effects of total pressure requires a separate set of simulations using RRKM theory (Marcus, 1952) or molecular dynamics simulations. The cluster kinetics calculations described here implicitly assume all clusters to be thermalised. In other words, for example evaporation and collision rates correspond to the high-pressure limits, and effects of the pressure of the nature of the bath gas are typically neglected.

5. Commonly Applied Quantum Chemical Methods

We assume that the reader is familiar with basic quantum chemical methods and refer to the "Introduction to Computational Chemistry" book by Frank Jensen (Jensen, 2006) if refreshing is needed. To calculate the cluster Gibbs free energies the equilibrium geometries and vibrational frequencies are required. This require the calculation of the gradient of the energy for obtaining stationary points, and the second derivative of the energy (the Hessian) for obtaining the vibrational frequencies. These calculations are significantly more computationally demanding than simple energy evaluation, which has limited the calculations of cluster geometries to density functional theory (DFT) and in some cases MP2 (Kurtén et al., 2007a; Temelso et al., 2012a,b). Applying more accurate methods has been limited to very small systems and are only used for benchmarking purposes. For instance, accurate CCSD(T)-F12/cc-pVDZ-F12 optimizations have only been carried out for small $(H_2SO_4)(H_2O)$ and $(H_2SO_4)(NH_3)$ clusters (Long et al., 2016). Even applying conventional MP2 is too computationally demanding for medium sized clusters. This leaves DFT as the only possibility when modelling the structures of atmospheric clusters.

The binding free energy calculated from equation (33b) can be divided into a pure electronic contribution (ΔE_{bind}) and a thermal contribution ($\Delta G_{\text{bind, thermal}}$) to the Gibbs free binding energy:

$$\Delta G_{\text{bind}} = \Delta E_{\text{bind}} + \Delta G_{\text{bind, thermal}} \tag{47}$$

It should be mentioned that the sign of ΔE_{bind} depends slightly on convention. We calculate ΔE_{bind} analogously to ΔG_{bind} in (33b). Care should also be taken with whether or not the zero point energy (ZPE) contribution is included

in the ΔE_{bind} or as part of the thermochemistry calculation. Occasionally, the ΔE_{bind}+ZPE is denoted the cluster dissociation energy (D_0) which might cause some confusion. Out of the two contributions in equation (47), it has been shown that the DFT ΔE_{bind} term is the largest source of errors in the Gibbs free energy calculations and that the $\Delta G_{\text{bind, thermal}}$ term is fairly similar when using different density functionals (Elm and Mikkelsen, 2014). A pragmatic approach is to use a cheaper computational method such as DFT with a small basis set for obtaining the geometry and vibrational frequencies, hence the $\Delta G_{\text{bind, thermal}}$ contribution, and then use a higher level of theory such as coupled cluster to obtain the ΔE_{bind}-value, calculated on top of the DFT geometry. This is usually denoted using a double slash (//) notation, where the method before the // denotes the quantum chemical method used to obtain the single point energy of the molecule/cluster, and the method after the // is used to obtain the molecular equilibrium geometry and vibrational frequencies. This allows for the calculation of an approximate high level (HL) binding Gibbs free energies as follows:

$$\Delta G_{\text{bind}}^{\text{HL, approx}} = \Delta E_{\text{bind}}^{\text{HL}} + \Delta G_{\text{bind, thermal}}^{\text{DFT}} \tag{48}$$

Unfortunately, different research groups have their own favourite functional, basis set and higher level correction to apply, which makes direct comparison between studies difficult. In the following we will briefly review the methods that have been applied to obtain the cluster structures, vibrational frequencies and higher level corrections for the binding energies.

5.1. Methodologies

5.1.1. B3LYP

Earliest studies applied the B3LYP functional to obtain the cluster binding free energies. This was based on computational limitations at the time and the general popularity of the method. Primarily, the hydration of sulfuric acid clusters have been studied in the early work of Bandy and Ianni (Bandy and Ianni, 1998; Ianni and Bandy, 2000) and by Re *et al.* (Re et al., 1999) using the B3LYP functional. These studies mainly revolved around obtaining the cluster structures and to resolve potential proton transfers in the clusters. A benchmark study by Kurtén *et al.* in 2006 showed that applying B3LYP leads to too weakly bound clusters (Kurtén et al., 2006). This has been further corroborated by several benchmark studies (Elm et al., 2013a; Leverentz et al., 2013) and the B3LYP functional is thus not recommended for cluster studies. However, the empirical dispersion corrected version B3LYP-D3 has shown promising performance in comparison with higher level methods (Elm and Kristensen, 2017; Schmitz and Elm, 2020)

5.1.2. PW91

Based on comparison with experimental free energy data of the first two hydration reactions of sulfuric acid (Nadykto and Yu, 2007), as well as vibrational spectra and structural parameters of the bisulfate ion, sulfuric acid and the sulfuric acid monohydrate (Nadykto et al., 2009), Nadykto and co-workers have applied the PW91/6-311++G(3df,3pd) level of theory. A benchmark study by Elm *et al.* corroborated that out of 22 tested functionals, PW91 was only surpassed by M06-2X in yielding reaction free energies in agreement with the experimental results (Elm et al., 2012). However, the large 6-311++G(3df,3pd) basis set is commonly applied which will make it difficult to apply this methodology to very large clusters. Being a GGA, the PW91 density functional with a small basis set might be the only choice for modelling extremely large systems such as clusters consisting of large organic compounds (Elm, 2019b). Depalma *et al.* (DePalma et al., 2014) showed that the PW91/6-31++G(d,p) level of theory yielded slightly too negative formation free energies of $(H_2SO_4)_x(NH_3)_x$ clusters, with $x = 2$, 3 or 4 compared to PW91/6-311++G(3df,3pd) calculations. However, the difference between the PW91/6-31++G(d,p) and PW91/6-311++G(3df,3pd) free energies were consistent for the studied clusters and the trend in the Gibbs free energies were correctly described by the lower level of theory.

5.1.3. B3RICC2

The RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7 level of theory (termed B3RICC2) has routinely been applied to study atmospheric relevant cluster systems (Ortega et al., 2012; Paasonen et al., 2012; Kupiainen-Määttä et al., 2012; Henschel et al., 2014, 2016; Ortega et al., 2016; Olenius et al., 2017; Li et al., 2020b; Lu et al., 2020). The choice was based on that the B3LYP/CBSB7 level of theory is applied in the composite complete basis set method CBS-QB3 (Montgomery, Jr.). The composite methods have been developed for accurate thermochemistry of molecules, but are highly empirical and not necessarily accurate for atmospheric relevant cluster systems. The B3RICC2 level was compared to measured cluster evaporation rates and found to perform well (Ortega et al., 2012). Large differences

between the B3RICC2 and PW91/6-311++G(3df,3pd) calculated thermochemistry have been pointed out (Nadykto et al., 2014). However, this is solely based on the fact that there exists a difference between the numbers calculated by the two methods and with no further comparison with higher level methods, it was difficult to assess which method is actually at fault. The RI-CC2/aug-cc-pV(T+d)Z single point energy correction was most likely chosen as it was the highest level of theory applicable at the time. However, CC2 (Christiansen et al., 1995) has been developed primarily for excited state properties, so there is little assurance that it should work well for ground state properties. Furthermore, the leading terms in the CC2 equations are dominated by MP2 terms, and thus higher accuracy than regular MP2 cannot be assumed. In fact, benchmark studies have shown that the RI-CC2/aug-cc-pV(T+d)Z single point energies severely over-predicts the stability of atmospheric molecular clusters (Schmitz and Elm, 2020). This will lead to B3RICC2 representing an upper bound of the simulated particle formation rates. Thus, in general, the RI-CC2/aug-cc-pV(T+d)Z single point energy correction cannot be recommended for studying atmospheric clusters.

5.1.4. M06-2X

Based on the benchmark study of Elm *et al.* (Elm et al., 2012), the M06-2X functional should perform more reliably than other functionals for obtaining the structures and reaction Gibbs free energies for small clusters compared to experimental methods. It should be mentioned that the PW6B95-D3 and ωB97X-D functionals were not among the functionals tested in the benchmark set. While it should perform well compared to experimental results, the M06-2X has been found to severely over-predict the binding energy of clusters consisting of sulfuric acid and highly oxidized organic molecules (HOMs) (Elm et al., 2015, 2016b) compared to Coupled Cluster methods. This shows that the M06-2X functional is not well transferable to other cluster systems, which might originate from the fact that it is heavily parameterized.

5.1.5. PW6B95-D3

Studying the $(H_2SO_4)(NH(CH_3)_2)$ and $(H_2SO_4)_2(NH_3)$ clusters, Leverentz *et al.* (Leverentz et al., 2013) found that the binding energies calculated at the PW6B95-D3/MG3S level of theory agreed the best with a higher level CCSD(T)-F12a/jun-cc-pV(T+d)Z reference. To the best of our knowledge the PW6B95-D3 functional has not been applied to any cluster studies. However, it does perform decently in recent benchmark studies on the binding energies with similar performance as M06-2X (Elm and Kristensen, 2017; Schmitz and Elm, 2020).

5.1.6. ωB97X-D

With the general lack of available experimental data, the best cause of action is to compare the approximate DFT methods with a higher level of theory. Using a test set of 107 atmospherically relevant clusters Elm *et al.* (Elm et al., 2013a) studied the performance of the B3LYP, CAM-B3LYP, M06-2X, PW91, LC-PW91, PBE0 and ωB97X-D with a 6-311++G(3df,3pd) basis set in calculating the binding energies compared to a high level DF-LCCSD(T)-F12/cc-pVDZ-F12 reference. Out of the tested functionals, the ωB97X-D functional performed the best with a mean absolute deviations of 2.12 kcal/mol. The PW91 and M06-2X functionals performed slightly worse with mean absolute deviations of 3.28 and 3.46 kcal/mol, respectively. The superiority of the ωB97X-D functional compared to other density functionals have further been corroborated by comparison of the binding energies with CCSD(T) complete basis set estimates for test sets up to 45 atmospheric relevant clusters (Elm and Kristensen, 2017; Schmitz and Elm, 2020). The fact that the ωB97X-D functional performs the best compared CCSD(T)/CBS estimates should currently make it the most reliable density functional to apply for studying atmospheric molecular clusters.

5.2. Higher Level Corrected Binding Energies

As mentioned in section 5, the DFT binding energies constitute the largest source of errors. A common approach is to correct the electronic single point energy of the clusters with a more reliable method such as CCSD(T). It should be noted that the DFT structures are not true minima on the higher level energy surface, though it is assumed that the surfaces are to some extend parallel. Using different DFT functionals (PW91, M06-2X and ωB97X-D) for geometry optimizing the same cluster conformations and calculating the binding energies with LCCSD(T)-F12/cc-pVDZ-F12 on top of each, has shown that there is very little difference (mean absolute error (MAE) of 0.58 kcal/mol for ωB97X-D) in the calculated binding energies (Elm and Mikkelsen, 2014). In a similar manner, it was shown that there was little difference (MAE of 0.40 kcal/mol for ωB97X-D) in the binding energy when changing the basis set used to optimize the DFT geometry from a large 6-311++G(3df,3pd) basis set to a small 6-31++G(d,p) basis set. This illustrates that the DFT geometries are quite decent and that it is less important which of the PW91, M06-2X and ωB97X-D functionals was used for obtaining the geometry and whether the 6-311++G(3df,3pd) or 6-31++G(d,p) basis set was

utilized, as long as a higher level correction is applied. Incidentally, this also implies that older studies that have been performed only at the DFT level can easily improved upon by simply computing a higher level calculation on the structure. Applying high level methods has several limitations:

1) The scaling of, for instance, CCSD(T) with respect to the system size is on the order of N^7.
2) The slow convergence of the correlation energy with respect to the basis set makes it difficult to obtain converged energies.

The scaling with respect to system size can be solved by using localized orbitals. Recently, it was shown by Myllys *et al.* (Myllys et al., 2016a) that Domain based Local Pair Natural Orbital Coupled Cluster methods (DLPNO-CCSD(T)) could be applied for studying clusters consisting of up to 10 molecules. This has been far beyond reach using conventional CCSD(T) methods. The slow basis set convergence can be circumvented by applying explicitly correlated (F12) wave functions. Recently, the explicitly correlated version of the DLPNO-CCSD(T) method has been developed. Schmitz *et al.* (Schmitz and Elm, 2020) studied the performance of these methods and found that the DLPNO-CCSD(T)-F12/cc-pVTZ-F12 level of theory with a TightPNO criterion could yield binding energies in good agreement (MAE < 0.2 kcal/mol) with CCSD(T) complete basis set estimates. In general, care should be taken to use a sufficiently tight PNO criterion. For very weakly bound systems a tighter PNO criterion is needed.

5.3. Basis set Convergence and Superposition Errors

When calculating binding energies, the emergence of basis set superposition errors (BSSE) might become a concern. BSSE arise in the calculation of binding energies as the interacting monomers come into the vicinity of each other. This leads to a lowering of the total electronic energy of the system as each fragment is stabilized by the overlapping basis function from another fragment. This leads to an over-prediction of the binding energy. As the clusters of interest are usually quite large, it is more or less impossible to apply a large enough basis set to completely suppress the BSSE. One approach to suppress BSSE is to apply the counterpoise (CP) correction (Boys and Bernardi, 1970), but this is very time consuming and no guarantee to lead to more reliable results. It has been shown in several studies (Elm et al., 2017a; Myllys et al., 2016b; Elm and Kristensen, 2017), that when limited to small/medium basis sets the uncorrected calculations actually agree better with the complete basis set limit compared to the CP corrected values (see Figure 9 for an illustration). This has been shown to be valid for both the DFT (Myllys et al., 2016b) and DLPNO (Elm et al., 2017a) binding energies. There exist several cost efficient approaches to correct for BSSE. Galano and Alvarez-Idaboy (Galano and Alvarez-Idaboy, 2006) developed an atom by atom approach denoted CP^{aa}. Frank Jensen further extended this approach into a generalized atomic counterpoise correction (ACP(n)) (Jensen, 2010), where n denotes the number of bonds between atoms for defining the subspace of basis functions. In this framework the ACP(1) method also included neighbouring atoms, making it identical to the CP^{aa} method. Kruse and Grimme (Kruse and Grimme, 2012) further developed this concept into a geometric counterpoise correction (gCP) that only depends on the input geometry of the molecule/cluster. However, to the best of our knowledge, none of these approaches have been tested on atmospheric molecular cluster systems.

Besides the potential BSSE, binding energy calculations might also suffer from basis set incompleteness errors (BSIE). This is definitely an issue when calculating the binding energy of a cluster using a small basis set. Benchmark calculations such as performed by Myllys *et al.* (Myllys et al., 2016b) are valuable in quantifying the potential errors one make when applying a too small basis set for calculating the binding energies. At least triple zeta basis sets are required for obtaining properly converged binding energies using DFT. It should be mentioned that geometries and the thermal contribution to the free energy require a significantly smaller basis set than the binding energies and one can get away with a double zeta basis set as long as the binding energy is corrected with a larger basis set. Explicitly correlated (F12) methods highly suppress both the BSSE and BSIE, and results in a good agreement with the complete basis set limit can be obtained even when using a cc-pVDZ-F12 basis set.

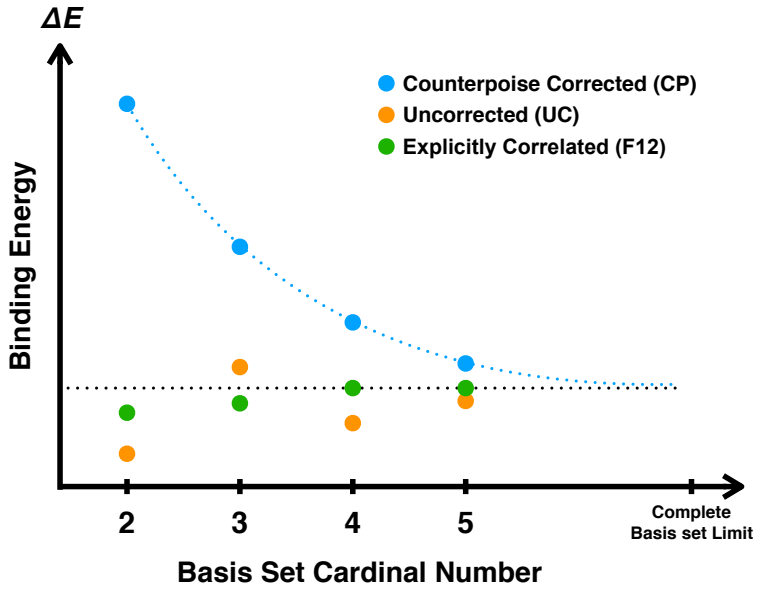

Figure 9: Sketch of the convergence of the DLPNO-CCSD(T) or DFT electronic binding energies of atmospheric molecular clusters.

5.4. Remaining issues

While there has been a clear advancement in the accuracy of the applied quantum chemical methods to atmospheric clusters, there are some central issues that remain unsolved or are limited by current technology/development. For instance, we are still limited to DFT with a relatively small basis set for obtaining the cluster structures and vibrational frequencies. As experimental data on cluster geometries are extremely scarce, ideally the structures should be obtained using more accurate methods. However, even MP2 has previously been limited to medium sized systems. Recently, analytical gradients were implemented for DLPNO-MP2 enabling the geometry optimizations of larger clusters at the MP2 level. However, to the best of our knowledge, the method has not yet been applied to any cluster studies yet. Potential future development of gradients for the DLPNO-CCSD(T) methods would permit accurate computations of the structures of relatively large clusters.

Current calculations of the cluster vibrational frequencies are usually based on the rigid-rotor harmonic approximations. Especially for weakly bound systems such as water clusters, vibrational anharmonicity might be important (Temelso et al., 2011). However, this is very rarely considered as anharmonic vibrational frequency calculations are extremely computationally expensive.

There also remain small contributions to the electronic energy calculations that might slightly influence the cluster binding energies. These include correlation beyond the perturbative triple (T) correction, core correlation and relativistic effects. However, these effects are unlikely to be important compared to the errors in the thermal contribution or basis set incompleteness/superposition errors.

5.5. Overall Recommendations

There has been a clear progression in the applied quantum chemical methods over the past decade. However, there is still not a universally applied approach that all groups follow. Based on the literature and the arguments in the previous sections, there is no doubt that the PW91, M06-2X and ωB97X-D functionals exhibits the best performance (compared to other density functionals) relative to higher level Coupled Cluster calculations.

As DFT remains the only choice for obtaining the cluster geometries, it is recommended that the PW91, M06-2X and/or ωB97X-D functionals are applied for calculating the structures and vibrational frequencies. Ideally, more than one functional should be applied to alleviate and estimate the potential error inherent in each functional.

It is also recommended that the single point energy is corrected with a higher level method such as DLPNO-CCSD(T). For instance, applying the DLPNO-CCSD(T)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p) level of theory has

been shown to be a good compromise between cost and efficiency and the methodology can consistently be applied to relatively large systems. Performing DLPNO-CCSD(T) calculations also alleviates the difference between the functionals used for the geometry optimization.

5.6. Typical Free Energy Values

The cluster formation process involves loss in the number of degrees of freedom as hydrogen bonds are formed. This will inevitably lead to a negative entropy contribution and hence to a positive -$T\Delta S$-value. For atmospheric relevant systems the ΔH-value will be negative and is associated with attractive intermolecular interactions such as H-bonds. To illustrate typical ΔG_{add}-values for atmospheric relevant systems we look further into the following three types of cluster formation reactions recently presented by Schmitz *et al.* (Schmitz and Elm, 2020):

$$\text{acid} + \text{base} \leftrightharpoons (\text{acid})(\text{base}) \tag{R1}$$

$$\text{acid} + \text{acid} \leftrightharpoons (\text{acid})(\text{acid}) \tag{R2}$$

$$\text{acid} + \text{water} \leftrightharpoons (\text{acid})(\text{water}) \tag{R3}$$

Here the acids are: formic-, acetic-, nitric-, sulfuric- and methanesulfonic acid and the bases are: ammonia, methyl-, dimethyl-, trimethyl- and ethylenediamine. All the combinations or the acids and bases were included in the test set leading to 45 addition reactions that illustrate some of the key interactions relevant for atmospheric molecular clusters. These reactions correspond to the formation of cluster dimers only, and they are roughly representative for the ΔE_{add}, ΔH_{add}, ΔS_{add} and ΔG_{add} values that you get when adding a monomer to a larger cluster. Here these reaction are given by equation (36):

$$\Delta G_{\text{add}} \quad = \quad G[\text{dimer}] - \Big(G[\text{monomer},1] + G[\text{monomer},2] \Big) \tag{49}$$

Other thermochemical functions, such as the enthalpy, entropy and their associated thermal contributions can calculated from, by simply exchanging G with either E, H, S, G_{thermal} or H_{thermal}. It should be mentioned here that the thermal contributions also include the zero point vibrational energy.

All the structures were obtained at the ωB97X-D/aug-cc-pVTZ level of theory and the single point energy has been calculated at the DLPNO-CCSD(T_0)-F12/cc-pVTZ-F12 level of theory with a TightPNO criterion. All the data are taken from the Schmitz *et al.* (Schmitz and Elm, 2020) paper. To decompose the influence of the different contributions to the free energy, Figure 10 shows box-and-whisker plots (minimum, first quartile, median, third quartile and maximum) of the ΔE_{add}, ΔH_{add} and ΔG_{add} values calculated for the 45 cluster addition reactions.

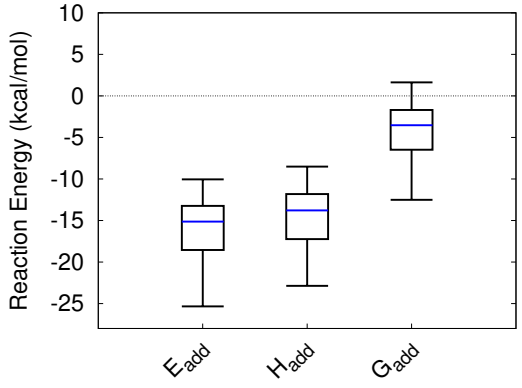

Figure 10: Distribution of the energetics for reactions R1-R3 for a test-set of 45 cluster addition reactions.

The ΔG_{add}-values span a large range from -12.5 kcal/mol (sulfuric acid - dimethylamine) to +1.6 kcal/mol (nitric acid dimer). Naturally, this very large range originates from the fact that the interactions are specific to what molecules actually interact. This large range emerges primarily from the pure electronic energy ΔE_{add} contribution, with only

a minor contribution from the $\Delta H_{\text{add, thermal}}$ term, which lies in the range of +0.9 kcal/mol to +2.7 kcal/mol. The thermal contribution to the free energy ($\Delta G_{\text{add, thermal}}$) also vary very little between the different systems with a range from +10.3 kcal/mol (nitric acid - water) to 14.1 kcal/mol (methanesulfunic acid - ethylenediamine). Again it should be mentioned that the $\Delta G_{\text{add, thermal}}$ term as given by equation (47) includes both the entropy contribution and the $\Delta H_{\text{thermal}}$ contribution. The changes in the $\Delta G_{\text{add, thermal}}$ contribution is primarily caused by the minor variation in the entropy contribution from -29.3 cal/mol·K to -39.8 cal/mol·K. All these values are very sensitive to the applied level of theory and can easily change by up to a few kcal/mol if a different method is used either for optimizing the cluster structures or to calculate the binding energy. This further illustrates that obtaining accurate binding energies are likely more important than the thermal contribution to the free energy, in the (quasi)harmonic oscillator framework. This conclusion is valid under the assumption that the (quasi)harmonic approximations yield reasonable entropies. The fact that different DFT methods tend to give similar entropies support this assumption. Furthermore, this clearly shows that strong intermolecular interactions are most important for the cluster formation process as the entropy penalty is within a similar range for most atmospheric relevant cluster addition reactions. However, it should be noted that for large clusters the rigid rotor harmonic oscillator approximations used in this example will most likely lead to a larger range in the ΔS_{add}-values, especially if the clusters are "fluffy".

6. Configurational Sampling

6.1. Potential/ Free Energy Surface Exploration

Previous sections discuss quantum chemistry calculations conducted on molecular atmospheric clusters. To compute these, the computer must be given a concrete cluster structure (configuration = set of atom nuclei coordinates). For a specific cluster consisting of a set of molecules, infinite configurations can be created by shifting some atoms, and each configuration has an associated potential energy. The entirety of all possible configurations shapes the potential energy surface on which a minima/configuration with lower (free) energy represents a more stable structure. The aim of configurational sampling/exploration processes is to identify the most stable cluster structures (global minimum and low-lying local minima). Local minimum can be found by minimization of the potential energy as a function of atomic coordinates. Generally, it is straightforward to locate the most stable structures for simple molecules consisting of only a few atoms. However, the potential energy functions for many-body systems, such as molecular clusters, become very complicated as numerous possible energy minima appear. Figure 11 shows an example of the potential energy function of configurational space known as the potential energy surface. To simplify, the configurational space represents all possible nuclei positions, which correspond to a multi-dimensional space (overall $3N - 6$ degrees of freedom, where N is number of atoms).

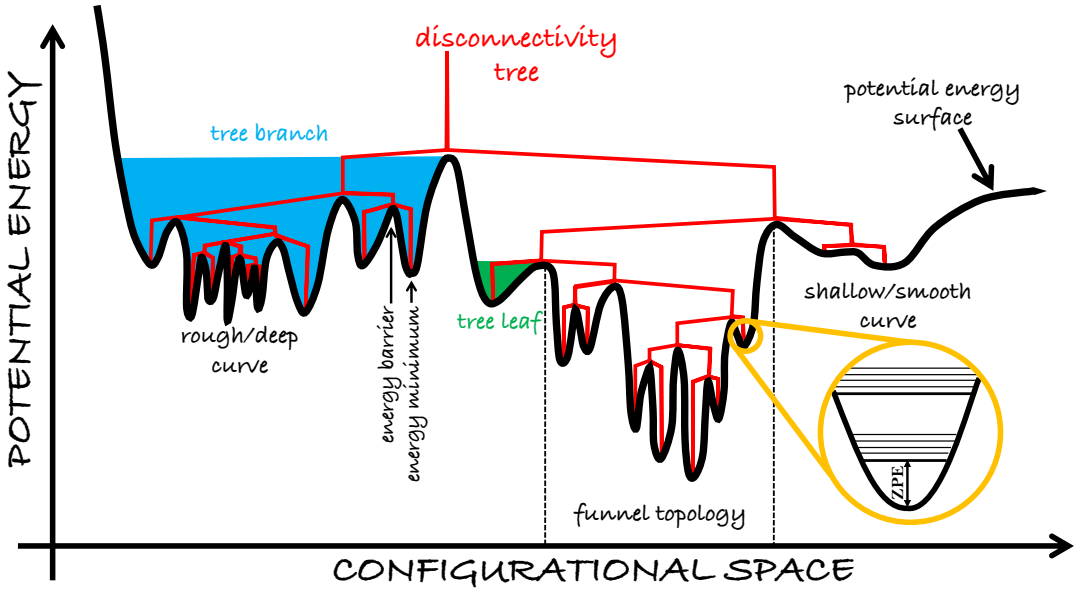

Figure 11: An illustrative potential energy surface of molecular clusters showing a disconnectivity tree and emphasizing some surface properties. (The Figure has been inspired by (Griffiths and Wales, 2019; Kubečka et al., 2019; Plotkin and Onuchic, 2002; Bryngelson et al., 1994; Zhou et al., 2019)).

Figure 11 illustrates several local minima which are connected by a so-called disconnectivity tree (or graph). (Bryngelson et al., 1994) The disconnectivity tree aids in guiding which minima structures are similar (*e.g.*, two structures which are part of the same tree branch or sub-branch) and which minima are very different, *i.e.*, those what are "disconnected" by long distance on the configurational space. To move from a given starting configuration to a distant minima on configurational space, you have to move the atoms significantly or completely reorganize the entire system to reach it. Figure 11 displays diverse depth of energy minima, variously high energy barriers, and different topologies and shapes of the potential energy surface (PES). Combined this complexity makes the configurational search for the global (free) energy minimum of molecular clusters extremely challenging.

Many different approaches have been utilized to sample the configurational space of atomic clusters (Huang et al., 2010; Wales and Doye, 1997; Zhang and Dolg, 2015) or exploring complicated PES in for instance protein folding (Bryngelson et al., 1994; Smellie et al., 2003; Wales, 2018; Plotkin and Onuchic, 2002). Due to the computational cost of quantum chemistry calculations, we recommend the "built-up" approach (Jensen, 2006). This implies that the PES exploration is initially performed at a low level of theory such as molecular-mechanics or semi-empirical methods. Subsequently, the selection of reliable configurations which represent energetically low-lying minima (tree leaves) are re-optimized and characterized at a higher level of theory.

Configurational exploration of small molecular clusters can be performed by simple combinatorics or by forming cluster guesses based on chemical intuition (Herb et al., 2011; Nadykto et al., 2011, 2014). In order to explore all thermodynamically stable minima, several studies have utilized various thermodynamic simulations such as sampling from molecular dynamics and Monte Carlo simulations (Shields et al., 2010; Husar et al., 2012) and its variations such as simulated annealing (Loukonen et al., 2010), umbrella sampling (Loukonen et al., 2010; Xie et al., 2019) and basin hopping (Chen et al., 2017; Peng et al., 2015; Miao et al., 2015; Jiang et al., 2014; Wales and Doye, 1997). However, the disconnectivity tree grows exponentially with cluster size and for systems containing more than 3-4 molecules thermodynamic simulations usually do not have enough time to explore all essential tree branches. Thus, some low-lying minima will likely be overlooked. Moreover, these sampling methods can usually not offer a systematic approach for configurational sampling of various molecular clusters.

Therefore, to maximize the number of initial guesses and cover positions far away from each other in the config-

urational space, several random distribution sampling techniques have been utilized. Elm *et al.* (Elm et al., 2013b,c) randomly place a new molecule around cluster minima of size $N-1$ to obtain the cluster N. Another approach is using a grid of points created around the $N-1$ cluster. Using Fibonacci spheres (González, 2009) the grid points can be evenly distributed around the cluster and a new molecule is placed and reoriented to form the N cluster (Kildgaard et al., 2018a,b). For clusters containing molecules with several rotamers a systematic rotor approach coupled with random sampling can be used for a more thorough exploration of the PES. Combining these two methods is particular handy for molecular clusters containing organic molecules such as pinic acid (Elm et al., 2014). However, even these systematic approaches do not effectively sample the tremendous number of molecular cluster configurations.

Recently, several genetic algorithm programs (OGOLEM (Dieterich and Hartke, 2010), ABCluster (Zhang and Dolg, 2015, 2016) and CLUSTER (Kanters and Donald, 2014)) for configurational sampling of molecular clusters have been introduced. Their primary focus is searching for a global minimum, but they also save all identified local minima during the PES exploration. ABCluster, which uses the Artificial Bee Colony algorithm (Karaboga and Basturk, 2008), has already been utilized in many configurational sampling procedures of atmospherically relevant molecular clusters (Liu et al., 2018b; Hou et al., 2017; Malloum et al., 2019; Chen et al., 2020c; Wang et al., 2019; Myllys et al., 2019a,c). Temelso *et al.* (Temelso et al., 2018) applied a genetic algorithm to generate the initial configurations and narrow down the configurations initially with a semi-empirical method and subsequently using density functional theory. Kubečka *et al.* (Kubečka et al., 2019) have developed the Jammy Key for Configurational Sampling (JKCS) program, which couples the ABCluster program with other 3rd-party quantum chemistry programs. Moreover, they have introduced an approach of including various isomers/protonation states of cluster molecules in the initial guess, which consequently overcome the problem of non-reactive potentials of force-field methods. Recently, Odbadrakh *et al.* (Odbadrakh et al., 2020) proposed a similar protocol based on generating the initial configurations using the OGOLEM program. While these genetic algorithms show promising applications in cluster formation studies involving acids and bases, they rely on treating the molecules as rigid in the global optimization using a force field, which makes the methods difficult to apply to systems containing organic molecules with many rotamers. Recently, Rasmussen *et al.* (Rasmussen et al., 2020)coupled the approach outlined by Kubečka *et al.* (Kubečka et al., 2019) with the systematic hydrate sampling approach by Kildgaard *et al.* (Kildgaard et al., 2018a,b) for studying the shallow potential free energy surface of sulfuric acid - water clusters. This approach appears to be a viable approach to study multicomponent atmospheric molecular clusters involving water.

Inspiration for proper configurational exploration could be obtained from various artificial neural networks or other machine learning methods already applied in several protein folding studies. (Fleetwood et al., 2020; Chen and Ferguson., 2018; Galvelis and Sugita, 2017) However, the main focus of configurational exploration is to provide enough initial guesses that cover the entire PES. Therefore, utilizing computationally cheap molecular mechanics methods in the initial step is assumed to be the most viable option. Subsequently, the molecules have to be optimized at a higher level of theory. Furthermore, we usually search for thermodynamically stable structures, *i.e.*, structure minima not with respect to the lowest potential energy, but to the lowest Gibbs free energy. This implies that the search for the lowest free energy minimum, will depend on the chosen temperature. The potential energy surface depends on the computational method of choice. This implies that different methods might locate a different number of stable structures in the sampling process. Generally, the main bottleneck in configurational sampling remains the selection of important configurations for further calculations at a higher level of theory.

6.2. Uniqueness, Filtering, Selection and Descriptors

Configurational exploration can easily provide an in-comprehensive number of molecular structures. We have to be able to handle, organize, and characterize thousands to millions of them. That is why we need ***descriptors*** that express variables, functions or features which provide a way to construct a unique representation for each cluster structure. A good descriptor should fulfil certain requirements, *e.g.*, invariance with respect to the rotation of the cluster and permutation of atom indices. (Huo and Rupp, 2017; Himanen et al., 2020) The choice of descriptor should always depend on the application at hand. First, we define three processes which can be applied to reduce the number of found configurations:

Uniqueness: Two molecular clusters can be assumed to be the same (one of them is redundant) if a (set of) descriptor(s) for them differ less than some predefined threshold. For example, when two configurations differ in energy less than 0.001 Hartree and, at the same time, differ in dipole moment less than 0.1 Debye, then, they are very likely to be identical and the difference is caused by numerical precision. Another option would be utilizing root mean square

distances (RMSD) (Kildgaard et al., 2018a) of two clusters with ordering implemented according to Temelso *et al.* (Temelso et al., 2017) The disadvantage here is that it is only a relative value between two structures and therefore lacks interpretability. The RMSD has to be minimized in order to allow for a selection of configurations.

Filtering: Configurations could be removed if their descriptors lies out of some predefined range. For example, all configurations which have (free) energy 20 kcal/mol or more higher than the configuration with the lowest energy. Filtering cut-off depends on the step in the configurational sampling procedure. For instance, the cut-off in free energy predicted by a semi-empirical method should be significantly higher than the cut-off in DFT free energy.

Selection: If the number of configurations remains too large after reduction by uniqueness and filtering, some sort of selection procedure has to be applied. One option is to randomly select a subset of conformers, but more effective is selection of a representative set based on various descriptors.

Above processes utilize molecular cluster descriptors, which represent each configuration. Inspired by polymer science Kubečka *et al.* (Kubečka et al., 2019) used the collective variable radius of gyration (Kubečka et al., 2019) as a descriptor for molecular clusters. The radius of gyration is a measure for cluster size. Moreover, they use dipole moments and the electronic energy. The conformers are narrowed down at the semi-empirical level of theory, which bring the difficulty that these can vary greatly depending on the choice of method and even implementations can differ between programs. To use descriptors based just on Cartesian coordinates, Yao *et al.* (Yao et al., 2018) use in their work neural networks.

The recent progressions in the field of Machine Learning have also accelerated the development of different descriptors, which has resulted in a large and diverse selection. Some of them have been implemented as libraries in Python such as Quantum Machine Learning (QML) (Christensen et al., 2017) and DScribe (Himanen et al., 2020). The descriptors range from fingerprint-style (Cereto-Massagué et al., 2015) to those that represent the whole geometry of the structure (Huo and Rupp, 2017), with the latter being the preference when modelling non-covalently bound cluster PES. Geometry representing descriptors include Coulomb Matrix (Rupp et al., 2012; Montavon et al., 2015) , Bag of Bonds (Hansen et al., 2015), Many-Body Tensor Representation (Huo and Rupp, 2017), Atom-Centered Symmetry Functions (Behler, 2011), or Smooth Overlap of Atomic Positions (Bartók et al., 2013; Caro, 2019). Currently, we have not found any reports on testing these descriptors on atmospheric clusters specifically, but they, among others, are widely used in *e.g.* biomaterials research. Jäger *et al.* (Jäger et al., 2018) has applied Coulomb matrix, MBTR, ACSF, and SOAP on hydrogen absorbed onto metallic nanoclusters. Stuke *et al.* (Stuke et al., 2019) used Coulomb matrix and MBTR in the training of a kernel ridge regression model for predicting molecular orbital energies. It is likely that some of these descriptors will be utilized for atmospheric molecular clusters in the future.

The purpose of well-chosen descriptors is to enable the use of algorithms that allow for a sophisticated selection of molecular structures to calculate. In the past, this was commonly performed via a manual selection of certain "best" energies. Kubečka *et al.* (Kubečka et al., 2019) used a density algorithm on energies, radius of gyration, and dipole moments (Chaudhuri, 1994). In data science, this procedure is known as clustering. Clustering can be used to address the aforementioned bottleneck in configurational sampling as a more sophisticated filtering method (removal of redundant structures) and can be applied in different fields (Lo et al., 2018). K-means clustering is already used in medical science/biochemistry for gene expression profiling (Shai et al., 2003), molecular dynamics data (Wolf and Kirschner, 2013), or drug design (Lu et al., 2016). Hierarchical clustering has been been applied in studies protein folding and design (Boczko and Brooks, 1995; Troyer and Cohen, 1995; Gorham et al., 2011).

6.3. The Effect of Conformers on the Free Energies

The occurrence of local free energy minima affects the calculation cluster free energies. A common, but erroneous approach for accounting higher energy local minima when calculating free energies has been the application of Boltzmann averaging. This will lead to an increase in the cluster Gibbs free energy when accounting for multiple conformers and implies that the cluster becomes less stable which corresponds to a lowering in the number of available microstates in the system. The existence of multiple conformers should always lead to an increase in the number of available microstates and hence a lowering of the free energy. To correctly account for the effect that multiple conformers has on

the free energy, the following relation should be used (Partanen et al., 2016b,a; Ho et al., 2016):

$$\Delta G_{\text{multi-conf}} = -RT \ln \left[\sum_k \exp\left(\frac{-\Delta G_k}{RT} \right) \right] \tag{50}$$

This will correctly lead to a lowering of the free energy as more conformers k are taken into account. The effect of including higher free energy conformers has been shown to be relatively small on the order of -1 kcal/mol or lower for sulfuric acid - ammonia and sulfuric acid - pinic acid clusters (Partanen et al., 2016b), sulfuric acid - guanidine clusters (Kubečka et al., 2019) and sulfuric acid - water clusters (Rasmussen et al., 2020). It has also been illustrated that conformers 3 kcal/mol higher in energy compared to the lowest one will not contribute significantly to the free energy and can safely be neglected (Partanen et al., 2016b). This implies that other sources of errors are more pronounced and effort is better spend in improving other calculation parameters compared to performing an exhaustive search of all available conformers. For instance, not finding the global minimum yields a significantly larger error than not considering higher energy local minima. While the effect of including higher energy conformers is small, proper sampling techniques as outlined in section 6.1 will automatically generate a large set of relevant conformers. Albeit being a minor effect, it is worth considering as all the required data are available after sampling the relevant cluster structures.

7. Cluster Systems

In this section we present some of the cluster systems that have been studied in the literature. The purpose is not to comprehensively review all individual cluster formation studies in detail, but to outline the general picture of which compounds that have been studied and might be important in the formation process. We have only included studies that involve strong acids, such as sulfuric acid or methoanesulfonic acid. We have also excluded experimental work that have simply used a few quantum chemical calculations to back up their results.

7.1. The Two Component Sulfuric Acid and Water System

From a historical point of view the two-component sulfuric acid (sa) - water (w) cluster system was the first to be studied in detail. Many research groups have studied $(\text{sa})_a(\text{w})_b$ clusters using quantum chemical methods. Table 1 outlines quantum chemical studies that have involved the formation of sulfuric acid - water clusters, as well as which quantum chemical methods that have been applied.

Table 1: Cluster Formation Studies Involving the Two Component Sulfuric Acid - Water System

Cluster	Reference	Methods
$(\text{sa})_1(\text{w})_1$	(Kurdi and Kochanski, 1989)	SCF-MO-LCGO approach
$(\text{sa})_1(\text{w})_{1-3}$	(Arstila et al., 1998)	LDA, LDA with Becke correction and BLYP, plane wave basis
$(\text{sa})_1(\text{w})_{0-7}$	(Bandy and Ianni, 1998)	B3LYP/6-311++G(2d,2p)
$(\text{sa})_1(\text{w})_{1-5}$	(Re et al., 1999)	B3LYP/D95(d,p), B3LYP/D95++(d,p)
$(\text{sa})_2(\text{w})_{0-6}$	(Ianni and Bandy, 2000)	B3LYP/6-311++G(2d,2p)
$(\text{sa})_2(\text{w})_{0-8}$	(Ding et al., 2003b)	PW91/DNP
$(\text{sa})_1(\text{w})_{6-9}$	(Ding and Laasonen, 2004)	BLYP/DNP, PW91/DNP, RI-MP2/aTZVP
$(\text{sa})_1(\text{w})_{1-3}$	(Natsheh et al., 2004)	PW91/TZP
$(\text{sa})_1(\text{w})_{1-4}$	(Kurtén et al., 2007a)	MP2/aug-cc-pV(T+d)Z//MP2/aug-cc-pV(D+d)Z, including MP4/aug-cc-pV(D+d)Z higher level correlation
$(\text{sa})_{1-2}(\text{w})_{0-5}$	(Loukonen et al., 2010)	MP2/aug-cc-pV(T+d)Z//BLYP/DNP
$(\text{sa})_1(\text{w})_{1-6}$	(Temelso et al., 2012a)	MP2/CBS(4-5 inv)//RI-MP2/aug-cc-pVDZ
$(\text{sa})_2(\text{w})_{1-6}$	(Temelso et al., 2012b)	MP2/CBS(4-5 inv)//MP2/6-31+G(d)
$(\text{sa})_{1-4}(\text{w})_{1-5}$	(Henschel et al., 2014)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(\text{sa})_1(\text{w})_{1-15}$	(Kildgaard et al., 2018a)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(\text{sa})_{2-4}(\text{w})_{1-5}$	(Rasmussen et al., 2020)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)

Abbreviations: sulfuric acid (sa) and water (w)

The earliest studies (Kurdi and Kochanski, 1989; Arstila et al., 1998; Bandy and Ianni, 1998; Re et al., 1999; Natsheh

et al., 2004) only considered a single sulfuric acid molecule clustered with few water molecules and simply focussed on obtaining the cluster structures and binding energetics. Particular emphasis was shown on how many water molecules are required for the first dissociation of sulfuric acid in the clusters. Ding and Laasonen (Ding and Laasonen, 2004) extended the cluster data set up to $(sa)_1(w)_{6-9}$ clusters to investigate whether sulfuric acid could be fully dissociated in the two component cluster system. They found that the fully dissociated clusters were close in energy to the partially dissociated clusters, but in no cases was a fully dissociated cluster found as the global minimum. These findings indicate that fully dissociated clusters could co-exist with their partial dissociated counterparts and as the produced SO_4^{2-} ion in the cluster will be encapsulated in water molecules, this suggested that sulfuric acid molecules at the aerosol surfaces are unlikely to be fully deprotonated. These studies are extremely valuable as they showed the initial proton transfer did not occur in the smallest clusters, indicating that the CNT approach of using the liquid drop model, implicitly assuming that the clusters behave as a bulk solution, is highly inaccurate for the smallest clusters. Using equilibrium distributions that obey the law of mass action Noppel *et al.* (Noppel et al., 2002) derived a scheme to correct the CNT by treating the sulfuric acid molecule as in equilibrium with the surrounding water vapour. Water can then be taken into account by the respective hydrate distribution. This approach has been used to estimate how many water molecules that can be expected to bind to a given cluster (Loukonen et al., 2010; Temelso et al., 2012a,b; Henschel et al., 2014; Rasmussen et al., 2020).

Despite its simplicity, and considerable effort that has been made in studying the sulfuric acid water system, to date there still do not exist any studies that have reported sulfuric acid cluster formation explicitly from individual molecules up to large measurable sizes (∼2 nm in diameter). As seen in Table 1 the largest number of sulfuric acid molecules in the clusters remain four and more than five water molecules have not been considered.

Being the simplest system relevant for atmospheric particle formation the two component system serves as a good benchmark for testing newly developed algorithms and for comparing theoretical approaches with experiments.

7.2. Three Component Systems

Based on both experimental and theoretical evidence it is the general consensus that some stabilizing species is required for sulfuric acid particle formation to occur in the boundary layer. Very early it was established that ammonia, being the most abundant base molecule in the atmosphere, had a significant enhancing effect on sulfuric acid cluster formation. Table 2 summarizes quantum chemical cluster formation studies involving sulfuric acid, bases and water.

Table 2: Cluster Formation Studies Involving the Three Component Sulfuric Acid - Base - Water System

Cluster	Reference	Methods
$(sa)_1(a)_1(w)_{0-5}$	(Ianni and Bandy, 1999)	B3LYP/ 6-311++G(2d,2p)
$(sa)_2(a)_1(w)_1$	(Ianni and Bandy, 1999)	B3LYP/ 6-311++G(2d,2p)
$(sa)_1(a)_1(w)_{0-2}$	(Larson et al., 1999)	B3LYP/6- 311++G(d,p), MP2/6- 311++G(d,p)
$(sa)_{1-2}(a)_1(w)_{0-3}$	(Nadykto and Yu, 2007)	PW91/6-311++G(3df,3pd)
$(sa)_{1-2}(a)_1(w)_{0-7}$	(Kurtén et al., 2007b)	PW91/DNP
$(sa)_{1-3}(a)_1(w)_{0-1}$	(Torpo et al., 2007)	RI-MP2/aug-cc-pV(T+d)Z//MPW1B95/aug-cc-pV(D+d)Z
$(sa)_2(a)_{0-4}$	(Kurtén et al., 2007c)	RI-MP2/CBS(D,T)//RI-MP2/aug-cc-pV(D+d)Z
$(sa)_1(Amine^a)_1$	(Kurtén et al., 2008)	RI-MP2/aug-cc-pV(T+d)Z//MP2/aug-cc-pV(D+d)Z and RI-CC2/aug-cc-pV(T+d)Z//MP2/aug-cc-pV(D+d)Z
$(sa)_{1-2}(a)(w)_{0-5}$	(Loukonen et al., 2010)	RI-MP2/aug-cc-pV(T+d)Z//BLYP/DNP
$(sa)_{1-2}(dma)(w)_{0-5}$	(Loukonen et al., 2010)	RI-MP2/aug-cc-pV(T+d)Z//BLYP/DNP
$(sa)_{1-3}(a)_{0-3}(w)_{0-3}$	(Herb et al., 2011)	PW91/6-311++G(3df,3pd)
$(sa)_{1-3}(ma/dma)_{0-2}(w)_{0-2}$	(Nadykto et al., 2011)	PW91/6-311++G(3df,3pd)
$(sa)_{1-4}(a/dma)_{1-4}$	(Ortega et al., 2012)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-2}(dma/tma)_{1-2}$	(Paasonen et al., 2012)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-3}(a/dma)_{2-3}$	(Kupiainen-Määttä et al., 2012)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-4}(a/dma)_{1-4}$	(Olenius et al., 2013)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_1(ma)(w)_{0-6}$	(Bustos et al., 2014)	MP2/CBS(4-5 inv)//RI-MP2/aug-cc-pVDZ
$(sa)_{1-2}(dma)_{1-2}(w)_{0-5}$	(Nadykto et al., 2014)	PW91/6-311++G(3df,3pd)
$(sa)_{1-3}(a/dma)_{1-2}(w)_{0-5}$	(Henschel et al., 2014)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{2-8}(dma)_{2-8}(w)_{0-10}$	(DePalma et al., 2014)	PW91/6-31++G(d,p)
$(sa)_{1-2}(ma)_{1-2}(w)_{0-4}$	(Nadykto et al., 2015)	PW91/6-311++G(3df,3pd)
$(sa)_{1-3}(a/dma)_{1-3}(w)_{0-4}$	(Henschel et al., 2016)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-4}(mono/diamines)_1$	(Elm et al., 2016a)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//DFT^b/6-31++G(d,p)
$(sa)_{1-2}(dma)_{1-2}$	(Ma et al., 2016)	DF-LMP2-F12/aug-cc-pV(T+d)Z//PW91/6- 311++G(3df,3pd)
$(sa)_{1-2}(amines/hydrazine)_1$	(Li et al., 2016)	B3LYP-D3/aug-cc-pV(T+d)Z
$(sa)_{1-4}(mea)_{1-4}$	(Xie et al., 2017)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(sa)_{1-4}(dma/putrescine)_{1-4}$	(Elm et al., 2017d)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(sa)_{1-4}(a/ma)_{1-4}$	(Elm, 2017)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(sa)_{1-2}(ma/tma)_{1-2}(w)_{0-3}$	(Olenius et al., 2017)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-3}(ma/dma/tma)_{1-4}$	(Temelso et al., 2018)	MP2-F12/VTZ-F12//MP2/aVDZ, DFT^b/6-311++G(d,p)
$(sa)_{1-4}(guanidine)_{1-4}$	(Myllys et al., 2018)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//DFT^b/6-31++G(d,p)
$(sa)_{1-3}(ma)_1(a)_1(w)_{0-4}$	(Wang et al., 2018)	PW91/6-311++G(3df,3pd)
$(sa)_{1-4}(piperazine)_{1-4}$	(Ma et al., 2019a)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(sa)_{1-4}(dma)_{x=0-4}(a)_{4-x}$	(Myllys et al., 2019a)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(sa)_1(mea/dma/tma)_1(w)_{0-4}$	(Ge et al., 2020)	M06-2X/6-311++G(3df, 3pd)
$(sa)_{1-4}(dma)_{x=0-4}(a)_{4-x}$	(Li et al., 2020b)	RI-CC2/aug-cc-pV(T+d)Z//M06-2X/6-311++G(3df,3pd)

[a] Amine = methyl-, ethyl-, dimethyl-, diethyl-, trimethyl-, triethyl- and ethylmethylamine.

[b] DFT = ωB97X-D, PW91 and M06-2X.

Abbreviations: ammonia (a), methylamine (ma), dimethylamine (dma), trimethylamine (tma), monoethanolamine (mea), water (w)

Similar to the two component system, the early studies focussed on the cluster structures and energetics. Torpo *et al.* (Torpo et al., 2007) investigated how ammonia stabilized the initial cluster formation by leading to a further enhancement in the attachment of additional sulfuric acid molecules to the cluster. This lead to the conclusion that ammonia (being a base) could be a key species that enables small sulfuric acid-water clusters to grow into larger sizes.

However, at typic atmospheric conditions and concentrations of sulfuric acid (10^5-10^8 molecules cm^{-3}) and ammonia (up to ppb level) the observed new particle formation rates in different regions worldwide cannot fully be explained by a sulfuic acid - ammonia - water mechanism. The work by Kurtén *et al.* (Kurtén et al., 2008) showed that amines might be an even more potent source to new particle formation compared to ammonia even when considering that amine concentrations (1-10 ppt) most likely are orders magnitudes lower than ammonia. From a purely acid - base chemistry point of view this makes sense as amines are stronger bases, and hence the interaction with sulfuric acid will be stronger.

The main focus in three component cluster formation studies has been the combination of sulfuric acid, water and

simple alkyl amines (methyl-, dimethyl- and trimethylamine) as these are usually present in the highest concentrations in the atmosphere. More than 150 different amines and 30 amino acids have been identified in the atmosphere from both anthropogenic and biogenic sources (Ge et al., 2011). Because of the complexity of quantum chemical calculations, most studies to date still only consider a single type of amine in the cluster. Essentially, the correct formation mechanism might include the combination of many different amines. Recent studies have also established that there exist a synergistic effect between different bases, which implies that not only the basicity of the base and their relative abundance is important, but also the exact molecular structure in the form of the number of available hydrogen bond donors/acceptors has a crucial influence on the ability to form particles. Including different types of bases in the cluster very quickly leads to a large number of different combinations of cluster compositions and the amount of relevant configurations rapidly gets out of hand.

While experimental and theoretical work shows that amines highly enhance sulfuric acid driven new particle formation compared to ammonia, it is still relatively unknown how high amine concentrations actually are in different environments. This is caused by the fact that atmospheric measurements of amine concentrations are extremely difficult to perform and requires very sensitive equipment. Amines might be emitted from very local sources and the concentration in the gas phase is always found to be low. This can very well be linked to their high potential to "stick" to sulfuric acid which might hinder their detection in the gas phase as they more or less instantly partition to the cluster phase.

As demonstrated by Table 2 there exist very few studies that actually consider the synergistic effect of different base molecules in the clusters (Kupiainen-Määttä et al., 2012; Temelso et al., 2018; Myllys et al., 2019a; Li et al., 2020b). Similarly, studies investigating the effect of water on sulfuric acid - base clusters also remain scarce. To obtain realistic datasets, synergistic effects should be further studied in the future, most likely combined with the effect of also considering hydrated clusters. While it will be important to study new sulfuric acid - base systems, at present, it is impossible to consider all possible combinations of relevant bases. This implies that some sort of structure-activity-relation needs to be identified to potentially lump some of the amines together as a single species when considered in cluster distribution models.

7.3. Organic Enhanced Cluster Formation

Zhang *et al.* (Zhang et al., 2004) inferred that organic acids greatly enhanced sulfuric acid particle formation. This has immensely sprouted the interest in studying cluster formation involving sulfuric acid and organic compounds. Unfortunately, this has to some extent also led to a pathological quest for finding the "missing" organic that promotes new particle formation. As volatile organic compounds are photo-oxidized in the atmosphere there exist a plethora of different species with varying amount of oxygen content. This implies that cluster formation studies involving organics suffer from the same general issue as amines in the sense that many different combinations of cluster compositions inevitably will exist. Table 3 outlines some of the quantum chemical studies that have involved the formation of sulfuric acid - organic clusters. For simplicity we do not list the exact compositions of the clusters, only the organic compounds and other precursor molecules which were studied.

Table 3: Cluster Formation Studies Involving Sulfuric Acid and Organics

Cluster	Reference	Methods
(sa), formic-, acetic acid	(Nadykto and Yu, 2007)	PW91/6-311++G(3df,3pd)
(sa), benzoic-, cis-pinoic acid, (w)	(Zhao et al., 2009)	B3LYP/6-311++G(2d,2p)
(sa), oxalic acid	(Xu et al., 2010b)	PW91/6-311+G(3df,3pd)
(sa), benzoic-, maleic-, malic-, pyruvic-, phenylacetic-, and tartaric acid, (a)	(Xu et al., 2010a)	PW91/6-311+G(3df,3pd)
(sa), oxalic-, malonic acid-, maleic-, phthalic- succinic acid, (a), (w)	(Xu and Zhang, 2012)	PW91/6-311++G(3df,3pd)//B3LYP/6-311++G(2d,2p)
(sa), succinic acid, (dma), (w)	(Xu and Zhang, 2013)	PW91/6-311++G(2d,2p)
(sa), glycine, (a), (w)	(Elm et al., 2013c)	M06-2X/6-311++G(3df,3pd)
(sa), acetic acid, (w)	(Zhu et al., 2014)	PW91/6-311++G(3df,3pd)
(sa), pinic acid	(Elm et al., 2014)	M06-2X/6-31+G(d)
(sa), $C_6H_8O_7$	(Elm et al., 2015)	DLPNO-CCSD(T_0)/def2-QZVPP//DFT[a]/6-311++G(3df,3pd)
(sa), oxalic acid, (w)	(Miao et al., 2015)	DF-MP2-F12/cc.pVDZ-F12//PW91/6-311++G(3df,3pd)
$(sa)_4(a)_4$, monoterpene markers	(DePalma et al., 2015)	PW91/6-31++G(d,p)
(sa), monoterpene markers	(Ortega et al., 2016)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
(sa), $C_6H_8O_7$, (a), (dma), (w)	(Elm et al., 2016b)	DLPNO-CCSD(T_0)/def2-QZVPP//DFT[a]/6-31++G(d,p)
(sa), alanine, (w)	(Wang et al., 2016)	M06-2X/6-311++G(3df,3pd)
(sa), glyoxylic-, oxalic-, pyruvic acid	(Zhao et al., 2016)	B3LYP-D3/aug-cc- pVTZ
(sa), tricarboxylic acid (MBTCA)	(Elm et al., 2017c)	DLPNO-CCSD(T_0)/Def2-QZVPP//DFT[a]/6-31++G(d,p)
(sa), mono-, di-, peroxy acids,	(Elm et al., 2017b)	CCSD(T)-F12a/VTZ-F12//DFT/6-31++G(d,p)
(sa), monoterpene markers	(Elm et al., 2017b)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//DFT/6-31++G(d,p)
(sa), lactic acid, (dma), (w)	(Li et al., 2017)	M06-2X/6-311++G(3df,3pd)
(sa), glycolic acid, (a)	(Zhang et al., 2017)	M06-2X/6-311++G(3df,3pd)
(sa), glyoxylic acid, (a)	(Liu et al., 2018a)	M06-2X/6-311++G(3df,3pd)
(sa), malonic acid, (a)	(Zhang et al., 2018a)	M06-2X/6-311++G(3df,3pd)
(sa), serine, threonine, (w)	(Ge et al., 2018)	M06-2X/6-311++G(3df,3pd)
(sa), formic acid, (dma), (w)	(Zhang et al., 2018b)	PW91/6-311++G(3df,3pd)
(sa), aldehydes	(Shi et al., 2018)	M06-2X/6-311++G(3df,3pd)
(sa), amides	(Ma et al., 2019b)	M06-2X/6-311++G(3df,3pd)
(sa), succinic acid, (dma), (a), (w)	(Lin et al., 2019)	PW91/6-311++G(2d,2p)
(sa), cis-pinonic acid (+hydroxylate)	(Shi et al., 2019)	M06-2X/6-311++G(3df,3pd)
(sa), aromatic acids, (a), (w)	(Wang et al., 2019)	PW91/6-311++G(3df,3pd)
(sa), tricarboxylic acid (MBTCA)	(Elm, 2019b)	PW91/6-31++G(d,p)
(sa), glycine, (a)	(Li et al., 2020a)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//M06-2X/6-31++G(d,p)
(sa), trifluoroacetic acid, (dma)	(Lu et al., 2020)	RI-CC2/aug-cc-pV(T+d)Z//M06-2X/6-311++G(3df,3pd)
(sa), (COAs), (a/ma/dma)	(Li et al., 2020c)	PW91/M06-2X with 6-311++G(3df,3pd), G3MP2

[a] DFT = ωB97X-D, PW91 and M06-2X.

MBTCA = 3-methyl-1,2,3-butanetricarboxylic

COA = formic, acetic, oxalic, malonic, succinic, glutaric acid, adipic, benzoic, phenylacetic, pyruvic, maleic acid, malic, tartaric and pinonic acids

Despite having received significant attention, it remains debatable whether or not organic compounds actually participate in the initial steps in new particle formation. This is caused by the observation that organics, such as highly oxygenated autoxidation products, generally bind too weakly to sulfuric acid and do not enhance the attachment of additional sulfuric acid molecules (Elm et al., 2015). This implies that the addition of organics to a cluster consisting of sulfuric acid and bases is not able to compete with the corresponding uptake of another sulfuric acid molecule, even at a high loading of organic compounds (Elm et al., 2016b). The most likely organics to be involved in cluster formation are multicarboxylic acids. Elm *et al.* (Elm et al., 2017b) showed that the direct interaction between sulfuric acid and various carboxylic acids (*i.e* forming two sets of hydrogen bonds) were more or less independent of the carbon backbone of the carboxylic acid with a ΔG-value close to -6 kcal/mol for all acids. Stronger binding was only possible if additional hydrogen bonds were possible between the sulfuric acid and organic acid. A recent study by Li *et al.* (Li et al., 2020a) leads to the same conclusion with a similar value of binding free energy between 14 studied organic acids and sulfuric acid.

A few studies exist that have applied cluster distribution dynamics simulations to study the influence of organics. For instance, it has been indicated that small chained organic acids (lactic-, glycolic-, glyoxylic- and malonic acid) (Li et al., 2017; Zhang et al., 2017; Liu et al., 2018a; Zhang et al., 2018a) can enhance sulfuric acid - base new particle formation rates. However, to archive a significant enhancing effect the temperature must be very low ($\sim$ 220 K). Such low temperatures are only attainable at higher altitudes, which would inevitably lead to lower concentrations of precursor compounds compared to the ground level. Again, it should be noted that given high enough concentration and low enough temperature anything will nucleate, but this gives no information about the processes in the atmosphere. We thus recommend cautious consideration to the simulation conditions when studying the enhancing effect of organics. On the other hand, studies on very large clusters consisting of up to six sulfuric acid molecules and four tricarboxylic acids (MBTCA) molecules (Elm et al., 2017c; Elm, 2019b) have indicated that the organics form particles by themselves instead of interacting with sulfuric acid. However, this might very well be a unique feature of the MBTCA molecule and does not necessarily transfer to other multicarboxylic acids. Experimental evidence suggests that ions are crucial to induce cluster formation involving only organic compounds (Kirkby et al., 2016). However, without complete data sets that include all relevant components (sulfuric acid - bases - organics - water and ions) it remains speculative whether organics actually stabilize the initial cluster formation, or are only involved in the growth of the clusters.

A key reason for the unknown role of organic compounds participating in new particle formation is related to the limited structural information about the relevant compounds. A common approach has been to study compounds of anthropogenic (aromatic acids) and biogenic (isoprene, monoterpene derived acids) origin which are identified in high concentrations either in the gas-phase or particle phase. While to some extend very logical, this approach is flawed in the sense that it remains unknown at what stage a given compound enters the particle. Presently, all known organics, containing only C-, H- and O-atoms, simply bind too weakly to sulfuric acid and do not appreciably enhance the sticking of further sulfuric acid molecules to the cluster in the manner that bases do. So at any realistic concentration and temperature their effect just appears to be too small. At atmospheric relevant conditions it might very well only be a few elusive exotic organic oxidation products that can act as the initial stabilizer of the sulfuric acid clusters.

The potential of a given organic compound to form new particles is assumed to be linked to its saturation vapour pressure. This has sprouted the interest in atmospheric covalently bound dimers as these should have low vapour pressures given their sheer size. For instance, monoterpene ($C_{10}H_{16}$) derived dimers can consist 19-20 carbon atoms and 11-14 oxygen atoms per dimer molecule (Ehn et al., 2014). Such large molecules have many rotational degrees of freedom, which makes quantum chemical studies of particle formation involving covalently bound organic dimers very challenging. The oxygen-to-carbon (O:C) ratio has also widely been used as a metric for the vapour pressure of organic compounds and inferred the plausibility that a given compound could be involved in particle formation. However, it has been explicitly shown that high O:C ratios cannot directly be linked to low saturation vapour pressures (Kurtén et al., 2016), the exact molecular structure *i.e.* available strong binding groups is more important (Elm et al., 2017b).

Despite the considerable attention that organics have received, not a single compound has been identified that is capable of promoting the binding of additional sulfuric acid molecules to the clusters, which is required to enable the growth and to have a substantial effect on new particle formation. The high molecular weight and numerous available rotamers of relevant organic compounds makes it impossible to address the problem head on using a brute force approach. Thus, studies that focus on specific combinations of binding patterns of organic compounds to other cluster precursors might become a more valuable approach in the future. The most likely approach to successfully identify organic compounds that might be involved in particle formation will be a combination of identifying exotic oxidation products from gas-phase kinetic studies that bind strongly to other organics or sulfuric acid.

7.4. Cluster Formation in the Marine Environment

While sulfuric acid is believed to be important for new particle formation in many inland environments, methanesulfonic acid (msa) is believed to play a role over the oceans and in coastal regions. It has been hypothesized that with forthcoming stricter regulation on SO_2 emissions, the relative concentration of methane sulfonic acid compared to sulfuric acid will increase (Perraud et al., 2015). Typical concentrations of methanesulfonic acid is of the same order as, or slightly lower than, that of sulfuric acid (Eisele and Tanner, 1993; Mauldin et al., 1999, 2003; Berresheim et al., 2002; Bardouki et al., 2003; Yokelson et al., 2009). At seashores iodine emissions have also been linked to new particle formation (O'Dowd et al., 2002), with iodic acid being believed to be the main contributor (Sipilä et al., 2016). Table 4 presents some of the literature studies that are relevant to the marine environment.

Table 4: Cluster Formation Studies Involving Methanesulfonic Acid And Iodic Acid

Cluster	Reference	Methods
$(msa)_1(w)_{1-5}$	(Wang, 2007)	B3LYP/6-311+G(2df,p), G3XMP2
$(msa)_1(a)_1(w)_{1-5}$	(Li et al., 2007)	B3LYP/6-311++G(d,p), MP2/6-311++G(d,p)
$(msa)_1(dma/tma)_1(w)_{1-2}$	(Dawson et al., 2012)	RI-MP2/aug-cc-pV(T+d)Z//B3LYP-D/6–31++G(d,p)
$(msa)_x(sa)_y(dma)_{1-2}$, $x+y \leq 3$	(Bork et al., 2014)	M06-2X/6-311++G(3df,3pd)
$(msa)_2(ma/dma/tma)_{1-2}$	(Chen et al., 2016)	B3LYP-D3/6-31++G(d,p) and MP2/aug-cc-pVDZ
(msa), organics [a]	(Zhao et al., 2017)	B3LYP-D3/aug-cc-pV(T+d)Z
(msa), oxalic acid, (ma), (w)	(Xu et al., 2017)	B3LYP-D3/aug-cc-pVDZ
$(msa)_4(ma)_4(w)_{1-12}$	(Xu et al., 2018)	BLYP-D/6-31+G(d)
$(msa)_{1-3}(sa)_{1-3}(w)_{1-3}$	(Wen et al., 2019)	DF-LMP2-F12/PVTZ//PW91/6-311++G(3df,3pd)
$(msa)_{1-4}(mea)_{1-4}$ and $(msa)_{1-2}(mea)_{1-2}(w)_{1-3}$	(Shen et al., 2019)	DLPNO-CCSD(T)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)
$(msa)_{1-4}(a)_{1-4}$	(Chen et al., 2020a)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//M06-2X/6-31++G(d,p)
$(msa)_{1-4}(ma)_{1-4}$ and $(msa)_{1-2}(ma)_{1-2}(w)_{1-4}$	(Chen et al., 2020b)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//M06-2X/6-31++G(d,p)
$(msa)_{1-3}(diethylamine)_{1-3}$	(Xu et al., 2020)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//M06-2X/6-31++G(d,p)
$(msa)_1(a/ma/dma,)_{1-2}$	(Chen et al., 2020d)	B3LYP-D3, M06-2X, PW91, ωB97X-D with aug-cc-pV(T+d)Z
$(ia)_{1-6}(a)_{0-1}$, $(sa)_{1-2}(ia)_{1-4}$	(Rong et al., 2020)	RI-CC2/aug-cc-pVTZ//ωB97X-D/6-311++G(3df,3pd), aug-cc-pVDZ-PP with ECP28MDF for iodine
$(msa)_n(a/ma/tma)_n(w)_n$ $(n = 1, 2)$	(Perraud et al., 2020)	B3LYP-D3/aug-cc-pVDZ

Abbreviations: sulfuric acid (sa), methanesulfonic acid (msa), iodic acid (ia) and water (w).
[a] organics = methanol, formic acid, acetone, dimethylether, formaldehyde and methyl formate.

It is clear that the studies involving methane sulfonic acid is clearly lacking behind compared to sulfuric acid. Only in the recent few years have cluster distribution dynamics simulations involving methane sulfonic acid been reported. In general, most studies indicate that methane sulfonic acid is a significantly worse clustering agent compared to sulfuric acid. Interestingly, the selectivity towards different bases follow a different pattern compared to sulfuric acid. For methane sulfonic acid the stabilizing effect of bases follow MA > TMA ≈ DMA > NH_3 (Chen et al., 2016) as opposed to sulfuric acid that follow a DMA ≈ TMA > MA > NH_3 pattern (Jen et al., 2014; Glasoe et al., 2015).

Most studies have investigated exclusively either methane sulfonic acid or sulfuric acid, but it has been indicated that considering clusters consisting of both molecules lead to a synergetic effect (Bork et al., 2014). Thus more studies with both components and with different bases are required to fully understand the synergistic effects of msa and sa together with various bases. To the best of our knowledge there is only a single study that has looked into the potential of particle formation caused by iodic acid (ia) (Rong et al., 2020). As iodic acid might be important for new particle formation at coastal regions further studies are definitely required.

7.5. Ion Induced Cluster Formation

As demonstrated in the previous sections there exist an abundance of cluster formation studies involving sulfuric/methanesulfonic acid, bases, organics and water. However, the effect of ions has been somewhat less studied and has been a subject of much debate. Studies have indicated that there is a direct link between cloud cover and cosmic rays, implying that ionic pathways might contribute to particle formation and growth processes (Svensmark and FrisisChristensen, 1997; Marsh and Svensmark, 2000; Svensmark et al., 2007; Enghoff and Svensmark, 2008; Svensmark et al., 2009). A recent study showed that condensation of ions could contribute several percent to the growth of particles (Svensmark et al., 2017). Field measurements in the remote boreal forest have indicated that neutral pathways dominate the particle formation mechanism with ions only playing a minor role (Kulmala et al., 2007)]. There is no doubt that when the clusters are otherwise weakly bound and the concentration of precursor species is low, ions will have a large effect on the total particle formation rate (Olenius et al., 2013). Table 5 presents some of the studies that have involved ionic clusters. It should be mentioned that some of the studies listed in Tables 1-4 also address the role of ions and the Table below primarily presents the studies where ions are in focus.

Table 5: Cluster Formation Studies Focussing on Ionic Systems

Cluster	Reference	Methods
$(sa)_{1-4}(a)^{-}_{0-1}$	(Ortega et al., 2008)	RI-CC2/aug-cc-pV(T+d)Z//BLYP/DZP
$(sa)(a)(w)^{-}_{0-5}$	(Nadykto et al., 2008)	PW91/6-311++G(3df,3pd)
$(sa)(w)^{-}_{1-5}$ and $(sa)_1(w)^{+}_{1-6}$	(Nadykto et al., 2009)	PW91/6-311++G(3df,3pd)
$(sa)_{0-2}(a/dma)^{+}_{1-4}$	(Kupiainen-Määttä et al., 2012)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-4}(a)_{0-1}(w)^{-}_{0-3}$	(Herb et al., 2013)	PW91/6-311++G(3df,3pd)
$(sa)_{1-4}(a/dma)^{+/-}_{1-4}$	(Olenius et al., 2013)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-5}(a)^{-}_{0-4}$ and $(sa)_{1-4}(dma)^{-}_{1-4}$	(Ortega et al., 2014)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-4}(a/dma)_{0-2}(w)^{-}_{0-5}$	(Tsona et al., 2015)	RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7
$(sa)_{1-3}(MBTCA)_{1-3}(X)^{+/-}$ and $(sa)_{1-2}(pinic)_{1-2}(X)^{+/-}$	(Myllys et al., 2017)	DLPNO-CCSD(T_0)/def2-QZVPP//DFT[a]/6-31++G(d,p)
$(sa)_{1-4}(a/dma/guanidine)^{+/-}_{1-4}$	(Myllys et al., 2019b)	DLPNO-CCSD(T_0)/aug-cc-pVTZ//ωB97X-D/6-31++G(d,p)

Abbreviations: sulfuric acid (sa) and water (w)
[a] DFT = ωB97X-D, PW91 and M06-2X.
MBTCA = 3-methyl-1,2,3-butanetricarboxylic
X = ammonia, ammonium or bisulfate

As seen from Table 5, there exist very few complete datasets that include both positive and negative ions (Olenius et al., 2013; Myllys et al., 2019b). For weakly bound systems, such as sulfuric acid - ammonia clusters and clusters consisting purely organics, ions are potentially very important for the overall formation rate. This also implies that in the pursuit of a cleaner atmosphere, with reduced emission of anthropogenic SO_2, the relative contribution from ionic pathways might become more important in the future.

8. Outlook

It is clear from the previous sections there already exists a plethora of different cluster studies, and hence there is an emerging need for unified databases for storing thermochemical data to ensure easy public availability (Elm, 2019a). It is also evident that solving all possible combinations of relevant acid - base cluster systems up to 4×4 or larger grid sizes and including the effect of hydration is a formidable task and most likely not a viable option. To progress the field forward we need to be clever about which systems that are relevant to study and improve the applied algorithms used narrow down the massive number of cluster configurations. The application of machine learning techniques appears to be a natural step in the process of screening relevant cluster configurations. There is also both a need for studying cluster systems involving new compounds and at the same time expanding to larger clusters. Screening of small relevant atmospheric systems consisting of combinations of inorganic acids, polyfunctional organic acids, bases, ions and water might be a valid approach towards identifying which systems are capable of binding strong enough to be able to grow to larger sizes. Once again, we stress that for a system to be relevant for new particle formation it not only has to be "stable" with respect to evaporation, but also has to enable the further growth into larger sizes. Such screening procedures would allow for the potential to lump several different chemical species with similar properties as a single entity in cluster distribution dynamics models.

In general, we anticipate that both the quantity and quality of input thermochemistry based on quantum chemical data will improve. This includes both the calculated free energies, and to extension the evaporation rates, but also other input parameters such as the collision rates. For modelling realistic atmospheric conditions there is a need for further quantifying the magnitude of sinks/losses, as shown in Figure 6 the effect of losses can easily be similar to that of the uncertainties in the calculated ΔG-values for low vapour concentrations.

Establishing knowledge of which cluster systems that can be regarded as stable and grow, will make it possible to study reactive uptake of volatile organic species on small clusters/particles. This will allow the possibility to study the early growth behaviour where explicit knowledge is severely lacking. In this regard multi-scale modelling will most likely be useful, where the important part (or "active site") of the large cluster/particle is described with QM methods, while the environment is described at a simpler level of theory.

Obtaining the free energy profile of cluster growth from single molecules to small clusters to ∼2 nm particles will allow the identification of the transition from the cluster regime (discrete QM thermodynamics) to the bulk particle regime (continuum thermodynamics). Promising work in this aspect is underway, for example by connecting sectional

models to cluster distribution dynamics which allows not only for the formation rates to be simulated, but also the full size distribution and growth rate (Kürten et al., 2018; Kürten, 2019; Carlsson et al., 2020). However, it remains difficult to obtain highly accurate thermochemistry from quantum chemical calculations when studying clusters larger than 10 molecules. Thus, new approaches or development of cost-efficient methodologies are required to effectively bridge the gap between discrete and continuum thermodynamics.

Without the accessibility of accurate methods for modelling of large liquid particles COnductor-like Screening MOdels (COSMO*therm* (COSMO)) appears to be a possible intermediate step to obtain bulk thermochemical properties using quantum chemical methods for complex mixtures of compounds where experimental data is lacking (Wang et al., 2015). While restricted to bulk thermodynamics, COSMO-type calculations can predict many bulk properties of complex molecules in multicomponent mixtures. This implies that not only pure compound saturation vapour pressures (p_{sat}) can be calculated, but also Setschenow constants, solubilities and activity coefficients in any solvent environment, including both water and water-insoluble organic matter (Toivola et al., 2017; Kurtén et al., 2018; Hyttinen et al., 2020). However, presently the error margins appears to be quite large.

When studying larger clusters the assumption that the structures remain crystal-like might not be valid anymore. To address this issue would require sophisticated *ab initio* molecular dynamics simulations to adequately describe the cluster, but current methods are too computationally expensive to allow the long simulation time it would take to equilibrate a large cluster. Potentially, molecular dynamics simulations using semi-empirical quantum chemical methods might be a valid option for studying relatively large systems, while maintaining an adequate accuracy.

Recently, Roldin *et al.* (Roldin et al., 2019) coupled a cluster distribution dynamics model with chemical transport model to study the boreal forest aerosol-cloud-climate system. This allowed, for the first time, the explicit inclusion of particle formation based on quantum chemical data in the model instead of simple paramerizations. Hence, there is a general need for more accurate and complete thermochemical dataset based quantum chemistry to further improve the description of particle formation in chemical transport models in various regions.

About this Review

The article is an editor-invited review article. Editor-Invited review articles began in 2020 to commemorate the 50th anniversary of the Journal of Aerosol Science.

Acknowledgements

J. E. thanks the Independent Research Fond Denmark grant number 0964-00001B and the Swedish Research Council Formas project number 2018-01745-COBACCA for financial support. We thank the European Research Council project 692891-Damocles, Academy of Finland, University of Helsinki, Faculty of Science ATMATH project for funding, and CSC - Finnish IT Centre for computational resources. T. K. thanks the Academy of Finland for funding.

References

Almeida, J., Schobesberger, S., Kürten, A., Ortega, I.K., Kupiainen-Määttä, O., Praplan, A.P., Adamov, A., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Donahue, N.M., Downard, A., Dunne, E., Duplissy, J., Ehrhart, S., Flagan, R.C., Franchin, A., Guida, R., Hakala, J., Hansel, A., Heinritzi, M., Henschel, H., Jokinen, T., Junninen, H., Kajos, M., Kangasluoma, J., Keskinen, H., Kupc, A., Kurtén, T., Kvashin, A.N., Laaksonen, A., Lehtipalo, K., Leiminger, M., Leppa, J., Loukonen, V., Makhmutov, V., Mathot, S., McGrath, M.J., Nieminen, T., Olenius, T., Onnela, A., Petäjä, T., Riccobono, F., Riipinen, I., Rissanen, M., Rondo, L., Ruuskanen, T., Santos, F.D., Sarnela, N., Schallhart, S., Schnitzhofer, R., Seinfeld, J.H., Simon, M., Sipilä, M., Stozhkov, Y., Stratmann, F., Tome, A., Troestl, J., Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Virtanen, A., Vrtala, A., Wagner, P.E., Weingartner, E., Wex, H., Williamson, C., Wimmer, D., Ye, P., Yli-Juuti, T., Carslaw, K.S., Kulmala, M., Curtius, J., Baltensperger, U., Worsnop, D.R., Vehkamäki, H., Kirkby, J., 2013. Molecular understanding of sulphuric acid-amine particle nucleation in the atmosphere. Nature 502, 359–363.

Arstila, H., Laasonen, K., Laaksonen, A., 1998. *Ab initio* study of gas-phase sulphuric acid hydrates containing 1 to 3 water molecules. J. Chem. Phys. 108, 1031.

Bandy, A.R., Ianni, J.C., 1998. Study of the hydrates of H_2SO_4 using density functional theory. J. Phys. Chem. A 102, 6533–6539.

Bardouki, H., Berresheim, H., Vrekoussis, M., Sciare, J., Kouvarakis, G., Oikonomou, K., Schneider, J., Mihalopoulos, N., 2003. Gaseous (DMS, MSA, SO_2, H_2SO_4 and DMSO) and particulate (sulfate and methanesulfonate) sulfur species over the northeastern coast of Crete. Atmos. Chem. Phys. 3, 1871–1886.

Bartók, A.P., Kondor, R., Csányi, G., 2013. On representing chemical environments. Phys. Rev. B Condens. Matter 87, 1–16.

Becker, R., Döring, W., 1935. Kinetische behandlung der keimbildung in übersättigten dämpfen. Ann. Phys. 416, 719–752.

Behler, J., 2011. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. J. Chem. Phys. 134.

Berndt, T., Richters, S., Jokinen, T., Hyttinen, N., Kurtén, T., Otkjær, R.V., Kjaergaard, H.G., Stratmann, F., Herrmann, H., Sipilä, M., Kulmala, M., Ehn, M., 2016. Hydroxyl radical-induced formation of highly oxidized organic compounds. Nat. Commun. 7, 13677.

Berresheim, H., Elste, T., Tremmel, H.G., Allen, A.G., Hansson, H.C., Rosman, K., Maso, M.D., Makela, J.M., Kulmala, M., O'Dowd, C.D., 2002. Gas-aerosol relationships of h_2so_4, MSA, and OH: Observations in the coastal marine boundary layer at mace head, ireland. J. Geophys. Res. 107, 1–12.

Bianchi, F., Kurtén, T., Riva, M., Mohr, C., Rissanen, M.P., Roldin, P., Berndt, T., Crounse, J.D., Wennberg, P.O., Mentel, T.F., et al., 2019. Highly oxygenated organic molecules (HOM) from gas-phase autoxidation involving peroxy radicals: A key contributor to atmospheric aerosol. Chem. Rev. 119, 3472–3509.

Bianchi, F., Tröstl, J., Junninen, H., Frege, C., Henne, S., Hoyle, C., Molteni, U., Herrmann, E., Adamov, A., Bukowiecki, N., Chen, X., Duplissy, J., Gysel, M., Hutterli, M., Kangasluoma, J., Kontkanen, J., Kürten, A., Manninen, H.E., Münch, S., Peräkylä, O., Petäjä, T., Rondo, L., Williamson, C., Weingartner, E., Curtius, J., Worsnop, D.R., Kulmala, M., Dommen, J., Baltensperger, U., 2016. New particle formation in the free troposphere: A question of chemistry and timing. Science 352, 1109—-1112.

Boczko, E.M., Brooks, C.L., 1995. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 269, 393–396.

Bork, N., Elm, J., Olenius, T., Vehkamäki, H., 2014. Methane sulfonic acid-enhanced formation of molecular clusters of sulfuric acid and dimethyl amine. Atmos. Chem. Phys. 14, 12023–12030.

Bowman, J.M., 1978. Self-consistent field energies and wavefunctions for coupled oscillators. J. Chem. Phys. 68, 608–610.

Boys, S., Bernardi, F., 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. Mol. Phys. 19, 553–566.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G., 1994. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21, 167–195.

Bustos, D.J., Temelso, B., Shields, G.C., 2014. Hydration of the sulfuric acid - methylamine complex and implications for aerosol formation. J. Phys. Chem. A. 118, 7430–7441.

Carlsson, P.T.M., Celik, S., Becker, D., Olenius, T., Elm, J., Zeuch, T., 2020. Neutral sulfuric acid-water clustering rates: Bridging the gap between molecular simulation and experiment. J. Phys. Chem. Lett. 11, 4239–4244.

Carlsson, P.T.M., Zeuch, T., 2018. Investigation of nucleation kinetics in H_2SO_4 vapor through modeling of gas phase kinetics coupled with particle dynamics. J. Chem. Phys. 148, 104303.

Caro, M.A., 2019. Optimizing many-body atomic descriptors for enhanced computational performance of machine learning based interatomic potentials. Phys. Rev. B 100, 024112.

Cereto-Massagué, A., Ojeda, M.J., Valls, C., Mulero, M., Garcia-Vallvé, S., Pujadas, G., 2015. Molecular fingerprint similarity search in virtual screening. Methods 71, 58–63.

Chai, J., Head-Gordon, M., 2008. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. Phys. Chem. Chem. Phys. 10, 6615–6620.

Charlson, R.J., Lovelockt, J.E., Andreae, M.O.., Warren, S.G., 1987. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. Nature 326, 655–661.

Chaudhuri, B.B., 1994. How to choose a representative subset from a set of data in multi-dimensional space. Pattern Recognit. Lett. 15, 893–899.

Chen, D., Li, D., Wang, C., Liu, F., Wang, W., 2020a. Formation mechanism of methanesulfonic acid and ammonia clusters: A kinetics simulation study. Atmos. Environ. 222, 117161.

Chen, D., Li, D., Wang, C., Luo, Y., Liu, F., Wang, W., 2020b. Atmospheric implications of hydration on the formation of methanesulfonic acid and methylamine clusters: A theoretical study. Chemosphere 244, 125538.

Chen, D., Wang, W., Li, D., Wang, W., 2020c. Atmospheric implication of synergy in methanesulfonic acid–base trimers: a theoretical investigation. RSC Adv. 10, 5173.

Chen, D., Wang, W., Li, D., Wang, W., 2020d. Atmospheric implication of synergy in methanesulfonic acid–base trimers: A theoretical investigation. RSC Adv. 10, 5173–5182.

Chen, H., Varner, M.E., Gerber, R.B., Finlayson-Pitts, B.J., 2016. Reactions of methanesulfonic acid with amines and ammonia as a source of new particles in air. J. Phys. Chem. B 120, 1526–1536.

Chen, J., Jiang, S., Liu, Y., Huang, T., Wang, C., Miao, S., Wang, Z., Zhang, Y., Huang, W., 2017. Interaction of oxalic acid with dimethylamine and its atmospheric implications†. RSC Adv. 7, 6374–6388.

Chen, W., Ferguson., A.L., 2018. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. J. Comp. Chem. 39, 2079–2102.

Christensen, A.S., Faber, F.A., Huang, B., Bratholm, L.A., Tkatchenko, A., Müller, K.R., von Lilienfeld, O.L., 2017. QML: A Python toolkit for quantum machine learning. URL: https://github.com/qmlcode/qml.

Christiansen, O., 2004a. A second quantization formulation of multi-mode dynamics. J. Chem. Phys. 120, 2140–2148.

Christiansen, O., 2004b. Vibrational coupled cluster theory. J. Chem. Phys. 120, 2149–2159.

Christiansen, O., Koch, H., Jørgensen, P., 1995. The second-order approximate coupled cluster singles and doubles model cc2. Chem. Phys. Lett. , 409–418.

COSMO, . COSMOtherm: version C3.0, Release 19, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2019.

Courtney, W.G., 1961. Remarks on homogeneous nucleation. J. Chem. Phys. 35, 2249–2250.

Davis, W.D., 1973. Surface ionization mass spectroscopy of airborne particulates. J. Vac. Sci. Technol 10, 278.

Dawson, M.L., Varner, M.E., Perraud, V., Ezell, M.J., Gerber, R.B., Finlayson-Pitts, B.J., 2012. Simplified mechanism for new particle formation from methanesulfonic acid, amines, and water via experiments and *abinitio* calculations. Proc. Natl. Acad. Sci. U.S.A 109, 18719–18724.

DePalma, J.W., Bzdek, B.R., Doren, D.J., Johnston, M.V., 2012. Structure and energetics of nanometer size clusters of sulfuric acid with ammonia and dimethylamine. J. Phys. Chem. A 116, 1030–1040.

DePalma, J.W., Doren, D.J., Johnston, M.V., 2014. Formation and growth of molecular clusters containing sulfuric acid, water, ammonia, and

dimethylamine. J. Phys. Chem. A 118, 5464–5473.
DePalma, J.W., Wang, J., Wexler, A.S., Johnston, M.V., 2015. Growth of ammonium bisulfate clusters by adsorption of oxygenated organic molecules. J. Phys. Chem. A 119, 11191—-11198.
Dieterich, J.M., Hartke, B., 2010. OGOLEM: Global cluster structure optimisation for arbitrary mixtures of flexible molecules. A multiscaling, object-oriented approach. Mol. Phys. 108, 279–291.
Ding, C., Taskila, T., Laasonen, K., Laaksonen, A., 2003a. Reliable potential for small sulfuric acid–water clusters. Chem. Phys. 287, 7–19.
Ding, G., Laasonen, K., 2004. Partially and fully deprotonated sulfuric acid in $H_2SO_4(H_2O)_n$ ($n = 6-8$) clusters. Chem. Phys. Lett. 390, 307–313.
Ding, G., Laasonen, K., Laaksonen, A., 2003b. Two sulfuric acids in small water clusters. J. Phys. Chem. A 107, 8648–8658.
Ehn, M., Junninen, H., Petäjä, T., Kurtén, T., Kerminen, V., Schobesberger, S., Manninen, H.E., Ortega, I.K., Vehkamäki, H., Kulmala, M., Worsnop, D.R., 2010. Composition and temporal behavior of ambient ions in the boreal forest. Atmos. Chem. Phys. 10, 8513–8530.
Ehn, M., Kleist, E., Junninen, H., Petäjä, T., Lönn, G., Schobesberger, S., Dal, M.M., Trimborn, A., Kulmala, M., Worsnop, D.R., Wahner, A., Wildt, J., Mentel, T.F., 2012. Gas phase formation of extremely oxidized pinene reaction products in chamber and ambient air. Atmos. Chem. Phys. 12, 5113–5127.
Ehn, M., Thornton, J.A., Kleist, E., Sipilä, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I.H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurtén, T., Nielsen, L.B., Jørgensen, S., Kjaergaard, H.G., Canagaratna, M., Maso, M.D., Berndt, T., Petäjä, T., Wahner, A., Kerminen, V.M., Kulmala, M., Worsnop, D.R., Wildt, J., 2014. A large source of low-volatility secondary organic aerosol. Nature 506, 476–479.
Eisele, F.L., Tanner, D.J., 1993. Measurement of the gas-phase concentration of h2so4 and methane sulfonic-acid and estimates of h2so4 production and loss in the atmosphere. J. Geophys. Res. 98, 9001–9010.
Elm, J., 2017. Elucidating the limiting steps in sulfuric acid - base new particle formation. J. Phys. Chem. A 121, 8288–8295.
Elm, J., 2019a. An atmospheric cluster database consisting of sulfuric acid, bases, organics, and water. ACS Omega 4, 10965–10974.
Elm, J., 2019b. Unexpected growth coordinate in large clusters consisting of sulfuric acid and $C_8H_12O_6$ tricarboxylic acid. J. Phys. Chem. A 123, 3170–3175.
Elm, J., Bilde, M., Mikkelsen, K.V., 2012. Assessment of density functional theory in predicting structures and free energies of reaction of atmospheric prenucleation clusters. J. Chem. Theory Comput. 8, 2071–2077.
Elm, J., Bilde, M., Mikkelsen, K.V., 2013a. Assessment of binding energies of atmospheric clusters. Phys. Chem. Chem. Phys. 15, 16442–16445.
Elm, J., Bilde, M., Mikkelsen, K.V., 2013b. Influence of nucleation precursors on the reaction kinetics of methanol with the OH radical. J. Phys. Chem. A 117, 6695–6701.
Elm, J., Fard, M., Bilde, M., Mikkelsen, K.V., 2013c. Interaction of glycine with common atmospheric nucleation precursors. J. Phys. Chem. A 117, 12990–12997.
Elm, J., Jen, C.N., Kurtén, T., Vehkamäki, H., 2016a. Strong hydrogen bonded molecular interactions between atmospheric diamines and sulfuric acid. J. Phys. Chem. A 120, 3693–3700.
Elm, J., Kristensen, K., 2017. Basis set convergence of the binding energies of strongly hydrogen-bonded atmospheric clusters. Phys. Chem. Chem. Phys 19, 1122–1133.
Elm, J., Kurtén, T., Bilde, M., Mikkelsen, K.V., 2014. Molecular interaction of pinic acid with sulfuric acid: Exploring the thermodynamic landscape of cluster growth. J. Phys. Chem. A 118, 7892–7900.
Elm, J., Mikkelsen, K.V., 2014. Computational approaches for efficiently modelling of small atmospheric clusters. Chem. Phys. Lett. 615, 26–29.
Elm, J., Myllys, N., Hyttinen, N., Kurtén, T., 2015. Computational study of the clustering of a cyclohexene autoxidation product $C_6H_8O_7$ with itself and sulfuric acid. J. Phys. Chem. A 119, 8414–8421.
Elm, J., Myllys, N., Kurtén, T., 2017a. Phosphoric acid – a potentially elusive participant in atmospheric new particle formation. Mol. Phys. 115, 2168–2179.
Elm, J., Myllys, N., Kurtén, T., 2017b. What is required for highly oxidized molecules to form clusters with sulfuric acid? J. Phys. Chem. A. 121, 4578–4587.
Elm, J., Myllys, N., Luy, J., Kurtén, T., Vehkamäki, H., 2016b. The effect of water and bases on the clustering of a cyclohexene autoxidation product $C_6H_8O_7$ with sulfuric acid. J. Phys. Chem. A 120, 2240–2249.
Elm, J., Myllys, N., Olenius, T., Halonen, R., Kurtén, T., Vehkamäki, H., 2017c. Formation of atmospheric molecular clusters consisting of sulfuric acid and $C_8H_{12}O_6$ tricarboxylic acid. Phys. Chem. Chem. Phys 19, 4877–4886.
Elm, J., Passananti, M., Kurtén, T., Vehkamäki, H., 2017d. Diamines can initiate new particle formation in the atmosphere. J. Phys. Chem. A 121, 6155–6164.
Enghoff, M.B., Svensmark, H., 2008. The role of atmospheric ions in aerosol nucleation - a review. Atmos. Chem. Phys. 8, 4911–4923.
Falcon-Rodriguez, C.I., Osornio-Vargas, A.R., Sada-Ovalle, I., Segura-Medina, P., 2017. Aeroparticles, composition, and lung diseases. Front. Immunol. 7, 1–9.
Farkas, L., 1927. Keimbildungsgeschwindigkeit in übersättigten dämpfen. Z. Phys. Chem 125, 236–242.
Fleetwood, O., Kasimova, M.A., Westerlund, A.M., Delemotte, L., 2020. Molecular insights from conformational ensembles via machine learning. Biophys. J. 118, 765–780.
Funes-Ardois, I., Paton, R., 2016. Goodvibes: Goodvibes v1.0.1. DOI: http://dx.doi.org/10.5281/zenodo.60811.
Galano, A., Alvarez-Idaboy, J.R., 2006. A new approach to counterpoise correction to bsse. J. Comput. Chem. 27, 1203–1210.
Galvelis, R., Sugita, Y., 2017. Neural network and nearest neighbor algorithms for enhancing sampling of molecular dynamics. J. Chem. Theory Comput. 13, 2489–2500.
Gan, W.Q., FitzGerald, J.M., Carlsten, C., Sadatsafavi, M., Brauer, M., 2013. Associations of ambient air pollution with chronic obstructive pulmonary disease hospitalization and mortality. Am. J. Respir. Crit. Care. Med. 187, 721–727.
Ge, P., Luo, G., Huang, W., Xie, H., Chen, J., Luo, Y., 2020. Theoretical study of the hydration effects on alkylamine and alkanolamine clusters and the atmospheric implication. Chemosphere 243, 125323.

Ge, P., Luo, G., Luo, Y., Huang, W., Xie, H., Chen, J., Qu, J., 2018. Molecular understanding of the interaction of amino acids with sulfuric acid in the presence of water and the atmospheric implication. Chemosphere 210, 215–223.

Ge, X., Wexler, A.S., Clegg, S.L., 2011. Atmospheric amines - Part I. A review. Atmos. Environ. 45, 524—546.

Glasoe, W.A., Volz, K., Panta, B., Freshour, N., Bachman, R., Hanson, D.R., McMurry, P.H., Jen, C., 2015. Sulfuric acid nucleation: An experimental study of the effect of seven bases. J. Geophys. Res. Atmos. 120, 1933–1950.

González, Á., 2009. Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices. Math. Geol. 42, 49–64.

Gorham, R.D., Kieslich, C.A., Dimitrios, D., 2011. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. Ann. Biomed. Eng. 39, 1252–1263.

Griffiths, M., Wales, D.J., 2019. Nested basin-sampling. J. Chem. Theory Comput. 15, 6865–6881.

Grimme, S., 2012. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. Chem. Eur. J. 18, 9955–9964.

Halonen, R., Zapadinsky, E., Kurtén, T., Vehkamäki, H., Reischl, B., 2019. Rate enhancement in collisions of sulfuric acid molecules due to long-range intermolecular forces. Atmos. Chem. Phys. 19, 13355–13366.

Hansen, K., Biegler, F., Ramakrishnan, R., Pronobis, W., von Lilienfeld, O.A., Müller, K.R., Tkatchenko, A., 2015. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. J. Phys. Chem. Lett. 6, 2326–2331.

Hemmilä, M., Makkonen, U., Virkkula, A., Panagiotopoulou, G., Aalto, J., Kulmala, M., Petäjä, T., Hakola, H., Hellén, H., 2020. Amine and guanidine emissions from a boreal forest floor. Atmos. Chem. Phys. Discuss. 2020, 1–20.

Henschel, H., Kurtén, T., Vehkamäki, H., 2016. Computational study on the effect of hydration on new particle formation in the sulfuric acid/ammonia and sulfuric acid/ dimethylamine systems. J. Phys. Chem. A. 120, 1886–1896.

Henschel, H., Navarro, J.C.A., Yli-Juuti, T., Kupiainen-Määttä, O., Olenius, T., Ortega, I.K., Clegg, S.L., Kurtén, T., Riipinen, I., Vehkamäki, H., 2014. Hydration of atmospherically relevant molecular clusters: Computational chemistry and classical thermodynamics. J. Phys. Chem. A. 118, 2599–2611.

Herb, J., Nadykto, A.B., Yu, F., 2011. Large ternary hydrogen-bonded pre-nucleation clusters in the Earth's atmosphere. Chem. Phys. Lett. 518, 7–14.

Herb, J., Xu, Y., Yu, F., Nadykto, A.B., 2013. Large hydrogen-bonded pre-nucleation $(HSO_4^-)(H_2SO_4)_m(H_2O)_k$ and $(HSO_4^-)(NH_3)(H_2SO_4)_m(H_2O)_k$ clusters in the Earth's atmosphere. J. Phys. Chem. A 117, 133–152.

Himanen, L., Jäger, M.O.J., Morooka, E.V., Canova, F.F., Ranawat, Y.S., Gao, D.Z., Rinke, P., Foster, A.S., 2020. DScribe: Library of descriptors for machine learning in materials science. Comput. Phys. Commun. 247, 106949.

Ho, J., Coote, M.L., Chris, C., Truhlar, D.G., 2016. Chapter 6: Theoretical calculation of reduction potentials. CRC Press, Boca Raton, FL, USA. pp. 231—261.

Hogan. Jr, C.J., Fernandez. de. la. Mora, J., 2010. Ion-pair evaporation from ionic liquid clusters. J. Am. Soc. Mass Spectrom. 21, 1382–1386.

Hou, G., Zhang, J., Valiev, M., Wang, X., 2017. Structures and energetics of hydrated deprotonated cis-pinonic acid anion clusters and their atmospheric relevance. Phys. Chem. Chem. Phys. 19, 10676.

Huang, W., Pal, R., Wang, L., Zeng, X.C., Wang, L., 2010. Isomer identification and resolution in small gold clusters. J. Chem. Phys. 132, 054305.

Huo, H., Rupp, M., 2017. Unified representation of molecules and crystals for machine learning. URL: http://arxiv.org/abs/1704.06439.

Husar, D.E., Temelso, B., Ashworth, A.L., Shields, G.C., 2012. Hydration of the bisulfate ion: Atmospheric implications. J. Phys. Chem. A 116, 5151–5163.

Hyttinen, N., Elm, J., Malila, J., Calderón, S.M., Prisle, N.L., 2020. Thermodynamic properties of isoprene and monoterpene derivedorganosulfates estimated with COSMO*therm*. Atmos. Chem. Phys. 20, 5679–5696.

Hyttinen, N., Kupiainen-Määttä, O., Rissanen, M.P., Muuronen, M., Ehn, M., Kurtén, T., 2015. Modeling the charging of highly oxidized cyclohexene ozonolysis products using nitrate-based chemical ionization. J. Phys. Chem. A. 119, 6339–6345.

Hyttinen, N., Rissanen, M.P., Kurtén, T., 2017. Computational comparison of acetate and nitrate chemical ionization of highly oxidized cyclohexene ozonolysis intermediates and products. J. Phys. Chem. A. 121, 2172–2179.

Ianni, J.C., Bandy, A.R., 1999. A density functional theory study of the hydrates of NH3·H2SO4 and its implications for the formation of new atmospheric particles. J. Phys. Chem. A. 103, 2801–2811.

Ianni, J.C., Bandy, A.R., 2000. A theoretical study of the hydrates of $(H_2SO_4)_2$ and its implications for the formation of new atmospheric particles. J. Mol. Struc (THEOCHEM) 497, 19–37.

III, R.D.J., . NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 20, August 2019, Editor: Russell D. Johnson III, http://cccbdb.nist.gov/, DOI:10.18434/T47C7Z.

IPCC, . IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Iyer, S., Lopez-Hilfiker, F., Lee, B.H., Thornton, J.A., Kurtén, T., 2016. Modeling the detection of organic and inorganic compounds using iodide-based chemical ionization. J. Phys. Chem. A. 120, 576–587.

Jäger, M.O.J., Morooka, E.V., Canova, F.F., Himanen, L., Foster, A.S., 2018. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. NPJ Comput. Mater. 4, 37.

Jen, C.N., Bachman, R., Zhao, J., McMurry, P.H., Hanson, D.R., 2016a. Diamine-sulfuric acid reactions are a potent source of new particle formation. Geophys. Res. Lett. 43, 867–873.

Jen, C.N., McMurry, P.H., Hanson, D.R., 2014. Stabilization of sulfuric acid dimers by ammonia, methylamine, dimethylamine, and trimethylamine. J Geophys. Res. Atmos. 119, 7502–7514.

Jen, C.N., Zhao, J., McMurry, P.H., Hanson, D.R., 2016b. Chemical ionization of clusters formed from sulfuric acid and dimethylamine or diamines. Atmos. Chem. Phys. 16, 12513–12529.

Jensen, F., 2006. Introduction to computational chemistry. John Wiley & Sons, Inc., USA.

Jensen, F., 2010. An atomic counterpoise method for estimating inter-and intramolecular basis set superposition errors. J. Chem. Theory Comput.

6, 100–106.

Jiang, S., Liu, Y., Huang, T., Wen, H., Xu, K., Zhao, W., Zhang, W., Huang, W., 2014. Study of $Cl^-(H_2O)_n (n = 1–4)$ using basin-hopping method coupled with density functional theory. J. Comput. Chem. 35, 159–165.

Jokinen, T., Sipilä, M., Junninen, H., Ehn, M., Lönn, G., Hakala, J., Petäjä, T., Mauldin III, R.L., Kulmala, M., Worsnop, D.R., 2012. Atmospheric sulphuric acid and neutral cluster measurements using CI-APi-TOF. Atmos. Chem. Phys. 12, 4117—4125.

Kanters, R.P.F., Donald, K.J., 2014. CLUSTER : Searching for unique low energy minima of structures using a novel implementation of a genetic algorithm. J. Chem. Theory Comput. 10, 5729–5737.

Karaboga, D., Basturk, B., 2008. On the performance of artificial bee colony (ABC) algorithm. Appl. Soft Comput. 8, 687–697.

Kathmann, S.M., Schenter, G.K., Garrett, B.C., 1999. Dynamical nucleation theory: Calculation of condensation rate constants for small water clusters. J. Chem. Phys. 111, 4688–4697.

Kildgaard, J.V., Mikkelsen, K.V., Bilde, M., Elm, J., 2018a. Hydration of atmospheric molecular clusters: A new method for systematic configurational sampling. J. Phys. Chem. A 122, 5026—-5036.

Kildgaard, J.V., Mikkelsen, K.V., Bilde, M., Elm, J., 2018b. Hydration of atmospheric molecular clusters II: Organic acid–water clusters. J. Phys. Chem. A 122, 8549–8556.

Kirkby, J., Curtius, J., Almeida, J., Dunne, E., Duplissy, J., Ehrhart, S., Franchin, A., Gagne, S., Ickes, L., Kürten, A., Kupc, A., Metzger, A., Riccobono, F., Rondo, L., Schobesberger, S., Tsagkogeorgas, G., Wimmer, D., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Downard, A., Ehn, M., Flagan, R.C., Haider, S., Hansel, A., Hauser, D., Jud, W., Junninen, H., Kreissl, F., Kvashin, A., Laaksonen, A., Lehtipalo, K., Lima, J., Lovejoy, E.R., Makhmutov, V., Mathot, S., Mikkilä, J., Minginette, P., Mogo, S., Nieminen, T., Onnela, A., Pereira, P., Petäjä, T., Schnitzhofer, R., Seinfeld, J.H., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Vanhanen, J., Viisanen, Y., Vrtala, A., Wagner, P.E., Walther, H., Weingartner, E., Wex, H., Winkler, P.M., Carslaw, K.S., Worsnop, D.R., Baltensperger, U., Kulmala, M., 2011. Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation. Nature 476, 429–433.

Kirkby, J., Duplissy, J., Sengupta, K., Frege, C., Gordon, H., Williamson, C., Heinritzi, M., Simon, M., Yan, C., Almeida, J., Tröstl, J., Nieminen, T., Ortega, I.K., Wagner, R., Adamov, A., Amorim, A., Bernhammer, A., Bianchi, F., Breitenlechner, M., Brilke, S., Chen, X., Craven, J., Dias, A., Ehrhart, S., Flagan, R.C., Franchin, A., Fuchs, C., Guida, R., Hakala, J., Hoyle, C.R., Jokinen, T., Junninen, H., Kangasluoma, J., Kim, J., Krapf, M., Kürten, A., Laaksonen, A., Lehtipalo, K., Makhmutov, V., Mathot, S., Molteni, U., Onnela, A., Peräkylä, O., Piel, F., Petäjä, T., Praplan, A.P., Pringle, K., Rap, A., Richards, N.A.D., Riipinen, I., Rissanen, M.P., Rondo, L., Sarnela, N., Schobesberger, S., Scott, C.E., Seinfeld, J.H., Sipilä, M., Steiner, G., Stozhkov, Y., Stratmann, F., Tomé, A., Virtanen, A., Vogel, A.L., Wagner, A.C., Wagner, P.E., Weingartner, E., Wimmer, D., Winkler, P.M., Ye, P., Zhang, X., Hansel, A., Dommen, J., Donahue, N.M., Worsnop, D.R., Baltensperger, U., Kulmala, M., Carslaw, K.S., Curtius, J., 2016. Ion-induced nucleation of pure biogenic particles. Nature 533, 521–526.

Kontkanen, J., Lehtipalo, K., Ahonen, L., Kangasluoma, J., Manninen, H.E., Hakala, J., Rose, C., Sellegri, K., Xiao, S., Wang, L., Qi, X., Nie, W., Ding, A., Yu, H., Lee, S., Kerminen, V., Petäjä, T., Kulmala, M., 2017. Measurements of sub-3 nm particles using a particle size magnifier in different environments: From clean mountain top to polluted megacities. Atmos. Chem. Phys. 17, 2163–2187.

Kruse, H., Grimme, S., 2012. A geometrical correction for the inter- and intra-molecular basis set superposition error in hartree-fock and density functional theory calculations for large systems. J. Chem. Phys. 136, 154101:1–16.

Kubečka, J., Besel, V., Kurtén, T., Myllys, N., Vehkamäki, H., 2019. Configurational sampling of noncovalent (atmospheric) molecular clusters: Sulfuric acid and guanidine. J. Phys. Chem. A 123, 6022–6033.

Kulmala, M., 2010. Dynamical atmospheric cluster model. Atmos. Res. 98, 201–206.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H.E., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Järvinen, E., Äijälä, M., Kangasluoma, J., Hakala, J., Aalto, P.P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T., Mauldin, R.L., Duplissy, J., Vehkamäki, H., Bäck, J., Kortelainen, A., Riipinen, I., Kurtén, T., Johnston, M.V., Smith, J.N., Ehn, M., Mentel, T.F., Lehtinen, K.E.J., Laaksonen, A., Kerminen, V., Worsnop, D.R., 2013. Direct observations of atmospheric aerosol nucleation. Science 339, 943–946.

Kulmala, M., Riipinen, I., Sipilä, M., Manninen, H.E., Petäjä, T., Junninen, H., Maso, M.D., Mordas, G., Mirme, A., Vana, M., Hirsikko, A., Laakso, L., Harrison, R.M., Hanson, I., Leung, C., Lehtinen, K.E.J., Kerminen, V., 2007. Toward direct measurement of atmospheric nucleation. Science 318, 89–92.

Kupiainen-Määttä, O., Ortega, I.K., Kurtén, T., Vehkamäki, H., 2012. Amine substitution into sulfuric acid - ammonia clusters. Atmos. Chem. Phys. 12, 3591–3599.

Kurdi, L., Kochanski, E., 1989. Theoretical studies of sulfuric acid monohydrate: Neutral or ionic complex? Chem. Phys. Lett. 158, 111–115.

Kürten, A., 2019. New particle formation from sulfuric acid and ammonia: Nucleation and growth model based on thermodynamics derived from CLOUD measurements for a wide range of conditions. Atmos. Chem. Phys. 19, 5033–5050.

Kürten, A., Jokinen, T., Simon, M., Sipilä, M., Sarnela, N., Junninen, H., Adamov, A., Almeida, J., Amorim, A., Bianchi, F., Breitenlechner, M., Dommen, J., Donahue, N.M., Duplissy, J., Ehrhart, S., Flagan, R.C., Franchin, A., Hakala, J., Hansel, A., Heinritzi, M., Hutterli, M., Kangasluoma, J., Kirkby, J., Laaksonen, A., Lehtipalo, K., Leiminger, M., Makhmutov, V., Mathot, S., Onnela, A., Petäjä, T., Praplan, A.P., Riccobono, F., Rissanen, M.P., Rondo, L., Schobesberger, S., Seinfeld, J.H., Steiner, G., Tomé, A., Tröstl, J., Winkler, P.M., Williamson, C., Wimmer, D., Ye, P., Baltensperger, U., Carslaw, K.S., Kulmala, M., Worsnop, D.R., Curtius, J., 2014. Neutral molecular cluster formation of sulfuric acid-dimethylamine observed in real time under atmospheric conditions. Proc. Natl. Acad. Sci. U.S.A. 111, 15019–15024.

Kürten, A., Li, C., Bianchi, F., Curtius, J., Dias, A., Donahue, N.M., Duplissy, J., Flagan, R.C., Hakala, J., Jokinen, T., Kirkby, J., Kulmala, M., Laaksonen, A., Lehtipalo, K., Makhmutov, V., Onnela, A., Rissanen, M.P., Simon, M., Sipilä, M., Stozhkov, Y., Tröst, J., Ye, P., McMurry, P.H., 2018. New particle formation in the sulfuric acid-dimethylamine-water system: Reevaluation of CLOUD chamber measurements and comparison to an aerosol nucleation and growth model. Atmos. Chem. Phys. 18, 845–863.

Kurtén, T., Hyttinen, N., D'Ambro, E.L., Thornton, J., Prisle, N.L., 2018. Estimating the saturation vapor pressures of isoprene oxidation products $c_5h_{12}o_6$ and $c_5h_{10}o_6$ using COSMO-RS. Atmos. Chem. Phys. 18, 17589–17600.

Kurtén, T., Kuang, C., Gómez, P., McMurry, P.H., Vehkamäki, H., Ortega, I., Noppel, M., Kulmala, M., 2010. The role of cluster energy nonac-

commodation in atmospheric sulfuric acid nucleation. J. Chem. Phys. 132, 024304.
Kurtén, T., Loukonen, V., Vehkamäki, H., Kulmala, M., 2008. Amines are likely to enhance neutral and ion-induced sulfuric acid-water nucleation in the atmosphere more effectively than ammonia. Atmos. Chem. Phys. 8, 4095–4103.
Kurtén, T., Noppel, M., Vehkamäki, H., Salonen, M., Kulmala, M., 2007a. Quantum chemical studies of hydrate formation of H_2SO_4 and HSO_4^-. Boreal Environ. Res. 12, 431–453.
Kurtén, T., Sundberg, M.R., Vehkamäki, H., Noppel, M., Blomquist, J., Kulmala, M., 2006. Ab initio and density functional theory reinvestigation of gas-phase sulfuric acid monohydrate and ammonia hydrogen sulfate. J. Phys. Chem. A 110, 7178–7188.
Kurtén, T., Tiusanen, K., Roldin, P., Rissanen, M.P., Luy, J., Boy, M., Ehn, M., Donahue, N.M., 2016. α-pinene autoxidation products may not have extremely low saturation vapor pressures despite high O:C ratios. J. Phys. Chem. A. 120, 2569–2582.
Kurtén, T., Torpo, L., Ding, G., Vehkamäki, H., Sundberg, M.R., Laasonen, K., Kulmala, M., 2007b. A density functional study on water-sulfuric acid-ammonia clusters and implications for atmospheric cluster formation. J. Geophys. Res. 112, D04210.
Kurtén, T., Torpo, L., Sundberg, M.R., Kerminen, V., Vehkamäki, H., Kulmala, M., 2007c. Estimating the NH_3:H_2SO_4 ratio of nucleating clusters in atmospheric conditions using quantum chemical methods. Atmos. Chem. Phys. 7, 2765–2773.
Larson, L.J., Largent, A., Tao, M., 1999. Structure of the sulfuric acid - ammonia system and the effect of water molecules in the gas phase. J. Phys. Chem. A. 103, 6786–6792.
Lee, B.H., Lopez-Hilfiker, F.D., Mohr, C., Kurtén, T., Worsnop, D.R., Thornton, J.A., 2014. An iodide-adduct high-resolution time-of-flight chemical-ionization mass spectrometer: Application to atmospheric inorganic and organic compounds. Environ. Sci. Technol. 48, 6309–6317.
Lehtipalo, K., Yan, C., Dada, L., Bianchi, F., Xiao, M., Wagner, R., Stolzenburg, D., Ahonen, L.R., Amorim, A., Baccarini, A., Bauer, P.S., Baumgartner, B., Bergen, A., Bernhammer, A., Breitenlechner, M., Brilke, S., Buchholz, A., Mazon, S.B., Chen, D., Chen, X., Dias, A., Dommen, J., Draper, D.C., Duplissy, J., Ehn, M., Finkenzeller, H., Fischer, L., Frege, C., Fuchs, C., Garmash, O., Gordon, H., Hakala, J., He, X., Heikkinen, L., Heinritzi, M., Helm, J.C., Hofbauer, V., Hoyle, C.R., Jokinen, T., Kangasluoma, J., Kerminen, V., Kim, C., Kirkby, J., Kontkanen, J., Kürten, A., Lawler, M.J., Mai, H., Mathot, S., Mauldin, R.L., Molteni, U., Nichman, L., Nie, W., Nieminen, T., Ojdanic, A., Onnela, A., Passananti, M., Petäjä, T., Piel, F., Pospisilova, V., Quéléver, L.L.J., Rissanen, M.P., Rose, C., Sarnela, N., Schallhart, S., Schuchmann, S., Sengupta, K., Simon, M., Sipilä, M., Tauber, C., Tomé, A., Tröstl, J., Väisänen, O., Vogel, A.L., Volkamer, R., Wagner, A.C., Wang, M., Weitz, L., Wimmer, D., Ye, P., Ylisirniö, A., Zha, Q., Carslaw, K.S., Curtius, J., Donahue, N.M., Flagan, R.C., Hansel, A., Riipinen, I., Virtanen, A., Winkler, P.M., Baltensperger, U., Kulmala, M., Worsnop, D.R., 2018. Multicomponent new particle formation from sulfuric acid, ammonia and biogenic vapors. Sci. Adv. 4, 1–9.
Leverentz, H.R., Siepmann, J.I., Truhlar, D.G., Loukonen, V., Vehkamäki, H., 2013. Energetics of atmospherically implicated clusters made of sulfuric acid, ammonia, and dimethyl amine. J. Phys. Chem. A 117, 3819–3825.
Li, D., Chen, D., Liu, F., Wang, W., 2020a. Role of glycine on sulfuric acid-ammonia clusters formation: Transporter or participator. J. Environ. Sci. 89, 125–135.
Li, G.X., Gao, T., Chen, D., Li, Y.X., Zhang, Y.G., Zhu, Z.H., 2006. The splitting of low-lying states for hydroxyl molecule under spin-orbit coupling. Chinese Phys. 15, 998–1003.
Li, H., Kupiainen-Määttä, O., Zhang, H., Zhang, X., Ge, M., 2017. A molecular-scale study on the role of lactic acid in new particle formation: Influence of relative humidity and temperature. Atmos. Environ. 166, 479–487.
Li, H., Ning, A., Zhong, J., Zhang, H., Liu, L., Zhang, Y., Zhang, X., Zeng, X.C., He, H., 2020b. Influence of atmospheric conditions on sulfuric acid-dimethylamine-ammonia-based new particle formation. Chemosphere 245, 125554.
Li, S., Qu, K., Zhao, H., Ding, L., Du, L., 2016. Clustering of amines and hydrazines in atmospheric nucleation. Chem. Phys. 472, 198–207.
Li, S., Zhang, L., Qin, W., Tao, F., 2007. Intermolecular structure and properties of the methanesulfonic acid-ammonia system in small water clusters. Chem. Phys. Lett. 447, 33–38.
Li, Y., Zhang, H., Zhang, Q., Xu, Y., Nadykto, A.B., 2020c. Interactions of sulfuric acid with common atmospheric bases and organic acids: Thermodynamics and implications to new particle formation. J. Environ. Sci. in press.
Lin, Y., Ji, Y., Li, Y., Secrest, J., Xu, W., Xu, F., Wang, Y., An, T., Zhang, R., 2019. Interaction between succinic acid and sulfuric acid–base clusters. Atmos. Chem. Phys. 19, 8003–8019.
Liu, L., Kupiainen-Määttä, O., Zhang, H., Li, H., Zhong, J., Kurtén, T., Vehkamäki, H., Zhang, S., Zhang, Y., Ge, M., Zhang, X., Li, Z., 2018a. Clustering mechanism of oxocarboxylic acids involving hydration reaction: Implications for the atmospheric models. J. Chem. Phys. 148, 214303.
Liu, L., Li, H., Zhang, H., Zhong, J., Bai, Y., Ge, M., Li, Z., Chen, Y., Zhang, X., 2018b. The role of nitric acid in atmospheric new particle formation. Phys. Chem. Chem. Phys. 20, 17406.
Liu, L., Zhong, J., Vehkamäki, H., Kurtén, T., Du, L., Zhang, X., Francisco, J.S., Zeng, X.C., 2019. Unexpected quenching effect on new particle formation from the atmospheric reaction of methanol with SO_3. Proc. Natl. Acad. Sci. U.S.A 116, 24966–24971.
Lo, Y., Rensi, S.E., Torng, W., Altman, R.B., 2018. Machine learning in chemoinformatics and drug discovery. Drug Discov. Today 23, 1538–1546.
Long, B., Tan, X., Wang, Y., Li, J., Ren, D., Zhang, W., 2016. Theoretical studies on reactions of OH with H2SO4 ⋯ NH3 complex and NH2 with H2SO4 in the presence of water. ChemistrySelect 1, 1421–1430.
Loukonen, V., Bork, N., Vehkamäki, H., 2014. From collisions to clusters: First steps of sulphuric acid nanocluster formation dynamics. Mol. Phys. 112, 1979–1986.
Loukonen, V., Kurtén, T., Ortega, I.K., Vehkamäki, H., Pádua, A.A.H., Sellegri, K., Kulmala, M., 2010. Enhancing effect of dimethylamine in sulfuric acid nucleation in the presence of water - A computational study. Atmos. Chem. Phys. 10, 4961–4974.
Lu, J., Chen, L., Yin, J., Huang, T., Bi, Y., Kong, X., Zheng, M., Cai, Y., 2016. Identification of new candidate drugs for lung cancer using chemical–chemical interactions, chemical–protein interactions and a K-means clustering algorithm. J. Biomol. Struct. Dyn. 34, 906–917.
Lu, Y., Liu, L., Ning, A., Yang, G., Liu, Y., Kurtén, T., Vehkamäki, H., Zhang, X., Wang, L., 2020. Atmospheric sulfuric acid-dimethylamine nucleation enhanced by trifluoroacetic acid. Geophys. Res. Lett. 47, e2019GL085627.
Ma, F., Xie, H., Elm, J., Shen, J., Chen, J., Vehkamäki, H., 2019a. Piperazine enhancing sulfuric acid-based new particle formation: Implications for the atmospheric fate of piperazine. Environ. Sci. Technol. , 8785–8795.

Ma, X., Sun, Y., Huang, Z., Zhang, Q., Wang, W., 2019b. A density functional theory study of the molecular interactions between a series of amides and sulfuric acid. Chemosphere 214, 781–790.

Ma, Y., Chen, J., Jiang, S., Liu, Y., Huang, T., Miao, S., Wang, C., Huang, W., 2016. Characterization of the nucleation precursor (H_2SO_4-$(CH_3)_2NH$) complex: Intra-cluster interactions and atmospheric relevance. RSC Adv. 6, 5824–5836.

Malloum, A., Fifen, J.J., Dhaouadi, Z., Engo, S.G.N., Conradie, J., 2019. Structures, relative stability and binding energies of neutral water clusters, $(H_2O)_{2-30}$. New J. Chem. 43, 13020.

Manninen, H.E., Petäjä, T., Asmi, E., Riipinen, I., Nieminen, T., Mikkilä, J., Hörrak, U., Mirme, A., Mirme, S., Laakso, L., Kerminen, V., Kulmala, M., 2009. Long-term field measurements of charged and neutral clusters using neutral cluster and air ion spectrometer (NAIS). Boreal Env. Res. 14, 591–605.

Marcus, R.A., 1952. Unimolecular dissociations and free radical recombination reactions. J. Chem. Phys. 20, 359–364.

Marsh, N.D., Svensmark, H., 2000. Low cloud properties influenced by cosmic rays. Phys. Rev. Lett. 85, 5004–5007.

Mauldin, R.L., Cantrell, C.A., Zondlo, M.A., Kosciuch, E., Ridley, B.A., Weber, R., Eisele, F.E., 2003. Measurements of OH, H_2SO_4, and MSA during tropospheric ozone production about the spring equinox (TOPSE). J. Geophys. Res. 108, 1–18.

Mauldin, R.L., Tanner, D.J., Heath, J.A., Huebert, B.J., Eisele, F.L., 1999. Observations of H_2SO_4 and MSA during PEM-Tropics-A. J. Geophys. Res. 104, 5801–5816.

McGrath, M.J., Olenius, T., Ortega, I.K., Loukonen, V., Paasonen, P., Kurtén, T., Kulmala, M., Vehkamäki, H., 2012. Atmospheric Cluster Dynamics Code: A flexible method for solution of the birth-death equations. Atmos. Chem. Phys. 12, 2345–2355.

McMurry, P.H., 2000. The history of CPCs. Aerosol Sci. Technol. 33, 297–322.

McMurry, P.H., 2000. Review of atmospheric aerosol measurements. Atmos. Environ. 34, 1959–1999.

Mei, M., Song, H., Chen, L., Hu, B., Bai, R., Xu, D., Liu, Y., Zhao, Y., Chen, C., 2018. Early-life exposure to three size-fractionated ultrafine and fine atmospheric particulates in Beijing exacerbates asthma development in mature mice. Part. Fibre Toxicol. 15, 13.

Merikanto, J., Spracklen, D.V., Mann, G.W., Pickering, S.J., Carslaw, K.S., 2009. Impact of nucleation on global CCN. Atmos. Chem. Phys. 9, 8601–8616.

Miao, S., Jiang, S., Chen, J., Ma, Y., Zhu, Y., Wen, Y., Zhang, M., Huang, W., 2015. Hydration of a sulfuric acid-oxalic acid complex: Acid dissociation and its atmospheric implication†. RSC Adv. 5, 48638–48646.

Montavon, G., Hansen, K., Fazli, S., Rupp, M., Biegler, F., Ziehe, A., Tkatchenko, A., von Lilienfeld, O.A., Müller, K., 2015. Learning invariant representations of molecules for atomization energy prediction. J. Phys. Chem. Lett. 6, 2326–2331.

Montgomery(Jr.), J.A., Frisch, M.J., Ochterski, J.W., Petersson, G.A., 1999. A complete basis set model chemistry. vi. use of density functional geometries and frequencies. J. Chem. Phys. 110, 2811–2827.

Myllys, N., Chee, S., Olenius, T., Lawler, M., Smith, J., 2019a. Molecular-level understanding of synergistic effects in sulfuric acid–amine–ammonia mixed clusters. J. Phys. Chem. A 123, 2420–2425.

Myllys, N., Elm, J., Halonen, R., Kurtén, T., Vehkamäki, H., 2016a. Coupled cluster evaluation of the stability of atmospheric acid-base clusters with up to 10 molecules. J. Phys. Chem. A 120, 621–630.

Myllys, N., Elm, J., Kurtén, T., 2016b. Density functional theory basis set convergence of sulfuric acid-containing molecular clusters. Comp. Theor. Chem. 1098, 1–12.

Myllys, N., Kubečka, J., Besel, V., Alfaouri, D., Olenius, T., Smith, J.N., Passananti, M., 2019b. Role of base strength, cluster structure and charge in sulfuric acid-driven particle formation. Atmos. Chem. Phys. 19, 9753—-9768.

Myllys, N., Olenius, T., Kurtén, T., Vehkamäki, H., Riipinen, I., Elm, J., 2017. Effect of bisulfate, ammonia, and ammonium on the clustering of organic acids and sulfuric acid. J. Phys. Chem. A 121, 4812–4824.

Myllys, N., Ponkkonen, T., Chee, S., Smith, J., 2019c. Enhancing potential of trimethylamine oxide on atmospheric particle formation. Atmosphere 11, 35.

Myllys, N., Ponkkonen, T., Passananti, M., Elm, J., Vehkamäki, H., Olenius, T., 2018. Guanidine: A highly efficient stabilizer in atmospheric new-particle formation. J. Phys. Chem. A 122, 4717–4729.

Nadykto, A.B., Herb, J., Yu, F., Xu, Y., 2014. Enhancement in the production of nucleating clusters due to dimethylamine and large uncertainties in the thermochemistry of amine-enhanced nucleation. Chem. Phys. Lett. 609, 42–49.

Nadykto, A.B., Herb, J., Yu, F., Xu, Y., Nazarenko, E.S., 2015. Estimating the lower limit of the impact of amines on nucleation in the Earth's atmosphere. Entropy 17, 2764–2780.

Nadykto, A.B., Yu, F., 2007. Strong hydrogen bonding between atmospheric nucleation precursors and common organics. Chem. Phys. Lett. 435, 14–18.

Nadykto, A.B., Yu, F., Herb, J., 2008. Effect of ammonia on the gas-phase hydration of the common atmospheric ion HSO_4^-. Int. J. Mol. Sci. 9, 2184–2193.

Nadykto, A.B., Yu, F., Herb, J., 2009. Theoretical analysis of the gas-phase hydration of common atmospheric pre-nucleation $(HSO_4^-)(H_2O)_n$ and $(H_3O^+)(H_2SO_4)(H_2O)_n$ cluster ions. Chem. Phys. 360, 67–73.

Nadykto, A.B., Yu, F., Jakovleva, M.V., Herb, J., Xu, Y., 2011. Amines in the Earth's atmosphere: A density functional theory study of the thermochemistry of pre-nucleation clusters. Entropy 13, 554–569.

Natsheh, A.A., Nadykto, A.B., Mikkelsen, K.V., Yu, F., Ruuskanen, J., 2004. Sulfuric acid and sulfuric acid hydrates in the gas phase: A DFT investigation. J. Phys. Chem. A 108, 8914–8929.

Noppel, M., Vehkamäki, H., Kulmala, M., 2002. An improved model for hydrate formation in sulfuric acid-water nucleation. J. Chem. Phys. 116, 218–228.

Noziére, B., Kalberer, M., Claeys, M., Allan, J., abd S. Decesari, B.D., Finessi, E., M, G., Grgić, I., Hamilton, J.F., et al., 2015. The molecular identification of organic compounds in theatmosphere: State of the art and challenges. Chem. Rev. 115, 3919–3983.

Ochterski, J.W., 2000. Thermochemistry in gaussian, https://gaussian.com/thermo/.

Odbadrakh, T.T., Gale, A.G., Ball, B.T., Temelso, B., Shields, G.C., 2020. Computation of atmospheric concentrations of molecular clusters from

ab initio thermochemistry. J. Vis. Exp. 158, e60964.

O'Dowd, C.D., Jimenez, J.L., Bahreini, R., Flagan, R.C., Seinfeld, J.H., Hämeri, K., Pirjola, L., Kulmala, M., Jennings, S.G., Hoffmann, T., 2002. Marine aerosol formation from biogenic iodine emissions. Nature 417, 632–636.

Olenius, T., Halonen, R., Kurtén, T., Henschel, H., Kupiainen-Määttä, O., Ortega, I.K., Jen, C.N., Vehkamäki, H., Riipinen, I., 2017. New particle formation from sulfuric acid and amines: Comparison of mono-, di-, and trimethylamines. J. Geophys. Res. Atmos 122, 7103–7118.

Olenius, T., Kupiainen-Määttä, O., Ortega, I.K., Kurtén, T., Vehkamäki, H., 2013. Free energy barrier in the growth of sulfuric acid-ammonia and sulfuric acid-dimethylamine clusters. J. Chem. Phys. 139, 084312.

Ortega, I.K., Donahue, N.M., Kurtén, T., Kulmala, M., Focsa, C., Vehkamäki, H., 2016. Can highly oxidized organics contribute to atmospheric new particle formation? J. Phys. Chem. A 120, 1452–1458.

Ortega, I.K., Kupiainen-Määttä, O., Kurtén, T., Olenius, T., Wilkman, O., McGrath, M.J., Loukonen, V., Vehkamäki, H., 2012. From quantum chemical formation free energies to evaporation rates. Atmos. Chem. Phys. 12, 225–235.

Ortega, I.K., Kurtén, T., Vehkamäki, H., 2008. The role of ammonia in sulfuric acid ion induced nucleation. Atmos. Chem. Phys. 8, 2859–2867.

Ortega, I.K., Olenius, T., Kupiainen-Määttä, O., Loukonen, V., Kurtén, T., Vehkamäki, H., 2014. Electrical charging changes the composition of sulfuric acid-ammonia/dimethylamine clusters. Atmos. Chem. Phys. 14, 7995–8007.

Paasonen, P., Olenius, T., Kupiainen-Määttä, O., Kurtén, T., Petäjä, T., Birmili, W., Hamed, A., Hu, M., Huey, L.G., Plass-Duelmer, C., Smith, J.N., Wiedensohler, A., Loukonen, V., McGrath, M.J., Ortega, I.K., Laaksonen, A., Vehkamäki, H., Kerminen, V., Kulmala, M., 2012. On the formation of sulphuric acid – amine clusters in varying atmospheric conditions and its influence on atmospheric new particle formation. Atmos. Chem. Phys. 12, 9113–9133.

Partanen, L., Hänninen, V., Halonen, L., 2012. Ab initio structural and vibrational investigation of sulfuric acid. J. Chem. Phys. A 116, 2867–2879.

Partanen, L., Hänninen, V., Halonen, L., 2016a. Effects of global and local anharmonicities on the thermodynamic properties of sulfuric acid monohydrate. J. Chem. Theory Comput. 12, 5511–5524.

Partanen, L., Vehkamäki, H., K. Hansen, . J. Elm, .H.H., Kurtén, T., Halonen, R., Zapadinsky, E., 2016b. Effect of conformers on free energies of atmospheric complexes. J. Phys. Chem. A 120, 8613–8624.

Passananti, M., Zapadinsky, E., Zanca, T., Kangasluoma, J., Myllys, N., Rissanen, M.P., Kurtén, T., Ehn, M., Attouid, M., Vehkamäki, H., 2019. How well can we predict cluster fragmentation inside a mass spectrometer? Chem.Commun. 55, 5946–5949.

Peng, X., Liu, Y., Huang, T., Jiang, S., Huang, W., 2015. Interaction of gas phase oxalic acid with ammonia and its atmospheric implications. Phys. Chem. Chem. Phys. 17, 9552–9563.

Perraud, V., Horne, J.R., Martinez, A.S., Kalinowski, J., Meinardi, S., Dawson, M.L., Wingen, L.M., Dabdub, D., Blake, D.R., Gerber, R.B., Finlayson-Pitts, B.J., 2015. The future of airborne sulfur-containing particles in the absence of fossil fuel sulfur dioxide emissions. Proc. Natl. Acad. Sci. USA 112, 13514–13519.

Perraud, V., Xu, J., Gerber, R.B., Finlayson-Pitts, B.J., 2020. Integrated experimental and theoretical approach to probe the synergistic effect of ammonia in methanesulfonic acid reactions with small alkylamines. Environ. Sci.-Processes Imp. 22, 305–328.

Pfaendtner, J., Yu, X., Broadbelt, L.J., 2007. The 1-D hindered rotor approximation. Theor. Chem. Account. 118, 881–898.

Plotkin, S.S., Onuchic, J.N., 2002. Understanding protein folding with energy landscape theory Part I: Basic concepts. Q. Rev. Biol. 35, 111–167.

Pople, J.A., Schlegel, H.B., Krishnan, R., Defrees, D.J., Binkley, J.S., Frisch, M.J., Whiteside, R.A., Hout, R.F., Hehre, W., 1981. Molecular orbital studies of vibrational frequencies. J.Int. J.Quantum Chem. Symp. 15, 269–278.

Pöschl, U., Shiraiwa, M., 2015. Multiphase chemistry at the atmosphere-biosphere interface influencing climate and public health in the anthropocene. Chem. Rev. 115, 4440–4475.

Rasmussen, F.R., Kubečka, J., Besel, V., Vehkamäki, H., Mikkelsen, K.V., Bilde, M., Elm, J., 2020. Hydration of atmospheric molecular clusters iii: Procedure for efficient free energy surface exploration of large hydrated clusters. J. Phys. Chem. A , 10.1021/acs.jpca.0c02932.

Re, S., Osamura‡, Y., Morokuma, K., 1999. Coexistence of neutral and ion-pair clusters of hydrated sulfuric acid $H_2SO_4(H_2O)_n$ ($n = 1 - 5$) - A molecular orbital study. J. Phys. Chem. A 103, 3535–3547.

Reid, J.P., Bertram, A.K., Topping, D.O., Laskin, A., Martin, S.T., Petters, M.D., Pope, F.D., Rovelli, G., 2018. The viscosity of atmospherically relevant organicparticles. Nat. Commun 9, 956.

Ribeiro, R.F., Marenich, A.V., Cramer, C.J., Truhlar, D.G., 2011. Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation. J. Phys. Chem. B 115, 14556–14562.

Riccobono, F., Schobesberger, S., Scott, C.E., Dommen, J., Ortega, I.K., Rondo, L., Almeida, J., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Downard, A., Dunne, E.M., Duplissy, J., Ehrhart, S., Flagan, R.C., Franchin, A., Hansel, A., Junninen, H., Kajos, M., Keskinen, H., Kupc, A., Kürten, A., Kvashin, A.N., Laaksonen, A., Lehtipalo, K., Makhmutov, V., Mathot, S., Nieminen, T., Onnela, A., Petäjä, T., Praplan, A.P., Santos, F.D., Schallhart, S., Seinfeld, J.H., Sipilä, M., Spracklen, D.V., Stozhkov, Y., Stratmann, F., Tomé, A., Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Vrtala, A., Wagner, P.E., Weingartner, E., Wex, H., Wimmer., D., Carslaw, K.S., Curtius, J., Donahue, N.M., Kirkby, J., Kulmala, M., Worsnop, D.R., Baltensperger, U., 2014. Oxidation products of biogenic emissions contribute to nucleation of atmospheric particles. Science 344, 717–721.

Riva, M., Ehn, M., Li, D., Tomaz, S., Bourgain, F., Perrier, S., George, C., 2019. CI-Orbitrap: An analytical instrument to study atmospheric reactive organic species. Anal. Chem. 91, 9419—9423.

Roldin, P., Ehn, M., Kurtén, T., Olenius, T., Rissanen, M.P., Sarnela, N., Elm, J., Rantala, P., Hao, L., Hyttinen, N., et al., 2019. The role of highly oxygenated organic molecules in the boreal aerosol-cloud-climate system. Nat. Commun. 10, 4370.

Rong, H., Liu, J., Zhang, Y., Du, L., Zhang, X., Li, Z., 2020. Nucleation mechanisms of iodic acid in clean and polluted coastal regions. Chemosphere 253, 126743.

Rose, C., Zha, Q., Dada, L., Yan, C., Lehtipalo, K., Junninen, H., Mazon, S.B., Jokinen, T., Sarnela, N., Sipilä, M., Petäjä, T., Kerminen, V., Bianchi, F., Kulmala, M., 2018. Observations of biogenic ion-induced cluster formation in the atmosphere. Sci. Adv. 4, 1—10.

Rupp, M., Tkatchenko, A., Müller, K., von Lilienfeld, O.A., 2012. Fast and accurate modeling of molecular atomization energies with machine learning. Phys. Rev. Lett. 108, 058301.

Schenter, G.K., Kathmann, S.M., Garrett, B.C., 1999. Dynamical nucleation theory: A new molecular approach to vapor-liquid nucleation. Phys. Rev. Lett. 82, 3484–3487.

Schmitz, G., Elm, J., 2020. Assessment of the DLPNO binding energies of strongly non-covalent bonded atmospheric molecular clusters. ACS Omega 5, 7601–7612.

Schobesberger, S., Junninen, H., Bianchi, F., Lönn, G., Ehn, M., Lehtipalo, K., Dommen, J., Ehrhart, S., Ortega, I.K., Franchin, A., Nieminen, T., Riccobono, F., Hutterli, M., Duplissy, J., Almeida, J., Amorim, A., Breitenlechner, M., Downard, A.J., Dunne, E.M., Flagan, R.C., Kajos, M., Keskinen, H., Kirkby, J., Kupc, A., Kürten, A., Kurtén, T., Laaksonen, A., Mathot, S., Onnela, A., Praplan, A.P., Rondo, L., Santos, F.D., Schallhart, S., Schnitzhofer, R., Sipilä, M., Tomé, A., Tsagkogeorgas, G., Vehkamäki, H., Wimmer, D., Baltensperger, U., Carslaw, K.S., Curtius, J., Hansel, A., Petäjä, T., Kulmalau, M., Donahueand, N.M., Worsnop, D.R., 2013. Molecular understanding of atmospheric particle formation from sulfuric acid and large oxidized organic molecules. Proc. Natl. Acad. Sci. U.S.A. 110, 17223–17228.

Shai, R., Shi, T., Kremen, T.J., Horvath, S., Liau, L.M., Cloughesy, T.F., Mischel, P.S., Nelson, S.F., 2003. Gene expression profiling identifies molecular subtypes of gliomas. Oncogene 22, 4918–4923.

Shen, J., Xie, H., Elm, J., Ma, F., Chen, J., Vehkamäki, H., 2019. Methanesulfonic acid-driven new particle formation enhanced by monoethanolamine: A computational study. Environ. Sci. Technol. 53, 14387–14397.

Shi, X., Zhang, R., Sun, Y., Xu, F., Zhang, Q., Wang, W., 2018. A density functional theory study of aldehydes and their atmospheric products participating in nucleation. Phys. Chem. Chem. Phys. 20, 1005–1011.

Shi, X., Zhao, X., Zhang, R., Xu, F., Cheng, J., Zhang, Q., Wang, W., 2019. Theoretical study of the cis-pinonic acid and its atmospheric hydrolysate participation in the atmospheric nucleation. Sci. Total Environ. 674, 234–241.

Shields, R.M., Temelso, B., Archer, K.A., Morrell, T.E., Shields, G.C., 2010. Accurate predictions of water cluster formation, $(H_2O)_{n=2-10}$. J. Phys. Chem. A 114, 11725–11737.

Shiraiwa, M., Ammann, M., Koop, T., Pöschl, U., 2011. Gas uptake and chemical aging of semisolid organic aerosol particles. Proc. Natl. Acad. Sci. USA 108, 11003–11008.

Shiraiwa, M., Li, Y., Tsimpidi, A.P., Karydis, V.A., Berkemeier, T., Pandis, S.N., Lelieveld, J., Koop, T., Pöschl, U.U., 2017. Global distribution of particle phase statein atmospheric secondary organic aerosols. Nat. Commun 8, 15002.

Sipilä, M., Berndt, T., Petäjä, T., Brus, D., Vanhanen, J., Stratmann, F., Patokoski, J., Mauldin, R.L., Hyvärinen, A., Lihavainen, H., Kulmala, M., 2010. The role of sulfuric acid in atmospheric nucleation. Science 327, 1243–1246.

Sipilä, M., Sarnela, N., Jokinen, T., Henschel, H., Junninen, H., Kontkanen, J., Richters, S., Kangasluoma, J., Franchin, A., Peräkylä, O., Rissanen, M.P., Ehn, M., Vehkamäki, H., Kurtén, T., Berndt, T., Petäjä, T., Worsnop, D., Ceburnis, D., Kerminen, V., Kulmala, M., O'Dowd, C., 2016. Molecular-scale evidence of aerosol particle formation via sequential addition of HIO_3. Nature 537, 532–534.

Sipilä, M., Sarnela, N., Jokinen, T., Junninen, H., Hakala, J., Rissanen, M.P., Praplan, A., Simon, M., Kürten, A., Bianchi, F., Dommen, J., Curtius, J., Petäjä, T., Worsnop, D.R., 2015. Bisulfate - cluster based atmospheric pressure chemical ionization mass spectrometer for high-sensitivity (< 100 ppqV) detection of atmospheric dimethyl amine: proof-of-concept and first ambient data from boreal forest. Atmos. Meas. Tech. 8, 4001–4011.

Smellie, A., Stanton, R., Henne, R., Teig, S., 2003. Conformational analysis by intersection: CONAN. J. Comput. Chem. 24, 10–20.

Smoluchowski, M., 1916. Drei vorträge über diffusion, brownsche molekularbewegung und koagulation von kolloidteilchen. Phys. Z. 17, 557–571.

Stinson, J.L., Kathmann, S.M., Ford, I.J., 2016. A classical reactive potential for molecular clusters of sulphuric acid and water. Mol. Phys. 114, 172–185.

Stuke, A., Todorović, M., Rupp, M., Kunkel, C., Ghosh, K., Himanen, L., Rinke, P., 2019. Chemical diversity in molecular orbital energy predictions with kernel ridge regression. J. Chem. Phys. 150.

Su, T., Bowers, M.T., 1973. Theory of ion-polar molecule collisions. comparison with experimental charge transfer reactions of rare gas ions to geometric isomers of difluorobenzene and dichloroethylene. J. Chem. Phys. 58, 3027–3037.

Su, T., Chesnavich, W.J., 1982. Parametrization of the ion-polarmolecule collision rate constant by trajectory calculations. J. Chem.Phys. 76, 5183–5185.

Svensmark, H., Bondo, T., Svensmark, J., 2009. Cosmic ray decreases affect atmospheric aerosols and clouds. Geophys. Res. Lett. 36, L15101.

Svensmark, H., Enghoff, M.B., Shaviv, N.J., Svensmark, J., 2017. Increased ionization supports growth of aerosols into cloud condensation nuclei. Nat. Commun. 8, 2199.

Svensmark, H., FriisChristensen, E., 1997. Variation of cosmic ray flux and global cloud coverage - a missing link in solar-climate relationships. J. Atmos. Sol.-Terr. Phys 59, 1225–1232.

Svensmark, H., Pedersen, J.O.P., Marsh, N.D., Enghoff, M.B., Uggerhøj, U.I., 2007. Experimental evidence for the role of ions in particle nucleation under atmospheric conditions. Proc. R. Soc. A 463, 385–396.

Temelso, B., Archer, K.A., Shields, G.C., 2011. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. J. Phys. Chem. A 115, 12034–12046.

Temelso, B., Mabey, J.M., Kubota, T., Appiah-Padi, N., Shields, G.C., 2017. ArbAlign: A tool for optimal alignment of arbitrarily ordered isomers using the Kuhn-Munkres algorithm. J. Chem. Inf. Model. 57, 1045–1054.

Temelso, B., Morrell, T.E., Shields, R.M., Allodi, M.A., Wood, E.K., Kirschner, K.N., Castonguay, T.C., Archer, K.A., Shields, G.C., 2012a. Quantum mechanical study of sulfuric acid hydration: Atmospheric implications. J. Phys. Chem. A. 116, 2209–2224.

Temelso, B., Morrison, E.F., Speer, D.L., Cao, B.C., Appiah-Padi, N., Kim, G., Shields, G.C., 2018. Effect of mixing ammonia and alkylamines on sulfate aerosol formation. J. Phys. Chem. A 122, 1612–1622.

Temelso, B., Phan, T.N., Shields, G.C., 2012b. Computational study of the hydration of sulfuric acid dimers: Implications for acid dissociation and aerosol formation. J. Phys. Chem. A. 116, 9745–9758.

Thomas, J.M., He, S., Larriba-Andaluz, C., DePalma, J.W., Johnston, M.V., Hogan. Jr, C.J., 2016. Ion mobility spectrometry-mass spectrometry examination of the structures, stabilities, and extents of hydration of dimethylamine-sulfuric acid clusters. Phys. Chem. Chem. Phys. 18, 22962–22972.

Toivola, M., Prisle, N.L., Elm, J., Waxman, E.M., Volkamer, R., T, K., 2017. Can COSMO*therm* predict a salting in effect? J. Phys. Chem. A 121, 6288–6295.

Torpo, L., Kurtén, T., Vehkamäki, H., Laasonen, K., Sundberg, M.R., Kulmala, M., 2007. Significance of ammonia in growth of atmospheric nanoclusters. J. Phys. Chem. A. 111, 10671–10674.

Tröstl, J., Chuang, W.K., Gordon, H., Heinritzi, M., CYan, Molteni, U., Ahlm, L., Frege, C., Bianchi, F., Wagner, R., et al., 2016. The role of low-volatility organic compounds in initial particle growth in the atmosphere. Nature 533, 527–531.

Troyer, J.M., Cohen, F.E., 1995. Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory. Proteins Struct. Funct. Bioinf. 23, 97–110.

Truhlar, D.G., 1991. A simple approximation for the vibrational partition function of a hindered internal rotation. J Comp Chem 12, 266–270.

Tsona, N.T., Henschel, H., Bork, N., Loukonen, V., Vehkamäki, H., 2015. Structures, hydration, and electrical mobilities of bisulfate ion-sulfuric acid-ammonia/dimethylamine clusters: A computational study. J. Phys. Chem. A 119, 9670–9679.

Vanhanen, J., Mikkilä, J., Lehtipalo, K., Sipilä, M., Manninen, H.E., Siivola, E., Petäjä, T., Kulmala, M., 2011. Particle size magnifier for nano-CN detection. Arosol Sci. Technol. 45, 533–542.

Vehkamäki, H., 2006. Classical nucleation theory in multicomponent systems. Springer, United States.

Virtanen, A., Joutsensaari, J., Koop, T., Kannosto, J., Yli-Pirilä, P., Leskinen, J., Mäkelä, J.M., Holopainen, J.K., Pöschl, Kulmala, M., Worsnop, D.R., Laaksonen, A., 2010. An amorphous solid state of biogenic secondary organic aerosol particles. Nature 467, 824–827.

Wagner, R., Yan, C., Lehtipalo, K., Duplissy, J., Nieminen, T., Kangasluoma, J., Ahonen, L.R., Dada, L., Kontkanen, J., Manninen, H.E., Dias, A., Amorim, A., Bauer, P.S., Bergen, A., Bernhammer, A., Bianchi, F., Brilke, S., Mazon, S.B., Chen, X., Draper, D.C., Fischer, L., Frege, C., Fuchs, C., Garmash, O., Gordon, H., Hakala, J., Heikkinen, L., Heinritzi, M., Hofbauer, V., Hoyle, C.R., Kirkby, J., Kürten, A., Kvashnin, A.N., Laurila, T., Lawler, M.J., Mai, H., Makhmutov, V., Mauldin III, R.L., Molteni, U., Nichman, L., Nie, W., Ojdanic, A., Onnela, A., Piel, F., Quéléver, L.L.J., Rissanen, M.P., Sarnela, N., Schallhart, S., Sengupta, K., Simon, M., Stolzenburg, D., Stozhkov, Y., Tröstl, J., Viisanen, Y., Vogel, A.L., Wagner, A.C., Xiao, M., Ye, P., Baltensperger, U., Curtius, J., Donahue, N.M., Flagan, R.C., Gallagher, M., Hansel, A., Smith, J.N., Tomé, A., Winkler, P.M., Worsnop, D., Ehn, M., Sipilä, M., Kerminen, V., Petäjä, T., Kulmala, M., 2017. The role of ions in new particle formation in the CLOUD chamber. Atmos. Chem. Phys. 17, 15181–15197.

Wales, D.J., 2018. Exploring energy landscapes. Annu. Rev. Phys. 69, 401–425.

Wales, D.J., Doye, J.P.K., 1997. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. A 101, 5111–5116.

Waller, S.E., Yang, Y., Castracane, E., Racow, E.E., Kreinbihl, J.J., Nickson, K.A., Johnson, C.J., 2018. The interplay between hydrogen bonding and coulombic forces in determining the structure of sulfuric acid-amine clusters. J. Phys. Chem. Lett. 9, 1216–1222.

Wang, C., Goss, K.U., Lei, Y.D., Abbatt, J.P.D., Wania, F., 2015. Calculating equilibrium phase distribution during the formation ofsecondary organic aerosol using cosmotherm. Environ. Sci. Technol. 49, 8585–8594.

Wang, C., Jiang, S., Liu, Y., Wen, H., Wang, Z., Han, Y., Huang, T., Huang, W., 2018. Synergistic effect of ammonia and methylamine on nucleation in the Earth's atmosphere. a theoretical study. J. Phys. Chem. A 122, 3470–3479.

Wang, C., Ma, Y., Chen, J., Jiang, S., Liu, Y., Wen, H., Feng, Y., Hong, Y., Huang, T., Huang, W., 2016. Bidirectional interaction of alanine with sulfuric acid in the presence of water and the atmospheric implication. J. Phys. Chem. A 120, 2357–2371.

Wang, H., Zhao, X., Zuo, C., Ma, X., F. Xu, Y.S., Zhang, Q., 2019. A molecular understanding of the interaction of typical aromatic acids with common aerosol nucleation precursors and their atmospheric implications. RSC Adv. 9, 36171–36181.

Wang, L., 2007. Clusters of hydrated methane sulfonic acid $CH_3SO_3H{\cdot}(H_2O)_n (n = 1 - 5)$: A theoretical study. J. Phys. Chem. A 111, 3642–3651.

Wen, H., Wang, C., Wang, Z., Hou, X., Han, Y., Liu, Y., Jiang, S., Huang, T., Huang, W., 2019. Formation of atmospheric molecular clusters consisting of methanesulfonic acid and sulfuric acid: Insights from flow tube experiments and cluster dynamics simulations. Atmos. Environ. 199, 380—390.

WHO, . WHO (World Health Organization), *Public health, environmental and social determinants of health*, 2014.

Wilemski, G., Wyslouzil, B.E., 1995. Binary nucleation kinetics. I. Self-consistent size distribution. J. Chem. Phys. 103, 1127–1136.

Willetts, A., Handy, N.C., Green, W.H., Jayatilaka, D., 1990. Anharmonic corrections to vibrational transition intensities. J. Phys. Chem. 94, 5608–5616.

Wolf, A., Kirschner, K.N., 2013. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. J. Mol. Model. 19, 539–549.

Wyslouzil, B.E., Wilemski, G., 1995. Binary nucleation kinetics. II. Numerical solution of the birth–deathequations. J. Chem. Phys. 103, 1137–1151.

Xie, H., Elm, J., Halonen, R., Myllys, N., Kurtén, T., Kulmala, M., Vehkamäki, H., 2017. The atmospheric fate of monoethanolamine: Enhancing new-particle formation of sulfuric acid as an important removal process. Environ. Sci. Technol. 51, 8422–8431.

Xie, L., Cheng, H., Fang, D., Chen, Z., Yang, M., 2019. Enhanced QM/MM sampling for free energy calculation of chemical reactions: A case study of double proton transfer. J. Chem. Phys. 150, 044111.

Xu, C., Jiang, S., Liu, Y., Feng, Y., Wang, Z., Huang, T., Zhao, Y., Li, J., Huang, W., 2020. Formation of atmospheric molecular clusters of methanesulfonic acid–diethylamine complex and its atmospheric significance. Atmos. Environ. 226, 117404.

Xu, J., Finlayson-Pitts, B.J., Gerber, R.B., 2017. Proton transfer in mixed clusters of methanesulfonic acid, methylamine, and oxalic acid: Implications for atmospheric particle formation. J. Phys. Chem. A 121, 2377–2385.

Xu, J., Perraud, V., Finlayson-Pitts, B.J., Gerber, R.B., 2018. Uptake of water by an acid–base nanoparticle: Theoretical and experimental studies of the methanesulfonic acid–methylamine system. Phys. Chem. Chem. Phys. 20, 22249–22259.

Xu, W., Zhang, R., 2012. Theoretical investigation of interaction of dicarboxylic acids with common aerosol nucleation precursors. J. Phys. Chem. A. 116, 4539–4550.

Xu, W., Zhang, R., 2013. A theoretical study of hydrated molecular clusters of amines and dicarboxylic acids. J. Chem. Phys. 139, 064312.

Xu, Y., Nadykto, A.B., Yu, F., Herb, J., Wang, W., 2010a. Interaction between common organic acids and trace nucleation species in the Earth's atmosphere. J. Phys. Chem. A 114, 387–396.

Xu, Y., Nadykto, A.B., Yu, F., Jiang, L., Wang, W., 2010b. Formation and properties of hydrogen-bonded complexes of common organic oxalic acid with atmospheric nucleation precursors. J. Mol. Struc-THEOCHEM 951, 28–33.

Yang, Y., Waller, S.E., Kreinbihl, J.J., Johnson, C.J., 2018. Direct link between structure and hydration in ammonium and aminium bisulfate clusters implicated in atmospheric new particle formation. J. Phys. Chem. Lett. 9, 5647–5652.

Yao, K., Herr, J.E., Toth, D.W., Mckintyre, R., Parkhill, J., 2018. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. Chem. Sci. 9, 2261–2269.

Yokelson, R.J., Crounse, J.D., DeCarlo, P.F., T. Karl, S.U., Atlas, E., Campos, T., Shinozuka, Y., Kapustin, V., Clarke, A.D., Weinheimer, A., Knapp, D.J., Montzka, D.D., Holloway, J., Weibring, P., Flocke, F., Zheng, W., Toohey, D., Wennberg, P.O., Wiedinmyer, C., Mauldin, L., Fried, A., Richter, D., Walega, J., Jimenez, J.L., Adachi, K., Buseck, P.R., Hall, S.R., Shetter, R., 2009. Emissions from biomass burning in the yucatan. Atmos. Chem. Phys. 9, 5785–5812.

Yu, F., 2006a. Effect of ammonia on new particle formation: A kinetic H_2SO_4-H_2O-NH_3 nucleation model constrained by laboratory measurements. J. Geophys. Res 111, D01204.

Yu, F., 2006b. From molecular clusters to nanoparticles: Second-generationion-mediated nucleation model. Atmos. Chem. Phys. 6, 5193–5211.

Yu, F., Nadykto, A.B., Herb, J., Luo, G., Nazarenko, K.M., Uvarova, L.A., 2018. H2SO4-H2O-NH3 ternary ion-mediated nucleation (TIMN): Kinetic-based model and comparison with CLOUD measurements. Atmos. Chem. Phys. 18, 17451–17474.

Yue, G.K., Chan, L.Y., 1979. Theory of the formation of aerosols of volatile binary solutions through the ion-induced nucleation process. J. Colloid Interface Sci 68, 501–507.

Zapadinsky, E., Passananti, M., Myllys, N., Kurtén, T., Vehkamäki, H., 2019. Modeling on fragmentation of clusters inside a mass spectrometer. J. Phys. Chem. A 123, 611–624.

Zhang, H., Kupiainen-Määttä, O., Zhang, X., Molinero, V., Zhang, Y., Li, Z., 2017. The enhancement mechanism of glycolic acid on the formation of atmospheric sulfuric acid - ammonia molecular clusters. J. Chem. Phys. 146, 184308.

Zhang, H., Li, H., Liu, L., Zhang, Y., Zhang, X., Li, Z., 2018a. The potential role of malonic acid in the atmospheric sulfuric acid - ammonia clusters formation. Chemosphere 203, 26–33.

Zhang, J., Dolg, M., 2015. ABCluster: the artificial bee colony algorithm for cluster global optimization. Phys. Chem. Chem. Phys. 17, 24173–24181.

Zhang, J., Dolg, M., 2016. Global optimization of clusters of rigid molecules using the artificial bee colony algorithm. Phys. Chem. Chem. Phys. 18, 3003–3010.

Zhang, R., Jiang, S., Liu, Y., Wen, H., Feng, Y., Huang, T., Huang, W., 2018b. An investigation about the structures, thermodynamics and kinetics of the formic acid involved molecular clusters. Chem. Phys. 507, 44–50.

Zhang, R., Suh, I., Zhao, J., Zhang, D., Fortner, E.C., Tie, X., Molina, L.T., Molina, M.J., 2004. Atmospheric new particle formation enhanced by organic acids. Science 304, 1487–1490.

Zhao, H., Jiang, X., Du, L., 2017. Contribution of methane sulfonic acid to new particle formation in the atmosphere. Chemosphere 174, 689–699.

Zhao, H., Zhang, Q., Du, L., 2016. Hydrogen bonding in cyclic complexes of carboxylic acid–sulfuric acid and their atmospheric implications. RSC Adv. 6, 71733–71743.

Zhao, J., Khalizov, A., Zhang, R., McGraw, R., 2009. Hydrogen-bonding interaction in molecular complexes and clusters of aerosol nucleation precursors. J. Phys. Chem. A. 113, 680–689.

Zhou, C., Ieritano, C., Hopkins, W.S., 2019. Augmenting basin-hopping with techniques from unsupervised machine learning: Applications in spectroscopy and ion mobility. Front. Chem. 7, 519.

Zhu, Y., Liu, Y., Huang, T., Jiang, S., Xu, K., Wen, H., Zhang, W., Huang, W., 2014. Theoretical study of the hydration of atmospheric nucleation-precursors with acetic acid. J. Phys. Chem. A 118, 7959–7974.

Jonas Elm received a master's degree in nanoscience from the University of Copenhagen (2011) and a PhD degree in theoretical chemistry from the University of Copenhagen (2014) working with computational modelling of atmospheric molecular clusters under the supervision of Prof. Kurt V. Mikkelsen. Presently, he is employed as an assistant professor at Aarhus University leading the Computational Atmospheric Chemistry group. In his current work he is generally striving towards developing a unified atmospheric new particle formation model using a combination of quantum chemical calculations and machine learning. The overall goal is to effectively bridge the persisting gap between theory and experiments.

Jakub Kubečka completed his Bc thesis and Ing diploma in Physical and Analytical Chemistry at the University of Chemical Technology, Prague, and parallelly Bc thesis and Mgr diploma in Physical Chemistry at the Charles University in Prague. He is currently a Ph.D. student in computational aerosol physics in the Division of Atmospheric Sciences, Department of Physics at the University of Helsinki. His interests are computer calculations with main focus on theoretical and computational quantum chemistry. He has been studying configurational sampling and stability of molecular clusters in the context of atmospheric new-particle formation.

Vitus Besel obtained a Bachelor of Chemistry at the Ludwig-Maximilians University Munich and completed a Master's degree of Theoretical and Computational Methods at the University of Helsinki. Over the course of his Master studies he worked for the computational aerosol physics group, under the supervision of Professor Hanna Vehkamäki, studying configurational sampling of atmospheric molecular clusters and cluster distribution dynamics. He will continue working in the same research group as a doctoral student starting summer 2020.

Matias Jääskeläinen has a BSc in Chemistry and MSc in Theoretical and Computational Methods from the University of Helsinki. He has studied molecular descriptors and configurational sampling for atmospheric molecular clusters. His interests include Data Analysis and Machine Learning

Roope Halonen received his MSc. in computational aerosol physics in 2016, and is currently continuing his studies as a PhD. candidate at University of Helsinki. His main research interests include homogeneous nucleation, reaction kinetics and non-equilibrium processes related to nano-sized particle growth in gas phase.

Theo Kurtén is a Docent and University Lecturer at the Department of Chemistry of the University of Helsinki. His research group uses computational chemistry tools to study reactive sulfur, nitrogen and carbon compounds in the atmosphere, with focus on the gas-phase formation and degradation reactions of extremely low-volatility vapors.

Hanna Vehkamäki is a professor in computational aerosol physics in the Division of Atmospheric Sciences, Department of Physics at the University of Helsinki. She leads a research group of 10-15 people, and their research focuses on molecular level modeling of atmospheric cluster and particle formation as well as ice nucleation. She got her PhD 1998 at the University of Helsinki, was a Research Fellow at the University College London 1998-1999, and has received an ERC stating grant 2009 and an ERC advanced grant 2015. She has also been awarded for her equal opportunities and work well being work in academia. Photo courtesy of Veikko Somerpuro 2018.