Path sampling for atmospheric reactions: formic acid catalysed conversion of SO$_3$ + H$_2$O to H$_2$SO$_4$

Christopher D. Daub$^1$, Enrico Riccardi$^2$, Vesa Hänninen$^1$ and Lauri Halonen$^1$

$^1$Department of Chemistry, University of Helsinki, Helsinki, Finland
$^2$Department of Chemistry, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT

Atmospheric reactions, hitherto studied computationally mainly with static computations in conjunction with transition state theories, can be further described via path sampling calculations. Here we report on an exploratory study of the formic acid catalysed hydrolysis of SO$_3$ to produce H$_2$SO$_4$. We demonstrate that precise measurements of rate coefficients can be obtained in principle for such reactions with an acceptable expenditure of computational resources, and that new insights into the reaction can be obtained by the analysis of the path space explored via path sampling techniques.

Subjects Theoretical and Computational Chemistry, Kinetics and Reactions, Physical Chemistry (other), Thermodynamics and Statistical Mechanics

Keywords Path sampling, Catalysis, Kinetics, Acid rain, Sulphuric acid, Atmospheric chemistry, AIMD, DFT, Transition state theory

INTRODUCTION

SO$_3$ dispersed in the atmosphere reacts with water, generating H$_2$SO$_4$ which is one of the main contributors to acidic rain, as well as contributing to nucleation processes in the upper atmosphere. The SO$_3$ · H$_2$O complex can form readily, however early investigations of the hydrolysis of SO$_3$ to form H$_2$SO$_4$ found the uncatalysed rate to be too low to explain the concentration of atmospheric H$_2$SO$_4$ (Wang et al., 1988; Hofmann & Von Ragué Schleyer, 1994; Loerting & Liedl, 2000). Further studies revealed that the formation of H$_2$SO$_4$ is catalysed by collisions with other atmospheric molecules (Akhmatskaya et al., 1997; Hazra & Sinha, 2011; Long et al., 2012; Torrent-Sucarrat, Francisco & Anglada, 2012; Bandyopadhyay, Kumar & Biswas, 2017; Sarkar, Oram & Bandyopadhyay, 2019), where the common denominator is the facilitation of a double proton transfer via the catalysing molecule (Kumar, Sinha & Francisco, 2016). The relative importance of collisions with different molecules is determined by both the atmospheric concentration of the colliding molecule and the species-dependent reaction rate. For example, while water vapour has a high atmospheric concentration, the reaction barrier remains quite high when an additional water molecule acts as the catalyst (Akhmatskaya et al., 1997; Long et al., 2012). On the other hand, NH$_3$ is an effective catalyst, lowering the reaction barrier.
significantly, but its atmospheric concentration is comparatively low (Bandyopadhyay, Kumar & Biswas, 2017; Sarkar, Oram & Bandyopadhyay, 2019). Recent work has highlighted the importance of formic acid, HCOOH, in catalysing the SO$_3$ hydration reaction (Hazra & Sinha, 2011; Long et al., 2012). Static quantum chemical calculations of the initial and final states, combined with reaction kinetics and transition state theory (TST), show that the reaction becomes nearly barrier-less, and hence that HCOOH is an effective catalyst. Combined with the recent findings that formic acid concentrations in the atmosphere are higher than previously assumed (Millet et al., 2015; Chaliyakunnel et al., 2016), and the presence of formic acid in volatile organic compounds (VOCs) emitted in boreal forests (Rantala et al., 2015), it seems likely that the formic acid catalysed hydration of SO$_3$ is a significant contributor to sulphuric acid concentrations in the atmosphere.

Investigations of gas phase SO$_3$ hydration reactions, however, to our knowledge have heretofore only been performed by static calculations, with kinetics and dynamics only included via TST (Long et al., 2012) or Rice, Ramsperger, Kassel, Markus (RRKM) theory (Hazra & Sinha, 2011). In the present work, we aim to directly investigate the dynamics of the formic acid catalysed conversion of SO$_3$ · H$_2$O to H$_2$SO$_4$ via path sampling (Van Erp, 2007; Van Erp, Moroni & Bolhuis, 2003; Moroni, Bolhuis & Van Erp, 2004). The method has recently reached a sufficient level of maturity, both theoretically (Riccardi, Dahlen & Van Erp, 2017; Cabriolu et al., 2017) and computationally (Lervik, Riccardi & Van Erp, 2017; Riccardi et al., 2019a). Notable recent work using path sampling has focused on chemical or other transitions between states in aqueous systems, for example water auto-ionization (Moqadam et al., 2018), interactions between biomolecules (Riccardi et al., 2019c) and silicate condensation (Moqadam et al., 2017). The main focus of the present work is to demonstrate the application of path sampling to investigating a new class of gas-phase reactions.

The path sampling approach here chosen is the replica exchange transition interface sampling (RETIMS) (Van Erp, 2007; Cabriolu et al., 2017). It permits the investigation of the path space, and thus the determination of a reaction rate, orders of magnitude faster than conventional simulations. This computational enhancement does not come at the cost of resolution since the method provides the same results, when converged, as infinitely long standard simulations. The RETIMS methodology will be combined with ab initio molecular dynamics (AI-MD) simulations to predict the reaction rate of the uni-molecular SO$_3$ hydration reaction with formic acid, considering different density functional theory (DFT) functionals and basis sets, and comparing the results with data available in the literature produced by more approximate methods (e.g., TST).

**MATERIALS AND METHODS**

**Description of the reaction and DFT methods**

We perform DFT-based Born–Oppenheimer molecular dynamics (BO-MD) simulations using the QUICKSTEP module in CP2K (VandeVondele et al., 2005). Before beginning these simulations we considered several different combinations of DFT functionals and
basis sets, first investigating the relative energies of the product and reactant states using NWChem (Valiev et al., 2010).

Snapshots of optimised geometries in the reactant (A1 and A2) and product (B1 and B2) states are displayed in Fig. 1. Other computational studies of SO$_3$ hydration have considered the complete reaction pathway, including the formation of the pre-reactive complex via atmospheric collisions. In this study, we focus only on the uni-molecular reaction involving the double proton transfer from the water molecule via the formic acid to form sulphuric acid, that is

\[
\text{H}_2\text{O} \cdot \text{SO}_3 \cdot \text{HCOOH} \xrightarrow{k_{AB}} \text{H}_2\text{SO}_4 \cdot \text{HCOOH}
\]

This reaction involves no collisions, and can be treated as an isomerization reaction, with first-order kinetics described by a single rate coefficient $k_{AB}$ with units of inverse time. In previous work, this uni-molecular rate coefficient has been computed using TST (Long et al., 2012) and RRKM theory (Hazra & Sinha, 2011). The uni-molecular reaction barrier height was found to be low (0.3 kJ mol$^{-1}$ at the MP2/6-311++G(3df, 3pd) level of theory (Hazra & Sinha, 2011)) compared to the barrier for the water catalysed reaction.

Two different local minima in both reactant and product states are present, differing only in the orientation of one OH bond initially belonging to the water molecule. The minimum in the reactant state (A1) differs from that reported in other investigations.

Figure 1 Snapshots of optimised initial states (A–D) and final product states (E–H) of the system. Indicated atoms and inter-atomic distances are used in the definition of the order parameter $s(r)$ defined in Eq. (1).
of this reaction (A2), but is only $\sim 1$ kJ mol$^{-1}$ lower in energy. The minimum energy in the product state (B2) is generally several kJ mol$^{-1}$ lower in energy than the other local minimum (B1). We remark that we did not consider the final reorganization of the post-reaction $\text{H}_2\text{SO}_4 \cdot \text{HCOOH}$ complex into a lower energy conformation which is the true global minimum (Hazra & Sinha, 2011). The uni-molecular rate coefficient we measure is unaffected by this final rearrangement since it is a spontaneous process. We display some of these results along with comparisons with the literature in Table 1.

Hybrid functionals such as B3LYP or PBE0 which include some of the Hartree–Fock exchange correlation are necessary to obtain DFT based results which are in good agreement with fully ab initio methods. However, using hybrid functionals in BO-MD simulations would be prohibitively expensive, especially for an exploratory study such as this one. In the end, in our BO-MD simulations, we compare two Generalized Gradient Approximation (GGA) DFT methods, with three different basis sets (Goedecker, Teter & Hutter, 1996). First, we use the BLYP functional (Becke, 1988; Lee, Yang & Parr, 1988) with the DZVP-MOLOPT-GTH basis functions, which we have previously used to investigate formic acid in bulk water (Murdachaew et al., 2016; Hänninen et al., 2018; Daub & Halonen, 2019). Although the BLYP results lack accuracy in terms of the comparative energies of the initial and final states, it was a relatively inexpensive methodology to use for initial investigation. Thereafter, the PBE functional (Perdew, Burke & Ernzerhof, 1996) and both the TZVP-MOLOPT-GTH and TZV2P-MOLOPT-GTH basis sets have been used. There are some differences between the basis sets implemented in NWChem and the plane-wave basis sets employed in CP2K, but respective double-zeta and triple-zeta basis sets should be similar. In particular, we note that the 6-311++G(3df, 3pd) Gaussian basis set produced comparable results to the TZV2P-MOLOPT-GTH basis set. All simulations, both with NWChem and CP2K, used the Grimme D2 dispersion correction (Grimme, 2004, 2006). All BO-MD simulations were

| Theory/basis set                        | $E_{A1}$ | $E_{A2}$ | $E_{B1}$ | $E_{B2}$ |
|----------------------------------------|----------|----------|----------|----------|
| BLYP/DZVP                              | 0.0      | 1.1      | 8.1      | 3.5      |
| BLYP/TZVP                              | 0.0      | 1.2      | 0.3      | -2.9     |
| B3LYP/TZVP                             | 0.0      | 1.1      | -17.4    | -22.3    |
| PBE/TZVP                               | 0.0      | 1.3      | -9.2     | -13.0    |
| PBE/6-311++G(3df, 3pd)                  | 0.0      | 1.0      | -13.9    | -18.4    |
| PBE/aug-cc-pVTZ                        | 0.0      | 1.2      | -12.5    | -16.5    |
| PBE0/TZVP                              | 0.0      | 1.3      | -27.3    | -32.4    |
| MP2/6-311++G(3df, 3pd) (Hazra & Sinha, 2011) | -        | 0.0      | -        | -32.4    |
| CCSD(T)/MP2/aug-cc-pV(T + d)Z (Long et al., 2012) | -        | 0.0      | -        | -42.5 (-41.0) |
performed in the microcanonical (NVE) ensemble in a cubic periodic box of fixed geometry with a box length of $L = 8 \text{ Å}$ ($1 \text{ Å} = 10^{-10} \text{ m}$).

**Path sampling using RETIS**

The estimation of a reaction rate via path sampling requires a numerical descriptor able to distinguish the product and reactant states and indicate the reaction progress: the order parameter. A proper selection of this parameter allows a more efficient sampling of the reaction pathway. In this study, we identify three different inter-atomic distances which all must decrease as the reaction proceeds. The parameter we introduce is the negative of the sum of these three distances

$$s(r) = - (r_{SO} + r_{O_1H_1} + r_{O_2H_2}) \quad (1)$$

The negative value is a technical necessity so that the value of the order parameter, $s(r)$, will increase from the reactant to the product state. The atom indices indicated are labelled in Fig. 1, as well as the value of $s(r)$ in some of the configurations shown. The value of $s(r)$ in the optimised initial state(s) depends on the DFT method, changing from $s(r) \approx -5.65 \text{ Å}$ in the BLYP calculations to $s(r) \approx -5.3 \text{ Å}$ in the PBE calculations. Regardless of the DFT method, $s(r) \approx -3.75 \text{ Å}$ in the optimised final state after the reaction.

The RETIS (Van Erp, 2007; Cabriolu et al., 2017) method has been adopted to properly sample the transition region. We will briefly describe the method here, and point interested readers to the existing literature for more details (Van Erp, 2007; Riccardi, Dahlen & Van Erp, 2017; Cabriolu et al., 2017; Lervik, Riccardi & Van Erp, 2017; Van Erp et al., 2016).

We define a set of $n_{\text{inter}}$ interfaces $\lambda_i \in \lambda_0, \ldots, \lambda_{n_{\text{inter}} - 1}$, with $\lambda_i < \lambda_{i+1}$, sub-dividing the path between product and reactant states along the order parameter $s(r)$. Configurations with $s(r) < \lambda_0$ are defined as being in the reactant state $A$, while configurations with $s(r) > \lambda_{n_{\text{inter}} - 1}$ are defined as being in the product state $B$. Furthermore, $n_{\text{inter}} - 1$ sets of path ensembles $[i^+], i = 0, \ldots, n_{\text{inter}} - 2$ are defined which consist of all paths which start at $\lambda_0$, proceed in the positive direction to cross $\lambda_i$ at least once, and end by crossing either $\lambda_0$ again or by crossing the final interface $\lambda_{n_{\text{inter}} - 1}$.

An additional ensemble required to be defined in RETIS is the $[0^-]$ ensemble, containing all of the paths which begin at $\lambda_0$, proceed in a backwards direction, and cross $\lambda_0$ again.

Statistical mechanical arguments can be used to show that the reaction rate coefficient $k_{AB}$ can be expressed as

$$k_{AB} = f_A P_{\text{cross}} = f_A \prod_{i=0}^{n-1} P_A (\lambda_{i+1} | \lambda_i) \quad (2)$$

where the initial flux $f_A$ is determined from the average lengths of paths in the $[0^-]$ and $[0^+]$ ensembles

$$f_A = \frac{1}{\langle t^{0-} \rangle + \langle t^{0+} \rangle} \quad (3)$$

and the partial crossing probability term, $P_A (\lambda_{i+1} | \lambda_i)$, is the conditional probability that a path crosses interface $\lambda_{i+1}$ given it has already crossed $\lambda_i$. 

---

**References**

Daub et al. (2020), *PeerJ Physical Chemistry*, DOI 10.7717/peerj-pchem.7
The initial points from which to generate a new path are determined with different Monte Carlo (MC) moves. In PyRETIS, the MC moves used are called shooting, time reversal and ensemble swapping, and were selected with a relative probability of 0.5, 0.25 and 0.25, respectively, according to the developers’ guidelines (Lervik, Riccardi & Van Erp, 2017).

Table 2 summarises all RETIS simulations with the same temperature $T = 300$ K included in this study. We have also completed a number of simulations at different temperatures to investigate the temperature dependence of the reaction rates. Different initial configurations corresponding to the two different reactant minima A1 and A2 were used. All initial paths were generated by the PyRETIS kick method (Lervik, Riccardi & Van Erp, 2017), whereby a path is constructed starting from the initial configuration by altering velocities at each step and only accepting the step if it increases the value of $s(r)$. In some cases, after a short equilibration run simulations were restarted by using the PyRETIS load function (Riccardi et al., 2019a) to start from the final path accepted in the first simulation. In general, initial paths generated by kick (an approach based on a combination of Monte Carlo and BO-MD) can require many cycles to find physically realistic reaction paths; in this particular study, the initial paths agree closely with the real reaction pathways found using RETIS. To facilitate the reproduction of our findings, the inputs files will be provided to the PyRETIS developers to be uploaded on the software website, according to the Open Data initiative (Riccardi, Pantano & Potestio, 2019b).

| DFT + basis set   | Init. cfg. | $n_{\text{inter}}$ | $\lambda_{0}/\text{Å}$ | $\lambda_{n_{\text{inter}}-1}/\text{Å}$ | $n_{\text{cycles}}$ | $t_{\text{sim}}$/ps |
|-------------------|------------|---------------------|-------------------------|-----------------------------|---------------------|----------------------|
| BLYP, DZVP        | A1         | 7                   | −5.65                   | −4.0                        | 1,331               | 644                  |
| BLYP, DZVP        | A1         | 10                  | −5.7                    | −3.8                        | 1,015               | 916                  |
| BLYP, DZVP        | A2         | 7                   | −5.65                   | −4.0                        | 914                 | 458                  |
| PBE, TZVP         | A1         | 8                   | −5.25                   | −3.9                        | 999                 | 325                  |
| PBE, TZVP         | A2         | 8                   | −5.25                   | −3.9                        | 1,009               | 324                  |
| PBE, TZV2P        | A1         | 8                   | −5.25                   | −3.9                        | 996                 | 339                  |
| PBE, TZV2P        | A2         | 7                   | −5.3                    | −3.9                        | 993                 | 318                  |
| PBE, TZV2P        | A1         | 6                   | −5.7                    | −3.9                        | 756                 | 647                  |
| BLYP, DZVP (250 K)| A1         | 10                  | −5.8                    | −4.0                        | 975                 | 1,265                |
| BLYP, DZVP (250 K)| A2         | 6                   | −5.7                    | −4.0                        | 891                 | 447                  |
| PBE, TZV2P (250 K)| A1         | 6                   | −5.65                   | −3.9                        | 519                 | 533                  |
| PBE, TZV2P (275 K)| A1         | 6                   | −5.65                   | −3.9                        | 530                 | 496                  |
| PBE, TZV2P (325 K)| A1         | 6                   | −5.65                   | −3.9                        | 937                 | 539                  |
| PBE, TZV2P (350 K)| A1         | 6                   | −5.65                   | −3.9                        | 988                 | 520                  |

The initial points from which to generate a new path are determined with different Monte Carlo (MC) moves. In PyRETIS, the MC moves used are called shooting, time reversal and ensemble swapping, and were selected with a relative probability of 0.5, 0.25 and 0.25, respectively, according to the developers’ guidelines (Lervik, Riccardi & Van Erp, 2017).

In Table 2 we summarise all RETIS simulations with the same temperature $T = 300$ K included in this study. We have also completed a number of simulations at different temperatures to investigate the temperature dependence of the reaction rates. Different initial configurations corresponding to the two different reactant minima A1 and A2 were used. All initial paths were generated by the PyRETIS kick method (Lervik, Riccardi & Van Erp, 2017), whereby a path is constructed starting from the initial configuration by altering velocities at each step and only accepting the step if it increases the value of $s(r)$. In some cases, after a short equilibration run simulations were restarted by using the PyRETIS load function (Riccardi et al., 2019a) to start from the final path accepted in the first simulation. In general, initial paths generated by kick (an approach based on a combination of Monte Carlo and BO-MD) can require many cycles to find physically realistic reaction paths; in this particular study, the initial paths agree closely with the real reaction pathways found using RETIS. To facilitate the reproduction of our findings, the inputs files will be provided to the PyRETIS developers to be uploaded on the software website, according to the Open Data initiative (Riccardi, Pantano & Potestio, 2019b).
RESULTS

As shown in Eq. 2, the computed reaction rate coefficient, $k_{AB}$, is derived from the initial flux $f_A$ multiplied by the total crossing probability $P_{\text{cross}}$. The value of $f_A \times P_{\text{cross}}$ at the end of $n_{\text{cycles}}$ of the RETIS simulations is reported in Fig. 2. The value of $k_{AB}$ obtained via path sampling using the RETIS methodology is directly comparable with the uni-molecular rate coefficient describing the double proton transfer reaction forming $\text{H}_2\text{SO}_4$. We show the average result obtained at the end of $n_{\text{cycles}}$ of all RETIS simulations using a given DFT method, basis set and temperature. Our most accurate estimate at $T = 300 \text{ K}$ has been obtained via PBE calculations with the TZV2P basis set, $k_{AB} = 3.5 \times 10^{-4} \text{ fs}^{-1}$, with a statistical error of $\sim 50\%$. For comparison, a computation of the rate coefficient at $T = 300 \text{ K}$ based on TST and reaction kinetics calculations (see column 6 of Table S4 in the Supporting Information for Long et al., 2012) gave an estimate of $k_{AB} = 1.9 \times 10^{-3} \text{ fs}^{-1}$, while calculations based on the RRKM kinetic theory (see Table 7 in the Supporting Information for Hazra & Sinha, 2011) gave an estimate of $k_{AB} = 4.6 \pm 0.1 \times 10^{-4} \text{ fs}^{-1}$. Overall, our predicted rate coefficient is in reasonable agreement with previous estimates in the literature, despite the fact that we underestimate the energy difference between the reactant and product states (see Table 1). We might reasonably expect that a RETIS calculation using a hybrid functional such as PBE0 with a more accurate energy difference would result in a somewhat larger rate coefficient. Unfortunately, at the present time the computational resources required for
such calculations leave the use of hybrid functionals in path sampling calculations beyond our reach.

Our estimated rate coefficients $k_{AB}$ reported in Fig. 2 include a total of at least 300 ps of simulation data for each RETIS run. In Fig. 3 we show the final value of $k_{AB}$ (i.e., the value of $k_{AB}$ at the maximum value of $s(r)$) as a function of the RETIS cycle number for the PBE calculation using the TZV2P basis set. This plot shows that our results are well-converged, and in fact that reliable results could have been obtained from a simulation using only half the data, or less. More careful optimisation of the interface positions $\lambda_i$ would also help the results to converge with less total simulation time.

Comparison of results using the same basis set and level of theory (BLYP, DZVP) at two different temperatures (300 K and 250 K) showed that the RETIS simulations predict a smaller rate coefficient at lower temperature. While this may make intuitive sense, it disagrees with the TST results, which predict an increase in the unimolecular rate coefficient at lower temperature (Long et al. (2012), see column 6 of Table S4 of the Supporting Information). Subsequently, we performed a series of simulations using the most reliable basis set and level of theory available to us (i.e., PBE, TZV2P), at a series of temperatures ranging from $T = 250$ to 350 K. This approach should allow us to directly compare the temperature dependence observed in the TST calculation with the results of the RETIS method. We summarise these results in Fig. 4.

The rate coefficient seems to be lower at $T = 250$ K, consistent with the BLYP data shown in Fig. 2, however over the full range of temperature we have sampled we cannot
measure any statistically meaningful correlation between \( k_{AB} \) and \( T \). This lack of temperature dependence is in contrast with the monotonic decrease in the uni-molecular reaction rate predicted by TST. It is unlikely that this temperature dependence (or lack thereof) would qualitatively change if the basis set or level of theory were changed.

In the path sampling approach, the temperature dependency of the rate coefficient arises from entropic contributions which are directly computed, and complicating factors such as vibrational anharmonicity are implicitly included. In TST or RRKM, on the other hand, entropic contributions are considered, but in a way that depends on certain assumptions about, for example, the harmonic nature of the vibrational potentials used to construct the vibrational partition function. The rather different temperature dependence of the rate coefficient on we have observed points toward how direct methods like RETIS might lead to different insights into reaction pathways. At the same time, in this case the reaction barrier is low and the uni-molecular reaction rate is very fast compared with the initial collision rate, and therefore the uni-molecular rate coefficient does not play a major role in determining the overall reaction rate \((\text{Long et al., 2012})\).

To investigate the reaction pathway, frame density plots can be used to visualise, in path space, the values of different descriptors. The visualisation and analysis tool of PyRETIS, PyVisA (O. Aarøen, E. Riccardi, 2019, unpublished data), has been used to plot the transition in different path spaces. Correlations, even if partial, can then be detected allowing direct analysis of the transition pathways available from the RETIS simulations. Details about the entire reaction, and not just the static reactant and product states, can be extracted for the whole transition region.

Figure 4 Temperature dependence of the reaction rate coefficient \( k_{AB} \), comparing our results with RETIS and the PBE functional and the TZV2P basis set with previous literature results using TST and RRKM theories. Error bars are one standard deviation in the mean. The black line is from a linear regression of the RETIS data.
Figure 5 | Frame densities for trajectories beginning in the final RETIS ensemble (−3.9 Å > s(r) > −4.6 Å). Interface locations $\lambda_i$ are indicated by the vertical lines when $s(r)$ is plotted on the horizontal axis. (A) $r_{SOw}$ vs. $s(r)$, (B) $r_{O1H1}$ vs. $s(r)$, (C) $r_{O2H2}$ vs. $r_{SOw}$, and (D) $r_{O2H2}$ vs. $r_{O1H1}$. DOI: 10.7717/peerj-pchem.7/fig-5
We have examined the three interatomic distances which contribute to our order parameter $s(r)$, as well as the order parameter itself. Those interatomic distances are defined in accordance with the atom labels shown in Fig. 1. In Fig. 5 we display four different correlation plots considering only the accepted paths belonging to the outermost ensemble, i.e. with $-4.6 \, \text{Å} < s(r) < -3.9 \, \text{Å}$. In the top two figures, the variation in the distance between $S$ and $O_{\text{W}}$, $r_{\text{SO}_w}$, and the distance $r_{O_{\text{H}_1}}$ are shown. These are both plotted versus the order parameter $s(r)$. All three of the individual components of $s(r)$ are strongly correlated with the order parameter, as shown by the high values of the correlation variable $r^2$.

When looking at the correlation between the individual inter-atomic distances themselves (see bottom 2 plots of Fig. 5), the strongest correlation exists between $r_{\text{SO}_w}$ and $r_{O_{\text{H}_1}}$. Comparing to Fig. 1, this seems sensible; as the reaction proceeds the distance between the (initial) water oxygen and the sulphur atom varies at the same time as the water hydrogen atom nearest the formic acid begins to transfer over to the formic acid molecule. Correlations with the third component of $s(r)$, that is, $r_{O_{\text{H}_2}}$, the distance between the acidic hydrogen and the closest $\text{SO}_3$ oxygen, are considerably weaker. According to the bottom right plot of Fig. 5, it seems that $r_{O_{\text{H}_2}}$ is almost constant until $r_{O_{\text{H}_1}} < 1.1 \, \text{Å}$, at which point $r_{O_{\text{H}_1}}$ suddenly changes from $\sim 1.4 \, \text{Å}$ down to $1.2 \, \text{Å}$. In other words, we suggest that the final step in the double proton transfer reaction is the change in $r_{O_{\text{H}_2}}$, and it does not begin without the other parts of the reaction having occurred. These findings are consistent with, and further explain the limited data about the transition region available from a TST investigation, which showed that $r_{O_{\text{H}_1}} \sim 1.2 \, \text{Å}$ at the transition state, while $r_{O_{\text{H}_2}}$ is larger, $\sim 1.4 \, \text{Å}$ at the transition state (Long et al., 2012).

**CONCLUSIONS**

We have made use of path sampling BO-MD simulations to study the hydrolysis of $\text{SO}_3$ when catalysed by formic acid. We discuss in detail the reaction pathway(s) that can describe the conversion from $\text{SO}_3 + \text{H}_2\text{O}$ to $\text{H}_2\text{SO}_4$. The mechanism has been highlighted and the overall rate computed with different levels of theory.

Our estimate of the uni-molecular rate coefficient for the reaction compares well with previous estimates using TST or RRKM theory (Hazra & Sinha, 2011; Long et al., 2012). Both methods have strengths and weaknesses. On the one hand, the static TST and RRKM calculations can be done with higher levels of theory and larger basis sets than we can handle with BO-MD and the currently available computational resources. It is also impossible, at the current state-of-the-art, to include the contributions of zero-point energy and tunnelling in the path sampling simulations. On the other hand, the kinetic theories are limited by their use of harmonic approximations to compute the vibrational partition functions which enter into the kinetic calculations (Long et al., 2012; Duncan, Bell & Truong, 1998). Path sampling, instead, is inherently dynamical, and the impact of vibrational anharmonicity is implicitly included. Being able to visualise the entire reactive landscape with path sampling, and not just the transition state itself, can also lead to new physical insights unavailable in traditional analyses based on TST or other kinetic theories. A more concerted attempt to carefully compare methodologies for reaction rate calculations would be a worthwhile subject for further study.
We want to emphasise that our results do not consider the complete \( \text{SO}_3 \) hydrolysis reaction, for example, the formation of the pre-reactive complex by atmospheric collisions. Our promising results encourage a broader application of the RETIS approach, which will be a goal pursued in forthcoming works. As the outset of this study, our stated focus was to demonstrate the effectiveness of path sampling to study a relatively simple uni-molecular gas-phase reaction. We acknowledge that in this case the uni-molecular reaction rate is much faster than the initial collision rate, and so overall \( \text{H}_2\text{SO}_4 \) formation in the atmosphere is not controlled by the uni-molecular reaction. The overall reaction was studied in great detail by previous researchers (Hazra & Sinha, 2011; Long et al., 2012), and our current results do not challenge their general conclusions.

The path sampling approach has here been successfully applied with BO-MD to study an atmospheric reaction. Our work highlights the advantage of studying relatively small gas-phase clusters compared with bulk aqueous phases, as the application of methods for studying rare events in combination with BO-MD is significantly eased in such small systems. We showed that path sampling can be exploited to efficiently generate ensembles of reactive paths and precise estimates of the reaction rate in gas-phase reactions of interest in atmospheric chemistry. The promising approach and results here reported are intended as an initial step towards a broader study of atmospherically relevant reactions using these novel sampling methods, including their application in our ongoing investigations into formic acid deprotonation (Murdachaew et al., 2016; Daub & Halonen, 2019).

**ADDITIONAL INFORMATION AND DECLARATIONS**

**Funding**

Christopher Daub, Vesa Hänninen and Lauri Halonen are supported by the Academy of Finland (Grant Number 294752). Enrico Riccardi is supported by the Research Council of Norway (Project Number 267669). The Project HPC-Europa3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme, funded travel costs for Enrico Riccardi’s visit to Helsinki and computational resources. Additional computing resources were provided by Finland’s Center for Scientific Computing (CSC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Grant Disclosures**

The following grant information was disclosed by the authors:

Academy of Finland: 294752.
Research Council of Norway: 267669.
The Project HPC-Europa3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme.
Finland’s Center for Scientific Computing (CSC).
Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Christopher D. Daub conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Enrico Riccardi conceived and designed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the paper, and approved the final draft.
- Vesa Hänninen conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Lauri Halonen conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability
The following information was supplied regarding data availability:
PyRETIS code is available at http://www.pyretis.org.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-pchem.7#supplemental-information.

REFERENCES
Akhmatskaya EV, Apps CJ, Hillier IH, Masters AJ, Palmer IJ, Watt NE, Vincent MA, Whitehead JC. 1997. Hydrolysis of SO$_3$ and ClONO$_2$ in water clusters. Journal of the Chemical Society, Faraday Transactions 93(16):2775–2779 DOI 10.1039/a701768e.

Bandyopadhyay B, Kumar P, Biswas P. 2017. Ammonia catalyzed formation of sulfuric acid in troposphere: the curious case of a base promoting acid rain. Journal of Physical Chemistry A 121(16):3101–3108 DOI 10.1021/acs.jpca.7b01172.

Becke AD. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. Physical Review A 38(6):3098–3100 DOI 10.1103/PhysRevA.38.3098.

Cabriolu R, Refsnes KMS, Bolhuis PG, Van Erp TS. 2017. Foundations and latest advances in replica exchange transition interface sampling. Journal of Chemical Physics 147(15):152722 DOI 10.1063/1.4989844.

Chaliyakunnel S, Millet DB, Wells KC, Cady-Pereira KE, Shephard MW. 2016. A large underestimate of formic acid from tropical fires: constraints from space-borne measurements. Environmental Science & Technology 50(11):5631–5640 DOI 10.1021/acs.est.5b06385.

Daub CD, Halonen L. 2019. Ab initio molecular dynamics simulations of the influence of lithium bromide salt on the deprotonation of formic acid in aqueous solution. Journal of Physical Chemistry B 123(31):6823–6829 DOI 10.1021/acs.jpcb.9b04618.

Duncan WT, Bell RL, Truong TN. 1998. The rate: program for ab initio direct dynamics calculations of thermal and vibrational-state-selected rate constants. Journal of Computational Chemistry 19:1039–1052 DOI 10.1002/(SICI)1096-987X(19980715)19:9<1039::AID-JCC5>3.0.CO;2-R.
Goedecker S, Teter M, Hutter J. 1996. Separable dual-space Gaussian pseudopotentials. *Physical Review B* 54(3):1703–1710 DOI 10.1103/PhysRevB.54.1703.

Grimme S. 2004. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* 25(12):1463–1473 DOI 10.1002/jcc.20078.

Grimme S. 2006. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* 27(15):1787–1799 DOI 10.1002/jcc.20495.

Hazra MK, Sinha A. 2011. Formic acid catalyzed hydrolysis of SO$_3$ in the gas phase: a barrierless mechanism for sulfuric acid production of potential atmospheric importance. *Journal of the American Chemical Society* 133(43):17444–17453 DOI 10.1021/ja207393v.

Hofmann P, Von Ragué Schleyer P. 1994. Acid rain: ab initio investigation of the H$_2$O·SO$_3$ complex and its conversion into H$_2$SO$_4$. *Journal of the American Chemical Society* 116(11):4947–4952 DOI 10.1021/ja00090a045.

Hänninen V, Murdachaew G, Nathanson GM, Gerber RB, Halonen L. 2018. Ab initio molecular dynamics studies of formic acid dimer colliding with liquid water. *Physical Chemistry Chemical Physics* 20(36):23717–23725 DOI 10.1039/C8CP03857K.

Kumar M, Sinha A, Francisco JS. 2016. Role of double hydrogen atom transfer reactions in atmospheric chemistry. *Accounts of Chemical Research* 49(5):877–883 DOI 10.1021/acs.accounts.6b00040.

Lee C, Yang W, Parr RG. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* 37(2):785–789 DOI 10.1103/PhysRevB.37.785.

Lervik A, Riccardi E, Van Erp TS. 2017. PyRETIS: a well-done, medium-sized Python library for rare events. *Journal of Computational Chemistry* 38(28):2439–2451 DOI 10.1002/jcc.24900.

Loerting T, Liedl KR. 2000. Toward elimination of discrepancies between theory and experiment: the rate constant of the atmospheric conversion of SO$_3$ to H$_2$SO$_4$. *Proceedings of the National Academy of Sciences of the United States of America* 97(16):8874–8878 DOI 10.1073/pnas.97.16.8874.

Long B, Long ZW, Wang YB, Tan XF, Han YH, Long CY, Qin SJ, Zhang WJ. 2012. Formic acid catalyzed gas-phase reaction of H$_2$O with SO$_3$ and the reverse reaction: a theoretical study. *ChemPhysChem* 13(1):323–329 DOI 10.1002/cphc.201100558.

Millet DB, Baasandorj M, Farmer DK, Thornton JA, Baumann K, Brophy P, Chaliyakunnel S, De Gouw JA, Graus M, Hu L, Koss A, Lee BH, Lopez-Hiliker FD, Neuman JA, Paulot F, Peischl J, Pollack IB, Ryerson TB, Warneke C, Williams BJ, Xu J. 2015. A large and ubiquitous source of atmospheric formic acid. *Atmospheric Chemistry and Physics* 15(11):6283–6304 DOI 10.5194/acp-15-6283-2015.

Moqadam M, Lervik A, Riccardi E, Venkatraman V, Alsberg BK, Van Erp TS. 2018. Local initiation conditions for water autoionization. *Proceedings of the National Academy of Sciences of the United States of America* 115(20):E4569–E4576 DOI 10.1073/pnas.1714070115.

Moqadam M, Riccardi E, Trinh TT, Lervik A, Van Erp TS. 2017. Rare event simulations reveal subtle key steps in aqueous silicate condensation. *Physical Chemistry Chemical Physics* 19(20):13361–13371 DOI 10.1039/C7CP01268C.

Moroni D, Bolhuis PG, Van Erp TS. 2004. Rate constants for diffusive processes by partial path sampling. *Journal of Chemical Physics* 120(9):4055–4065 DOI 10.1063/1.1644537.
Murdachaew G, Nathanson GM, Gerber RB, Halonen L. 2016. Deprotonation of formic acid in collisions with a liquid water surface studied by molecular dynamics and metadynamics simulations. Physical Chemistry Chemical Physics 18(43):29756–29770 DOI 10.1039/C6CP06071D.

Perdew JP, Burke K, Ernzerhof M. 1996. Generalized gradient approximation made simple. Physical Review Letters 77(18):3865–3868 DOI 10.1103/PhysRevLett.77.3865.

Rantala P, Aalto J, Taipale R, Ruskanen TM, Rinne J. 2015. Annual cycle of volatile organic compound exchange between a boreal pine forest and the atmosphere. Biogeosciences 12(19):5753–5770 DOI 10.5194/bg-12-5753-2015.

Riccardi E, Dahlen O, Van Erp TS. 2017. Fast decorrelating Monte Carlo moves for efficient path sampling. Journal of Physical Chemistry Letters 8(18):4456–4460 DOI 10.1021/acs.jpclett.7b01617.

Riccardi E, Lervik A, Roet S, Aaroen O, Van Erp TS. 2019a. PyRETIS 2: an improbability drive for rare events. Journal of Computational Chemistry 41(4):370–377 DOI 10.1002/jcc.26112.

Riccardi E, Pantano S, Potestio R. 2019b. Envisioning data sharing for the biocomputing community. Interface Focus 9(3):20190005 DOI 10.1098/rsfs.2019.0005.

Riccardi E, Van Mastbergen EC, Navarre WW, Vreede J. 2019c. Predicting the mechanism and rate of H-NS binding to AT-rich DNA. PLOS Computational Biology 15(3):e1006845 DOI 10.1371/journal.pcbi.1006845.

Sarkar S, Oram BK, Bandyopadhyay B. 2019. Influence of ammonia and water on the fate of sulfur trioxide in the troposphere: theoretical investigation of sulfamic acid and sulfuric acid formation pathways. Journal of Physical Chemistry A 123(14):3131–3141 DOI 10.1021/acs.jPCA.8b09306.

Torrent-Sucarrat M, Francisco JS, Anglada JM. 2012. Sulfuric acid as autocatalyst in the formation of sulfuric acid. Journal of the American Chemical Society 134(51):20632–20644 DOI 10.1021/ja207523b.

Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, Van Dam HJJ, Wang D, Nieplocha J, Apra E, Windus TL, De Jong WA. 2010. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. Computer Physics Communications 181(9):1477–1489 DOI 10.1016/j.cpc.2010.04.018.

Van Erp TS. 2007. Reaction rate calculation by parallel path swapping. Physical Review Letters 98(26):268301 DOI 10.1103/PhysRevLett.98.268301.

Van Erp TS, Moqadam M, Riccardi E, Lervik A. 2016. Analyzing complex reaction mechanisms using path sampling. Journal of Chemical Theory and Computation 12(11):5398–5410 DOI 10.1021/acs.jctc.6b00642.

Van Erp TS, Moroni D, Bolhuis PG. 2003. A novel path sampling method for the calculation of rate constants. Journal of Chemical Physics 118(17):7762–7774 DOI 10.1063/1.1562614.

VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J. 2005. QUICKSTEP: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. Computer Physics Communications 167(2):103–128 DOI 10.1016/j.cpc.2004.12.014.

Wang X, Jin YG, Suto M, Lee LC, O’Neal HE. 1988. Rate constant of the gas phase reaction of SO3 with H2O. Journal of Chemical Physics 89(8):4853–4860 DOI 10.1063/1.455680.