



# Computational Estimation of the Binding Energies of $\text{PO}_x$ and $\text{HPO}_x$ ( $x = 2, 3$ ) Species

Elettra L. Piacentino and Karin I. Öberg

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

Received 2022 May 16; revised 2022 September 14; accepted 2022 September 29; published 2022 November 9

## Abstract

The distribution of molecules between the gas and solid phase during star and planet formation determines the trajectory of gas and grain surface chemistry, as well as the delivery of elements to nascent planets. This distribution is primarily set by the binding energies of different molecules to water ice surfaces. We computationally estimated the binding energies of 10 astrochemically relevant P-bearing species on water surfaces. We also validate our method for 20 species with known binding energies. We used Density Functional Theory (DFT) calculations (M06-2X/aug-cc-pVDZ) to calculate the energetics of molecules and water-molecule clusters (1–3  $\text{H}_2\text{O}$  molecules) and from this determined the binding energy by comparing the complex and the separate molecule and cluster energies. We also explore whether these estimates can be improved by first calibrating our computational method using experimentally measured binding energies. Using the 20 reference molecules we find that the  $2\text{H}_2\text{O}$  cluster size yields the best binding energy estimates and that the application of a calibration to the data may improve the results for some classes of molecules, including more-refractory species. Based on these calculations we find that small P-bearing molecules such as  $\text{PH}_3$ ,  $\text{PN}$ ,  $\text{PO}$ ,  $\text{HPO}$ ,  $\text{PO}_2$ , and  $\text{POOH}$  are relatively volatile and should desorb prior or concomitantly with water ice, while  $\text{H}_2\text{PO}$ ,  $\text{HPO}_2$ ,  $\text{PO}_3$ , and  $\text{PO}_2\text{OH}$  can strongly bind to any hydroxylated surface and will likely remain on the interstellar grains surface past the desorption of water ice. The depletion of P carriers on grains constitutes a pathway for the inclusion of phosphorous molecules in planets and planetesimals.

*Unified Astronomy Thesaurus concepts:* [Interstellar molecules \(849\)](#); [Theoretical models \(2107\)](#); [Astrochemistry \(75\)](#); [Interstellar abundances \(832\)](#)

## 1. Introduction

The relative abundance of chemical species in the ice and gas phase in astrochemical environments is of fundamental importance to predict the composition of planets and planetesimals (Öberg et al. 2009; Williams & Cieza 2011; Marboeuf et al. 2014). In the early stages of star formation the chemical distribution between ice and gas phase determines the chemical inventory that can be accessed for gas phase and surface chemistry on grains. In disks the balance between adsorption and desorption processes regulates which molecules can be incorporated into planets and planetesimals and which can only be delivered to planets through gas accretion. The temperature gradient present in disks results in the formation of condensation lines and therefore a radius dependent chemical distribution (Öberg et al. 2011; Bergin et al. 2015; Öberg & Bergin 2021). The location of these condensation lines depends on the strength of this physisorption interaction—the binding energy (BE)—between a molecule and a solid surface (i.e., water ice, silicate, and carbonaceous grains) and consequently its availability in condensed or gaseous phase at a particular disk radius.

Among the biogenic elements, the distribution of phosphorus (P) containing molecules between gas and solid phases is perhaps most uncertain. The availability of P is key to the formation of several biotic molecules (Pasek & Lauretta 2005) and it is quite abundant on Earth ( $\text{P}/\text{H} \sim 10^{-3}$ , Fagerbakke & Heldal 1996). In comparison the cosmic abundance of P relative to hydrogen is much lower ( $\text{P}/\text{H} \sim 2.57 \times 10^{-7}$ ,

Asplund et al. 2009). Phosphorus carried by  $\text{PN}$  and  $\text{PO}$  has been detected in the gas phase around evolved stars (Tenenbaum et al. 2007; Ziurys et al. 2007; Milam et al. 2008) and in star-forming regions (massive: Fontani et al. 2016; Rivilla et al. 2018; low mass: Yamaguchi et al. 2011; Lefloch et al. 2016; Bergner et al. 2019). In all these cases the P abundance is low accounting for a  $\text{P}/\text{H}$  of about  $10^{-10}$ – $10^{-9}$  (Lefloch et al. 2016; Rivilla et al. 2018; Bergner et al. 2019; Rivilla et al. 2020). In circumstellar envelopes, such as around IRC +10216, phosphorous, carried by  $\text{PH}_3$  and  $\text{HCP}$ , has been observed with abundances of  $10^{-8}$  with respect to molecular hydrogen (Agúndez et al. 2012). This accounts for about 7% of the phosphorous elemental abundance (Agúndez et al. 2014). These evidences suggest that the undetected phosphorous is likely incorporated into grains and that there must therefore be additional less-volatile carriers of P in the interstellar medium (ISM).

The nature of less-volatile P-containing compounds in the ISM is currently unclear, but solar system studies may provide some clues. Analysis of CI chondrites have shown an elemental P abundance similar to the solar phosphorus abundance (Lodders 2003). In stony meteorites most of the phosphorous is carried by Ca, Mg-phosphate minerals, while reduced phosphorous is more common in Fe-rich meteorites (Pasek et al. 2004). Very recently volatile phosphorous, mainly carried by  $\text{PO}$  fragments, was detected on comet 67P/ChuryumovGerasimenko during the Rosetta mission (Altwegg et al. 2016; Rubin et al. 2019; Gardner et al. 2020). This might suggest that, also in the ISM, P is incorporated into a relatively refractory phase, which likely consists of species containing  $\text{PO}_x$  moieties.

In this study we address the possible distributions of phosphorous oxides in the ISM and disks through a theoretical



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

investigation of the BEs of interstellar P carriers candidates with a  $\text{PO}_x$  moiety. To determine relevant BEs it is necessary to define a reasonable model system. Dust grains are composed of silicates aggregates (Jones et al. 2017) or carbonaceous materials. In molecular clouds and other cold and dense interstellar and circumstellar environments, dust grains are coated with ices that formed by the freeze-out of molecules in the gas phase, which can chemically evolve via surface chemistry. Due to its high abundance, water is the major constituent of the icy surface (Boogert et al. 2015) in clouds and throughout star and planet formation. As such, water ice represents the most relevant surface for the evaluation of the BE of volatile and semivolatile species. Additionally, the BEs on water surface may also serve to estimate the BE on minerals as mineral surfaces are often hydroxylated (Landmesser et al. 1997; Schaible & Baragiola 2014).

The BE of stable species can be determined in the laboratory via temperature programmed desorption (TPD) experiments (Collings et al. 2004; Fayolle et al. 2016; Chaabouni et al. 2020). In this study we use 20 astrochemically relevant species that have had their BEs determined through TPD studies to evaluate our computational approach. These species are generally stable small molecules, which include C, H, N, and O atoms. Some examples of experimentally measured BE of S-bearing species are also available (i.e.,  $\text{H}_2\text{S}$ ), while there is lack of information in regard to P-bearing molecules, especially unstable ones.

In the case of unstable and exotic species, the BE needs to be estimated computationally. A computational challenge is the accurate representation of the amorphous solid water (ASW). Several studies have focused on a periodic representation of the solid water surface (Karssemeijer & Cuppen 2014; Karssemeijer et al. 2014; Zamirri et al. 2019; Ferrero et al. 2020); In these studies the BE of a molecule is estimated from its interaction with a sizable ASW surface that contains multiple binding sites. As shown in Ferrero et al. (2020) these approaches can yield a range of BEs for each molecule dependent on the optimized binding site, thus providing a distribution of BEs to ASW for each molecule.

The ASW can also be approximated by medium- to large-sized water clusters. Shimonishi et al. (2018) used 20 water molecules to define the ASW geometry allowing for the definition of multiple binding sites. Very recently, Germain et al. (2022) developed a computational method to build a large (1–200  $\text{H}_2\text{O}$ ) ASW cluster to determine accurate BE distribution ranges for molecules on ASW.

Other studies have focused on a small-cluster representation of the ASW, which are more computationally affordable. Although the use of a small cluster does not provide for a comprehensive account of long-range molecule–ASW interactions and for periodic variation of the ASW, it constitutes a lighter computational investment for each new molecule while still providing BE estimations well in range of both the periodic representation studies and the experimental value (Wakelam et al. 2017; Das et al. 2018; Ferrero et al. 2020). The optimal size for a small cluster is still debatable. Das et al. (2018) tested the cluster size of 1, 3–6  $\text{H}_2\text{O}$  and showed that the uncertainty on the BE is reduced as the cluster size increases. On the other hand, single water cluster systems may provide similar uncertainties to BE estimated using periodic systems, when the calculated values are calibrated using experimental values (Wakelam et al. 2017).

In this work we aim to computationally constrain the BEs of three known ISM P carriers as well as seven proposed interstellar P carriers over ASW. We will use two approaches described in detail in Section 2: direct calculation and calculations calibrated against experimental data on C-, N-, O-, and S-bearing molecules. The results of the two methods are presented in Section 3 where we also discuss the reliability of these approaches. In Section 4 we present and comment on the application of these methods to P-bearing molecules. We include some astrophysical implications of our new BEs for P-bearing molecules (Section 4.3). Finally, in Section 5 we summarize our findings.

## 2. Methods

### 2.1. Computational Details and Cluster Size Choice

When computing BEs there are several aspect of the process, such as the computational tools, the molecular approximations, and data treatment, for which choices need to be made.

To model the binding interactions we chose to use electronic structure based methods. In particular, within the density functional theory (DFT) we chose to use the M06-2X functional for its good performances in modeling noncovalent interactions (Mardirossian & Head-Gordon 2017). All calculations are run using the Gaussian 16 suite of software (Frisch et al. 2016) at the M06-2X/aug-cc-pVDZ level of theory (Dunning 1989; Kendall et al. 1992; Zhao & Truhlar 2008) and included the optimization of the clusters geometry to a stationary point as well as vibrational frequency calculations for the identification of the energy minima.<sup>1</sup> Additionally we tested the relative performances of MollerPlesset (MP) methods (Frisch et al. 1990), in particular at the MP2/aug-cc-pVDZ level of theory, for BE determination and we concluded that the two methods yields equivalent results (Appendix A). Previous work by Wakelam et al. (2017) and Ferrero et al. (2020) also chose the M06-2X functional for the estimation of BEs.

Concerning the description of the surface–molecule interaction, we chose to focus our work on cluster systems rather than on a periodic representation of the water surface due to its computational affordability. The use of small-cluster systems rather than a more extended periodic representation of the binding surface could lead to less accurate results. In the work from Ferrero et al. (2020) the authors show, however, that the BEs calculated using small water clusters are comparable to more expensive and/or complex periodic calculations. In our study the cluster size was purposefully kept small in the attempt to minimize the computational cost to enable easy scaling to larger molecular data sets. Inspired by the work of Das et al. (2018), who found an improvement in accuracy going from the monomeric to the tetrameric representation of ASW; while 5 and 6  $\text{H}_2\text{O}$  clusters did not further improve the accuracy, we evaluate the impact that the size of the water cluster has on the BE estimation by using 1  $\text{H}_2\text{O}$ , 2  $\text{H}_2\text{O}$ , and 3  $\text{H}_2\text{O}$  water cluster sizes. We also follow Wakelam et al. (2017) and explore whether directly calculated BEs can be improved through calibration against experimental values using a molecular training set. We extend this work by applying a calibration to the BEs calculated using 1–3  $\text{H}_2\text{O}$  water cluster sizes. To obtain a calibration set of molecules, we searched the literature for experimental data on BEs to ASW surfaces and identified

<sup>1</sup> Data set is available at doi:10.5281/zenodo.6551710.

**Table 1**  
Calculated and Experimentally Determined Binding Energies (in Kelvin) for Our Reference Molecules

M06-2X		Direct Method			Calibrated Method			Exp. <sup>b</sup>
aug-cc-pVDZ								
Species		1H <sub>2</sub> O	2H <sub>2</sub> O	3H <sub>2</sub> O	1H <sub>2</sub> O	2H <sub>2</sub> O	3H <sub>2</sub> O	
(1)	N <sub>2</sub>	530	998	953	1803	1723	1898	1125 <sup>[1,2]</sup>
(2)	CO	504	1191	1107	1769	1867	2048	1165 <sup>[1,2,3,4]</sup>
(3)	CH <sub>4</sub>	672	1212	1168	1990	1882	2108	1370 <sup>[2,5]</sup>
(4)	CO <sub>2</sub>	1733	3104	2259	3395	3289	3170	2339 <sup>[4,5,6]</sup>
(5)	C <sub>2</sub> H <sub>4</sub>	1498	2760	2294	3084	3034	3204	2400 <sup>[7]</sup>
(6)	C <sub>2</sub> H <sub>6</sub>	960	1600	1321	2372	2171	2256	2495 <sup>[2,7]</sup>
(7)	H <sub>2</sub> S	1641	3606	3154	3273	3663	4042	2519 <sup>[8,9]</sup>
(8)	C <sub>2</sub> H <sub>2</sub>	1388	3458	2413	2938	3552	3320	3000 <sup>[7]</sup>
(9)	H <sub>2</sub> CO	3076	4623	3832	5172	4419	4702	3260 <sup>[9,10]</sup>
(10)	C <sub>3</sub> H <sub>8</sub>	976	1652	1606	2394	2210	2534	3500 <sup>[2,7]</sup>
(11)	C <sub>3</sub> H <sub>6</sub>	2071	2871	2629	3842	3116	3531	3800 <sup>[7]</sup>
(12)	CH <sub>3</sub> NH <sub>2</sub>	3660	6414	2065	5944	5750	2982	4269 <sup>[11]</sup>
(13)	CH <sub>2</sub> CCH <sub>2</sub>	1792	2957	2145	3472	3180	3059	4400 <sup>[7]</sup>
(14)	CH <sub>3</sub> CCH	2280	3422	2251	4118	3525	3162	4400 <sup>[7]</sup>
(15)	HCl	2836	5225	4743	4854	4866	5589	5170 <sup>[12]</sup>
(16)	CH <sub>3</sub> OH	2584	6183	3826	4520	5579	4696	5410 <sup>[13]</sup>
(17)	NH <sub>3</sub>	3464	5949	4805	5685	5405	5649	5530 <sup>[14]</sup>
(18)	CH <sub>3</sub> NC	2352	4484	3680	4213	4315	4555	5686 <sup>[15]</sup>
(19)	H <sub>2</sub> O	2633	6007	5416	4586	5448	6245	5773 <sup>[16]</sup>
(20)	CH <sub>3</sub> CN	2445	5093	4148	4337	4769	5010	6150 <sup>[15]</sup>

**References:** (1) Fayolle et al. (2016), (2) Smith et al. (2016), (3) Collings et al. (2003), (4) Noble et al. (2012a), (5) He et al. (2016), (6) Gálvez et al. (2007), (7) Behmard et al. (2019), (8) Wakelam et al. (2017), (9) Penteadó et al. (2017), (10) Noble et al. (2012b), (11) Chaabouni et al. (2020), (12) Olanrewaju (2011), (13) Bahr et al. (2008), (14) Hama & Watanabe (2013), (15) Bertin et al. (2017), (16) Fraser et al. (2001), (17) Fraser et al. (2001).

20 molecules with well-defined experimental BEs (Table 1). The same literature data were also used to benchmark the performance of our computational results by direct comparison to the experimental values.

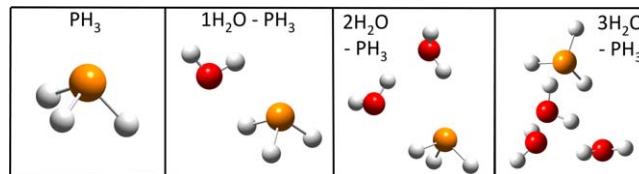
Concerning the uncertainties of the experimental BEs, we apply a 10% uncertainty to the BEs of species for which an uncertainty value of <10% was reported in the literature and we conservatively consider an uncertainty of 30% when the uncertainties were not reported in the literature (see Table 1).

The direct method, which estimates the BEs from the energetics of 1H<sub>2</sub>O, 2H<sub>2</sub>O, and 3H<sub>2</sub>O water cluster sizes calculation is described in detail in Section 2.2. Section 2.3 describes the application of the calibrated method to the direct method data.

## 2.2. Direct Method

We first calculate the BEs at different cluster sizes directly. The electronic energy of each molecule, water cluster, and water-molecule complex was calculated at the M06-2X/aug-cc-pVDZ level of theory for the 1, 2, and 3 H<sub>2</sub>O cluster representation. The geometry of the water-molecule complexes is optimized starting from a noninteraction configuration with the molecule placed at a minimum distance >5 Å from the water cluster. This is to ensure that no molecule-cluster interactions are present in the input geometry. For most molecules we performed a single optimization calculation (single initial configuration), which yielded geometries in agreement with the criteria described below. For molecules that allow multiple unique interaction configurations we repeated the optimization with two to three initial configurations.

The following criteria were used to chose the representative geometry: In the 1H<sub>2</sub>O clusters the representative geometry was chosen as the one where the main interaction was between H<sub>2</sub>O



**Figure 1.** Example of optimized geometries of the species, 1H<sub>2</sub>O systems, 2H<sub>2</sub>O systems, and 3H<sub>2</sub>O systems used in this work.

and the molecule. This is because the ASW surface is more rich in hydrogen than oxygen atoms and therefore the H<sub>2</sub>O-molecule interaction is more likely to occur (Wakelam et al. 2017). Generally the H<sub>2</sub>O interacts with an atom of the molecule but in the cases of unsaturated hydrocarbons the interaction is set between the H<sub>2</sub>O and the double or triple bond on the carbon chain. Exceptions to this criteria are CO<sub>2</sub>, for which the main interaction is always set between the O<sub>H<sub>2</sub>O</sub> and the C<sub>CO<sub>2</sub></sub>, and CH<sub>3</sub>OH, which sees the main interaction occurring between the hydroxylic H<sub>CH<sub>3</sub>OH</sub> and the O<sub>H<sub>2</sub>O</sub>. These exceptions in the final geometry are a result of the optimization calculation and are likely due to the nature of the molecules. The 2H<sub>2</sub>O and 3H<sub>2</sub>O complex geometries are chosen in a similar way while allowing for the second and third interaction to occur between the O<sub>H<sub>2</sub>O</sub> and the molecules. An example of the optimized geometries is shown in Figure 1 for the molecule PH<sub>3</sub>. A few more optimized geometries are shown in Appendix B.

In the direct method, the adsorption energies of the molecules in question is determined from the variation of the energy of an adsorbent molecule and an H<sub>2</sub>O cluster (in our case constituting 1–3 H<sub>2</sub>O water molecules) that arises when they are able to noncovalently coordinate with each other. The

**Table 2**  
Binding Energies of Selected P-bearing in Kelvin

M06-2X aug-cc-pVDZ Species	Direct Method (K)			Calibrated Method (K)		
	1H <sub>2</sub> O (±51.6%)	2H <sub>2</sub> O (±19.2%)	3H <sub>2</sub> O (±22.4%)	1H <sub>2</sub> O (±27.2%)	2H <sub>2</sub> O (±25.3%)	3H <sub>2</sub> O (±27.9%)
PH <sub>3</sub>	1117	2642	1228	2579	2945	2166
PN	2168	5119	2376	3970	4787	3284
PO	3791	9176	9221	6117	7804	9951
HPO	4176	9964	5304	6626	8390	6136
OPO	4327	10186	11031	6826	8555	11714
POOH	625	9512	4551	9382	8054	5403
PO <sub>2</sub> OH	7254	12290	<b>25502<sup>a</sup></b>	10698	10119	<b>25809<sup>a</sup></b>
HPO <sub>2</sub>	7457	12740	15388	10967	10454	15958
H <sub>2</sub> PO	9118	11112	8858	13164	9244	9597
PO <sub>3</sub>	11039	18343	<b>30441<sup>a</sup></b>	15706	14621	<b>30619<sup>a</sup></b>

**Note.**

<sup>a</sup> These structures react with the ASW to form a PO<sub>4</sub> moiety. The energy reported refer to a chemisorption events.

BE is calculated as follows:

$$BE = E_{\text{complex}} - (E_{\text{molecule}} + E_{\text{H}_2\text{Ocluster}}).$$

Where  $E_{\text{complex}}$  is the energy of one of the potential energy minima geometries of the water-molecule cluster when the species is physisorbed onto the surface of the cluster (2–5 Å),  $E_{\text{molecule}}$  is the energy of the species alone, and  $E_{\text{H}_2\text{Ocluster}}$  is the calculated energy of the water cluster.

The calculated electronic energies are used directly without accounting for the zero-point energy (ZPE) and the basis set superposition error (BSSE) similarly to Wakelam et al. (2017). Wakelam et al. (2017) tested whether the inclusion of ZPE and or BSSE significantly affected the accuracy of the resulting fit and found that, in the dimer case, the inclusion of the corrections slightly reduced the goodness of the fit. This suggests that, while the omission of either the ZPE or the BSSE may significantly affect the calculated energies, the omission of both corrections in the estimation of BEs yields values that better approximate the experimental BE (Das et al. 2018). This errors compensation is likely just a fortuitous but nonetheless advantageous balance. The direct method is evaluated for a reference set of 20 molecules (Table 1) and then applied to calculate the BEs of 10 P-bearing species (Table 2).

### 2.3. Calibrated Method

Wakelam et al. (2017) demonstrated that there is a systematic offset between calculated and experimentally determined BEs when the BEs are calculated using a single-water-molecule cluster. They also show that the accuracy of the calculated BE estimates can be improved if they are calibrated against experiments. Wakelam et al. (2017) calculated, using the equation in Section 2.2, the BEs of 16 molecules using a 1H<sub>2</sub>O ASW representation to build a calibration curve against the experimental BE of each molecule. From the fit, they evaluated the interaction correlation between a 1H<sub>2</sub>O representation and the ASW. Building upon the work of Wakelam et al. (2017), we extended the method by applying the calibrated method to the 1–3 H<sub>2</sub>O systems to evaluate the effects of the increased cluster size on the performances of the calibrated method. We also evaluate whether the use of the calibrated method provides a significant improvement in the BE estimation over the direct method.

The calibration curves were built by fitting the 20 BE's (20 for each of the 3 water cluster systems) obtained using the direct method against the experimental BE's values assuming a linear relationship between calculated and experimental values. The fit was then applied to the BE's calculated with the direct method resulting in the calibrated method estimation of the BE values. The calibrated method was then applied to estimate the BEs of the P species.

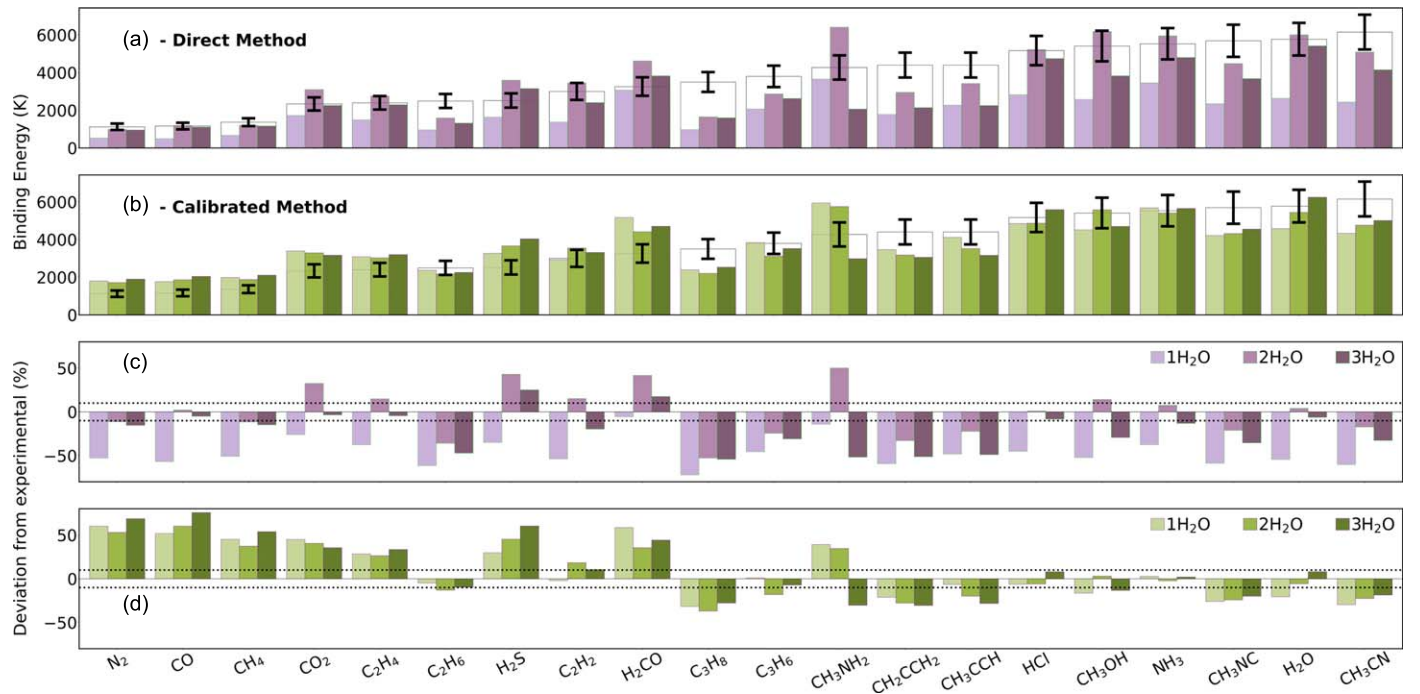
### 3. Methods Validation

Table 1 summarizes the calculated direct method and estimated calibrated method BE values for the calibration molecules. Figure 2 shows the BEs calculated using the direct method and the calibrated method as well as the percentage deviation from the experimental values. Below we present these results in detail. In Section 4 we present the application of the two methods to our selection of phosphorous molecules.

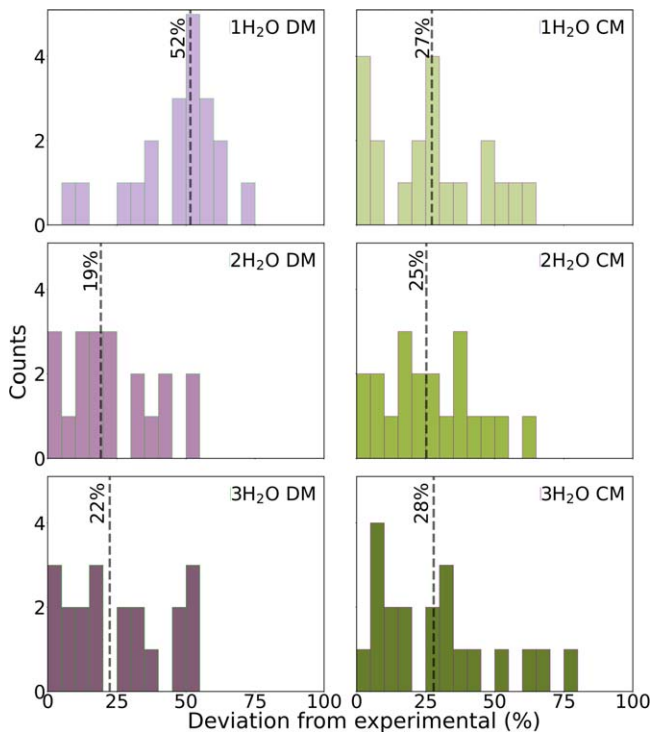
#### 3.1. Direct Method H<sub>2</sub>O Cluster versus ASW Binding Energies

Figure 2 and Table 1 show that the accuracy of the calculated direct method BEs is highly variable both between different molecules, and when using different cluster sizes. Calculations accounting for 1H<sub>2</sub>O interaction performed the worst, with deviations as high as 72% and a median deviation of 52%, while the 2 and 3H<sub>2</sub>O clusters performed better with deviations between a few percent and 54% and median deviations of 19% and 23%, respectively. These deviations can be compared to typical experimental errors of 10%. The performance improvement when increasing the cluster size from one to two or three H<sub>2</sub>O molecules is visualized in Figure 3, which shows histogram plots of the deviations in percentage from experimental values. In other words, there is a real increase in performance when increasing the cluster size from one to two, but not when increasing it to three for our sample of molecules. This suggests that increasing the cluster size beyond two H<sub>2</sub>O molecules may be of limited value when the focus is to computationally determine the mean value of a molecule BE using the cluster approach. However, this needs to be confirmed for a larger and more diverse sample of molecules, as well as for a larger range of cluster sizes. While the small-cluster approach is a useful and computationally inexpensive tool to determine the mean BE values, it provides limited information on the BE distribution and therefore cannot be





**Figure 2.** Binding energies (in kelvin) of the 20 chemical species used in this study calculated using three different H<sub>2</sub>O cluster sizes and direct method (panel (a)) and calibrated method (panel (b)) in relation to the experimental BE of the same species (black contour, 10% error bars). The deviation from the experimental values are shown in panel (c) and (d) respectively.



**Figure 3.** Bin distribution of the deviation from the experimental value of the computational BEs of 20 reference molecules calculated using the direct method (DM) and the calibrated method (CM) using a 1–3 H<sub>2</sub>O representation of the ASW surface. The dotted lines highlight the median values of the distributions.

applied in cases when the whole range of the BE distribution is of interest.

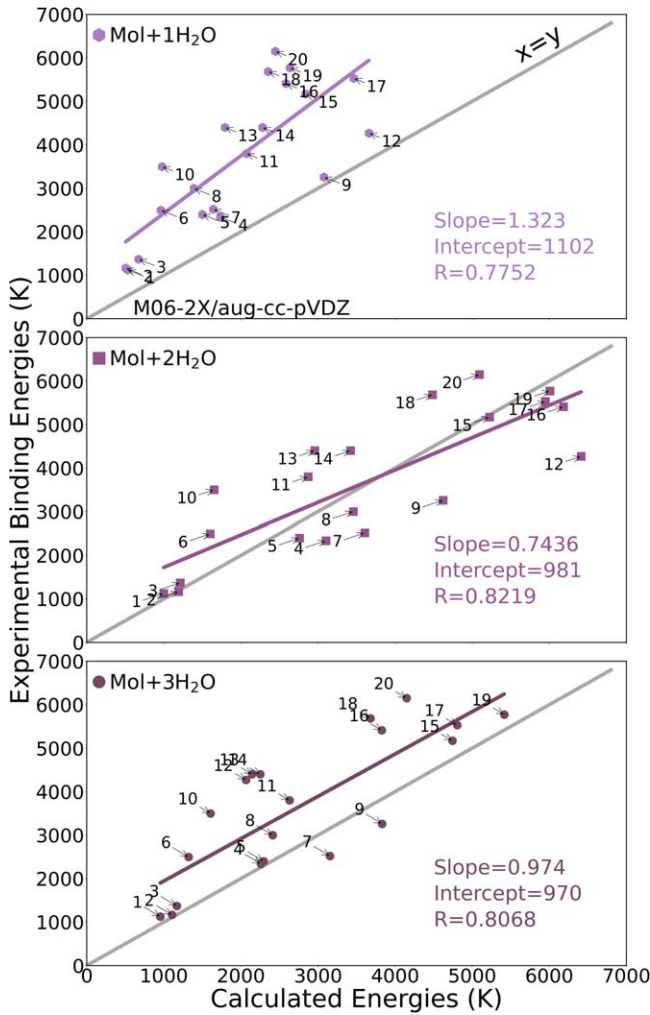
Interestingly the performance of the direct method approach appears to depend on the strength of the binding interaction.

Figure 2 shows that for molecules up a volatility of  $\sim 3000$  K the median deviations are only  $\sim 15\%$  for the 2H<sub>2</sub>O and 3H<sub>2</sub>O cluster calculations, while there is almost a 1.5 factor increase when considering the less-volatile species. There are also evidences that the accuracy of the direct method approach depends on the chemical nature of the molecule. The BEs prediction for hydrocarbons (i.e., C<sub>2</sub>H<sub>6</sub>, C<sub>3</sub>H<sub>8</sub>, C<sub>3</sub>H<sub>6</sub>, CH<sub>2</sub>CCH<sub>2</sub>, CH<sub>3</sub>CCH) seem to be less accurate than for other molecules with a deviation from the experimental values between 25% and 53% for the 2H<sub>2</sub>O cluster size (Figure 2, panels (a) and (c)).

Finally we note that there are a handful of molecules for which the deviation from experimental values increases between the 1H<sub>2</sub>O and 2H<sub>2</sub>O clusters; namely, CO<sub>2</sub>, H<sub>2</sub>S, H<sub>2</sub>CO, and CH<sub>3</sub>NH<sub>2</sub>. These molecules warrant further investigation, since this may be revealing something interesting about their interactions with water ice. For now we simply note that in each of these cases the BE proceeds from being slightly underpredicted in the case of 1H<sub>2</sub>O to overpredicted for the 2H<sub>2</sub>O cluster.

### 3.2. Consideration on the Direct Binding Energies Calculation

A first measure of the accuracy of the direct method BE calculations is the deviation from experimental values. As reported above the deviation from experimental BEs decreases from 52% to 19% (Figure 3) when increasing the cluster size from 1 to 2H<sub>2</sub>O, and no further improvement is seen when increasing the cluster size from 2 to 3H<sub>2</sub>O (22% median deviation). The improvement when going from 1 to 2/3H<sub>2</sub>O clusters is in line with the results from Das et al. (2018), who found an improvement in the deviation from the experimental values from 40% to 25% when increasing the cluster size from 1 to 3H<sub>2</sub>O molecules (2H<sub>2</sub>O cluster calculations were not included in the study). Das et al. (2018) also reported a consistent underestimation of the BE values compared to the

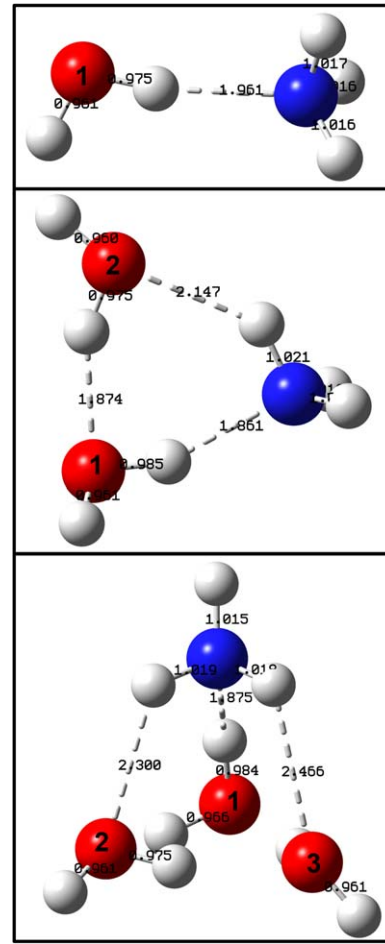


**Figure 4.** Linear fit of the 20 BEs calculated using the direct method for each of the 1–3H<sub>2</sub>O cluster systems against the experimental values from the literature. The data point numbers correspond to the indexes of the molecules in Table 1.

experimental values, which is present for both the monomeric and the trimeric representation of the ASW, additionally showing a trend of progressively less negative estimations as the cluster size increases. Our calculations for the 1 and 3H<sub>2</sub>O clusters show a similar trend where the 3H<sub>2</sub>O system generally underestimates the experimental BEs while providing a substantial improvement from the BEs derived from the 1H<sub>2</sub>O system.

However, we find that the 2H<sub>2</sub>O system does not fit in the same trend, underestimating only 45% of the BEs studied. The randomness of this distribution along with the good absolute median uncertainty of 19% suggests that, among the cluster size and geometries that are considered in this study, the 2H<sub>2</sub>O description of the ASW surface provides the best BEs prediction.

The improvement in the results when using 2/3H<sub>2</sub>O clusters versus 1H<sub>2</sub>O is intuitively due to the accounting for additional interactions between the molecule and the water cluster. The presence of more water molecules allows for additional binding constraints between the molecule and the water cluster yielding a binding geometry that better resembles the binding configuration on the ice. We can test this intuition by interrogating the molecule–cluster systems in detail. We find that the addition of



**Figure 5.** Optimized binding geometries (M06-2X/aug-cc-pVDZ) of NH<sub>3</sub> in combination with the 1, 2, and 3 H<sub>2</sub>O representation of the ASW surface. Bond lengths and long-range interactions (dotted lines) are in angstroms.

a second water molecule results in a geometry where the primary water molecule can bond more strongly with the molecule in question. Since in an ice system there are always neighboring molecules, accounting for this distortion is important to produce accurate BEs. Not accounting for this geometrical distortion yields underestimated BE: this is consistent with the underestimation (Figure 4) of the BEs in the 1H<sub>2</sub>O system where not all the fundamental interaction between the species and the ASW can be taken into account.

Figure 5 shows an example (NH<sub>3</sub>) of the effects on the molecules binding environment caused by the second H<sub>2</sub>O. The main interaction between the ASW and NH<sub>3</sub> is between the nitrogen on NH<sub>3</sub> and one of the water hydrogens. In the case of 2H<sub>2</sub>O the interaction distance is shorter (1.96 and 1.86 Å respectively in the 1 and 2H<sub>2</sub>O systems). A consequent elongation of the O–H bond in the primary water molecule is also observed. The second water molecule does not interact as strongly with the NH<sub>3</sub> but provides an additional anchoring point (2.15 Å) resulting in the reduction of the N–(H–O)<sub>H<sub>2</sub>O</sub> angle from 171° to 159°. A similar behavior is generally observed across the studied molecules.

Following the same intuition as above, we should observe an improved BE accuracy when increasing the H<sub>2</sub>O cluster size from 2 to 3H<sub>2</sub>O, but this is not what we find. The presence of the third water molecule introduces additional structural constraints (2.46 Å, Figure 5), which cause a weakening of

the primary and the secondary interaction between the molecule and the water. This results in a BE prediction for the 3H<sub>2</sub>O system very similar to the 2H<sub>2</sub>O system with the absolute error for ammonia going from 8% in the 2H<sub>2</sub>O to 13% in the 3H<sub>2</sub>O. The observed lack of increased precision when increasing the cluster size from 2 to 3 H<sub>2</sub>O molecules is surprising. It is typically expected for larger clusters to increase the accuracy of the BE estimation as more long-range interactions can be taken into account and less unique geometries become available. One possible explanation is that the use of a greater number of water molecules introduces significant freedom in regard to the arrangement of the water molecules themselves and this may produce geometries at odds with ASW. In our study we found that the BE estimations are dependent on the functional group that the molecule uses to bind to the water cluster. This is most apparent when comparing the isomers CH<sub>3</sub>CN and CH<sub>3</sub>NC (see Figure 9 in the Appendix B). For CH<sub>3</sub>CN and CH<sub>3</sub>NC the BE estimates reflect the interactions of the water with either the N or C atoms that are terminal to the molecule. In the case of CH<sub>3</sub>NC we have that the interaction distance is 2.18 Å, which becomes 2.07 Å for CH<sub>3</sub>CN, consequently increasing the BE of the molecule. This is consistent with the experimental BE and it reflects the affinity of the water for the functionalities present in the molecules. The binding distance is further shortened for molecules having a terminal oxygen; in the case of H<sub>2</sub>CO for example, the O<sub>CH<sub>2</sub>O</sub>-H<sub>2</sub>O distance is 1.98 Å.

We find a similar functionality dependency of the BE to water in cases where the main binding interaction occurs between the H<sub>2</sub>O and the double/triple bond in the molecules. In the case of CH<sub>2</sub>CCH<sub>2</sub>, and CH<sub>3</sub>CCH, for example, we find that CH<sub>3</sub>CCH binds strongly to the water cluster when compared to CH<sub>2</sub>CCH<sub>2</sub>. This is because the triple bond in CH<sub>3</sub>CCH constitutes a better binding functionality for hydrogen than the double bond in CH<sub>2</sub>CCH<sub>2</sub>. This effect of the molecule saturation on the BE prediction is observed also in the C2 and C3 series of hydrocarbons with smaller BEs as the saturation of the molecule increases (Table 1). This same trend has also been studied experimentally by Behmard et al. (2019) who observed a similar trend showing that the BE of C2 and C3 hydrocarbons decreases with the saturation of the molecules (i.e., BE<sub>C<sub>2</sub>H<sub>2</sub></sub> > BE<sub>C<sub>2</sub>H<sub>4</sub></sub> > BE<sub>C<sub>2</sub>H<sub>6</sub></sub>). In general, the application of the direct method to hydrocarbons seems to yield a worse approximation of their BEs compared to other molecules in the study. The median deviation for nonhydrocarbons is 16% and for hydrocarbons 23%, in the 2H<sub>2</sub>O direct method calculation. The underestimation of the BE of hydrocarbons may be associated to the poor natural affinity that this class of molecules has toward water. In such cases the size of the cluster may be more relevant than for other classes of molecules and the use of a periodic representation of the ASW may be of aid. The presence of a matrix may help producing a tighter packing of the water around the molecule, which would result in a binding structure that better resembles the experiments. The poor constrain that the direct method provides for hydrocarbons suggests that this class of molecules may also especially benefit from the calibrated method.

### 3.3. Calibrated Method Improvements over the Direct Method

We next evaluate whether the results obtained through the direct method can be improved upon application of the calibrated method using the same 20 molecules. Figure 4

shows the linear fit of the BE calculated using the direct method and the experimental values. The calculated BEs using 1 and 3H<sub>2</sub>O generally underestimate the BE. This effect is more pronounced and less uniform for the 1H<sub>2</sub>O compared to the 3H<sub>2</sub>O data set. The use of 2H<sub>2</sub>O appears to instead produce a close to random scatter around the expected values. We note, however, that as we go from more- to less-volatile molecules the 2H<sub>2</sub>O cluster method seems to be systematically over-predict the BEs.

We find that the calibrated method BE predictions are very similar to each other for all the cluster sizes. By contrast to the direct method, we find no improvement in the median deviation from the experimental value as the cluster size is increased (Figure 3). For 1, 2, and 3H<sub>2</sub>O clusters we find median deviations of 27%, 25%, and 28%, and deviation ranges of 1%–60%, 2%–60%, and 2%–69%, respectively. This implies that the calibrated method generally achieves a higher level of accuracy when considering single-water-molecule clusters, but a comparable level of accuracy when considering larger clusters in comparison with the direct method.

Using the calibrated method we see an opposite dependence of the results accuracy with the molecule volatility compared to the direct method (Figure 2, panel (d)). For volatile molecules having BE <3000 K, the median deviation from the experimental values calculated using the calibrated method–2H<sub>2</sub>O cluster size is 37%. The median deviation is reduced to 19% for the less-volatile group of molecules, which implies that the calibrated method outperforms the direct method for the more-refractory molecules. It also appears to do better with the 2 and 3C hydrocarbons compared to the direct method.

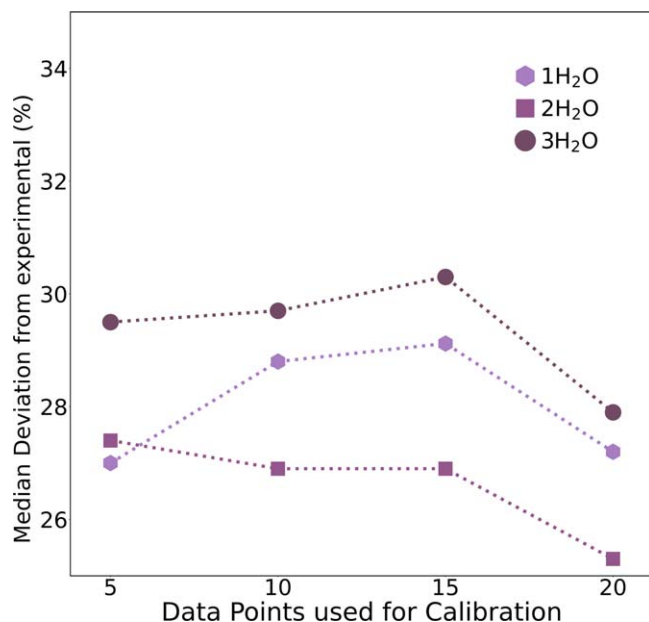
Additionally we tested the importance of the number of calibration points used to estimate the correlation for uncertainties minimization; we performed 1000 random selections of 5, 10, and 15 molecules from our list of 20 molecules and for each set we used the resulting fits to determine the BEs of the molecules not included in the selections. We finally calculated the median of the errors across the 1000 draws for each of the 5, 10, and 15 molecule sets (Figure 6). The median deviation decreases as the calibration set increases, but the improvement is small: about 2%. Though it will be interesting to revisit the calibrated method approach with more experimentally determined values, we may already be close to the limit where the peculiarities of each molecule–H<sub>2</sub>O system dominates the calculated uncertainty.

### 3.4. Utility of Calibrating Calculated Binding Energies?

The application of the calibrated method is more or less advantageous depending on the cluster size used and the volatility of the molecules in the exam. Without a calibration the 1H<sub>2</sub>O BEs are too inaccurate to be useful, and we hence recommend that such a calibration is always used for 1H<sub>2</sub>O BE calculations. The application of the calibrated method to the 2 and 3H<sub>2</sub>O cluster sizes do not appear to contribute to the accuracy of the results except for some specific classes of molecules.

The improvements to the predicted BEs observed when applying the calibration to the less-volatile species in contrast with the loss of accuracy for highly volatile molecules (<3000 K) suggests that the calibrated method does not, on average, improve the results. However, we suspect that the calibration could improve the results if applied in a more targeted way. More experimental BE values are needed to separately calibrate volatile and nonvolatile species.





**Figure 6.** Median of the deviation from the experimental values for the 1–3 H<sub>2</sub>O systems calculated after applying a 5, 10, 15, and 20 data point calibration.

In a similar way, we noticed that for some class of molecules such as hydrocarbons the BEs seem to be difficult to estimate compared to other types of molecules. This reduced accuracy for hydrocarbons further suggests that a blind calibration using heterogeneous collections of molecules might not be the best choice for all class of compounds. It is possible, however, that the use of more narrowly defined families of molecules, also in combination with the use of bigger cluster sizes, could improve the results.

The availability of more experimentally measured BEs will be crucial for the effective computational prediction of the BEs of species that show chemical class dependencies and the consequent application of computational methods to estimated the BEs of species that are not of easy access in a laboratory.

### 3.5. Result Summary and Recommendation

Overall we find that the direct method results obtained using the 2H<sub>2</sub>O cluster size provide BEs with the lowest median uncertainty on the whole range of reference molecules of 19%. The application of the calibrated method only improved on the 2H<sub>2</sub>O cluster size for the least-volatile molecules and for 2C and 3C hydrocarbons. The cutoff appears to be around a BE of 3000 K, but a larger molecular reference set is needed to explore this further.

Given these results, we recommend using the direct method 2H<sub>2</sub>O for volatile species and the calibrated method 2H<sub>2</sub>O for semivolatile and more-refractory species to determine the mean BE value using small water clusters.

## 4. Application of the Methods to P Molecules

### 4.1. BE of P-bearing Molecules

We calculated the BEs of 10 phosphorous molecules (Table 2) using the direct method and the calibrated method. The uncertainties on the calculated BEs values are derived from the median deviation from the experimental value obtained respectively for each of the methods (Figure 3). When estimating

the uncertainties for the calibrated method of the P molecular BEs below, we use the median deviations for the fiducial 20-molecule calibration set. As expected from the calibration set, the direct method 3H<sub>2</sub>O systems estimations provides, for most of the P-bearing molecules, BEs values that are in between the values obtained using the 1H<sub>2</sub>O and the 2H<sub>2</sub>O direct method calculations (Table 2). Exceptions are OPO, and HPO<sub>2</sub>, for which the direct method 3H<sub>2</sub>O BE values are higher than the 2H<sub>2</sub>O prediction. Additionally, the geometry optimization of PO<sub>2</sub>OH and PO<sub>3</sub> with 3H<sub>2</sub>O results in the coordination of the P species to one oxygen from the water cluster yielding a PO<sub>4</sub> moiety (see Appendix C). This complexation prevents the calculation of the physisorption energy for these two P molecules for the 3H<sub>2</sub>O cluster system. We note that the complexation is observed only for the 3H<sub>2</sub>O cluster setting; in all other cluster size neither covalent interactions nor deformation of the water cluster geometries are observed.

Similarly to what was observed for the reference set of molecules we find that the also in the case of P-bearing molecules the application of the calibrated method estimates BEs values that are equivalent across the three cluster systems. We also find similarities between the fiducial set and the P-bearing set of molecules when comparing the two methods' performance for each cluster size. In the case of 1H<sub>2</sub>O cluster size we find that, similarly to the fiducial set of molecules, the calibrated method estimations have higher BE values than the direct method by a factor of  $\sim 2.4$  for low-desorbing species, and  $\geq 1.5$  for the less-volatile species. The 2H<sub>2</sub>O cluster size yields a smaller discrepancy between the direct method and calibrated method prediction compared to the 1H<sub>2</sub>O cluster size with the 2H<sub>2</sub>O estimation yielding values within 15% of the direct method estimation. In the cases when the 3H<sub>2</sub>O clusters size did not yield additional coordination chemistry, the calibrated method estimation relates to the direct method estimation by a 1.8 and 1.2 factor respectively for volatile and less-volatile species.

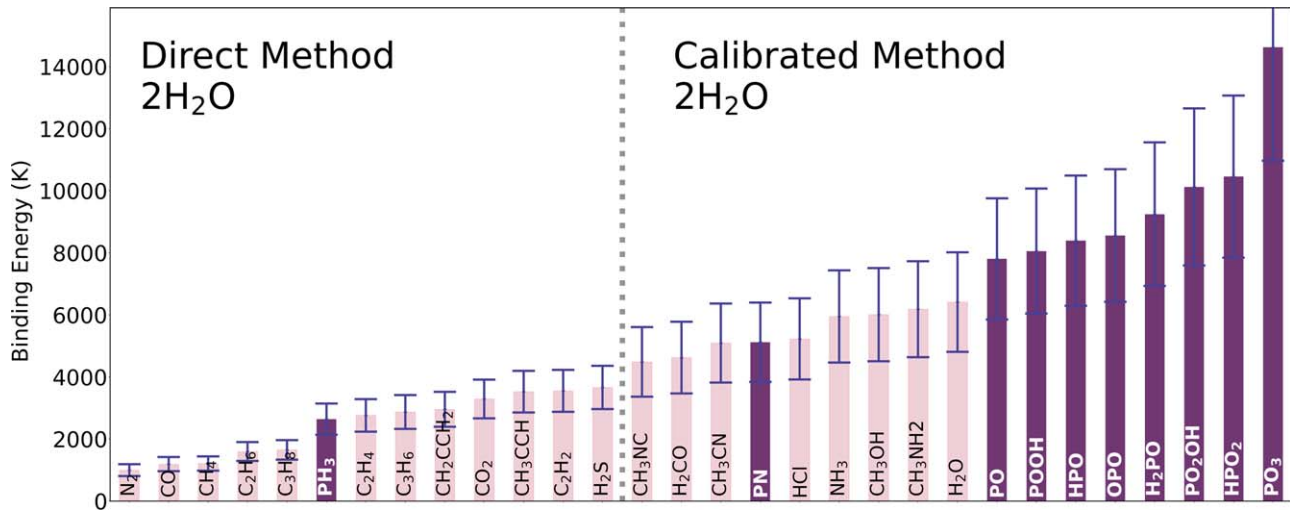
In Figure 7 the BE values calculated for the P-bearing species are shown in relation to the BEs calculated for the reference set of molecules using the 2H<sub>2</sub>O system. With the exception of PH<sub>3</sub>, PO, and PN the estimated BEs of the P-bearing species are found to be quite high (BE  $\sim 8000$  K). With a most of the molecules exceeding the range of the fiducial set of molecules (PO<sub>2</sub>OH, HPO<sub>2</sub>, H<sub>2</sub>PO, and PO<sub>3</sub>) with BE  $> 9000$  K.

Considering the results obtained for the calibration set of molecules, we divided the P-bearing species into three groups based on their BEs. The first group consists of PH<sub>3</sub> alone, which is the only highly volatile molecule with a BE below 3000 K when calculating energies for the P molecules. PH<sub>3</sub> BE is estimated using the direct method. PN, PO constitute the second group, defined by BEs similar to or lower than water ice (Fraser et al. 2001). For these molecules we recommend using the calibrated method results. In the highest range of BEs we find the refractory group constituted by HPO, OPO, POOH, H<sub>2</sub>PO, PO<sub>2</sub>OH, HPO<sub>2</sub>, and PO<sub>3</sub> for which the BEs exceed 8000 K. For these too we recommend using the calibrated method 2H<sub>2</sub>O but caution that the results are more uncertain since they extend beyond the calibration set.

### 4.2. P-bearing Binding Energy Calculation

We discuss the P molecule results by volatility grouping. The *volatile* species, PH<sub>3</sub>, PN, and PO, BEs fall within the





**Figure 7.** Binding energy values, in kelvin, calculated using the direct method (low-range BE) and calibrated method (high-range binding energy) in combination with the  $2\text{H}_2\text{O}$  cluster of both the reference species and the P-bearing species. We reported uncertainties of 19% for the direct method results and of 25% for the calibrated method results (Figure 3). Species are plotted in order of increasing calculated BE.

range of the experimentally determined BEs that we evaluated our methods against. We therefore expect that our error estimations for these species are reliable. The  $\text{PH}_3$  BE has been previously computed by Nguyen et al. (2021) at 1813–2690 K; this range is consistent with our calculation of  $2642\% \pm 19\%$ . Our values are also in agreement with the BE range calculated, using a 20-water-molecule cluster, by Molpeceres & Kastner (2021) who report 2189 K as the average BE and with the computational BE reported by Viana & da Silva (2015) of 3000 K calculated using a two-water-molecule cluster at the CCSD (T) and MP2 level of theory. Experimental works have shown that phosphine’s thermal sublimation occurs at around 60 K (Turner et al. 2015), consistent with a BE of  $\sim 1800$  K using the formalism of Hollenbach et al. (2009). In summary, our computational method appears to be accurate for phosphine.

To our knowledge, there have not been experimental or computational studies on PO, PN. Next we turn to the P species with calibrated  $2\text{H}_2\text{O}$  cluster BEs above 8000 K, which makes them effectively refractory in astrophysical environments. The calculated BEs for these species present two complications: First of all they fall outside of our calibration range and their BE error bars are therefore more uncertain. Second, in interstellar regions, they are not expected to desorb off water ice. However, we argue that the sublimation temperature on silicate grain should not substantially deviate from those expected on ASW. It has been shown that the BE of volatile species on ASW and on silicate surface falls in the same energy range (Suhasaria et al. 2017). In the case of  $\text{CO}_2$  and CO the experimentally measured BE on silicates differs from the BE on water by less than 20% (Noble et al. 2012a). This is likely due to the reactivity toward hydrogen of silicate surfaces, which results in them hosting -OH functionalities. This results in a binding behavior similar to the one of water (Landmesser et al. 1997; Schaible & Baragiola 2014).

We also found evidence that P-bearing species containing three oxygens can be further coordinated by a water molecule to form the  $\text{PO}_4$  moiety when enough water molecules are present in the environment (Figure 10). Furthermore, the high BE found for OPO and  $\text{HPO}_2$  using the  $3\text{H}_2\text{O}$  cluster size hints to possibility that these molecules might coordinate with the water molecules in the cluster to form  $\text{PO}_3$  species. The study

of these species’ reactivity may especially benefit from the use of bigger cluster size and from the use of dynamic models, this to better define the role that the ASW has in catalyzing the formation of additional PO bonds. We have not further explored this aspect in this work, but if this reactivity is proven to be viable in astrochemical environment, it could provide a plausible pathway for the formation of complex and even more-refractory P-bearing species.

#### 4.3. Astrochemical Implications

The abundance of phosphorous compounds detected in the gas phase at various stages of cloud evolution varies significantly. While only 1% of the expected phosphorous has been detected in star-forming regions (Rivilla et al. 2018; Bergner et al. 2019; Rivilla et al. 2020), phosphorous, in its ionic form, has been detected with solar abundances in diffuse clouds in the ISM (Jura & York 1978; Lebouteiller et al. 2006). This abundance discrepancy indicates that, during star and planet formation, the majority of the phosphorous is depleted on icy grains in semirefractory molecular carriers that have not yet been well constrained.

The majority of the P-bearing molecules that we explored in this work are found to be more refractory than water. This suggests that we should expect a semivolatile to refractory phosphorous reservoir that remains in the solid phase well after water sublimation. In disks, such species would remain solid interior to the water snow line.

The abundance of the PO-bearing species simulated in this work will largely depend on their condensed phase formation chemistry and on the specific environmental conditions. In the solar system, organic phosphonic acids (Cooper et al. 1992) as well as Ca-phosphate (Le Guillou et al. 2014) have been detected on the Murchison meteorite suggesting the possibility for the existence of a rich phosphorous chemistry in condensed phase. We are currently investigating different scenarios computationally to determine the fraction of locked P that is attainable at different ISM conditions (E. L. Piacentino et al., 2022 in preparation).

In comets,  $(\text{H})_x\text{PO}_y$  species have been previously considered as possible phosphorous carriers (Rivilla et al. 2020). Although the investigation on the ROSINA data collected on comet 67P

did not provide the direct detection of  $(\text{H})_x\text{PO}_y$  species, it showed the presence of PO fragments that were attributed solely to PO molecules (Rivilla et al. 2020). As the fragmentation pattern of  $(\text{H})_x\text{PO}_y$  species is not well constrained, it is also possible for the detected PO signal to include a contribution due to the fragmentation of bigger phosphorus molecules. In either case we agree with Rivilla et al.’s (2020) speculation that the PO signal is due to P-bearing molecules that were locked in the grains early on during star formation.

In our calculations we find that  $\text{PO}_2\text{OH}$  and  $\text{PO}_3$  can chemisorb on the water surface to coordinate with an additional oxygen atom. This can indicate the tendency of these PO-bearing species to further react with water molecules to form the likely more-refractory phosphate moiety. We speculate that the easiness with which phosphates may form could, depending on the environmental condition, lead to even a larger fraction of the phosphorous to be locked on grains in phosphate form.

In conclusion, while  $(\text{H})_x\text{PO}_y$  molecules have not yet being directly detected, the clues that we have indicate that these high desorbing species are good candidates for phosphorous carriers during star and planet formation, and may be the starting point of phosphate formation through their interaction with water ice.

## 5. Summary and Conclusions

We explored the performances of a direct ab initio  $\text{H}_2\text{O}$  cluster calculation, and a calibrated version of the same, for BE estimations. We tested our methods using 20 molecules for which the experimental BEs are well constrained in the literature and then applied these methods to 10 P-bearing molecules. We found the following:

1. The direct method– $2\text{H}_2\text{O}$  cluster-method combination performs better than any other method/cluster size that

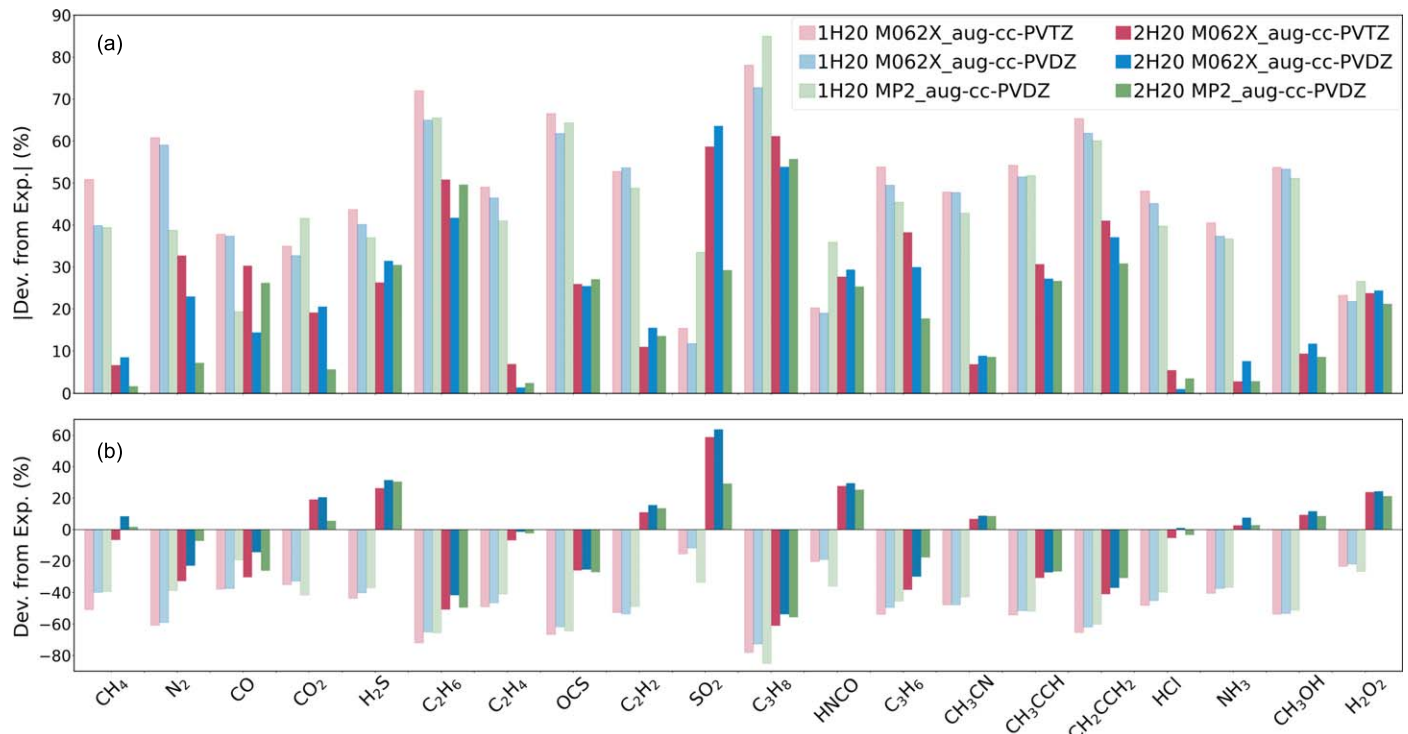
we evaluated. It provides a quick BE estimation that does not seem to carry systematical errors.

2. The application of the calibrated method improves the BE estimation for less-volatile molecules while it reduces the estimation accuracy for highly volatile molecules. This suggests that a targeted selection of the calibration set may be needed.
3. While the application of the calibrated method improves the results, the estimation of the BE for hydrocarbons seems to be difficult. This suggests that the use of a heterogeneous calibration set may not be optimal for all classes of molecules. A functional group-based study may help highlight the effect of molecular proprieties on the BE estimation.
4. The application of these computational methods to astrochemically interesting PO-containing species show that most of these species are more refractory than water.
5. The presence of semirefractory PO-containing species in disks beyond the water snow line could help explain the depletion of P in the ISM and it would supply a pathway for the inclusion of phosphorous in planets and planetesimals.

This work was supported by a grant from the Simons Foundation 686302, KÖ, and by an award from the Simons Foundation 321183FY19, KÖ.

## Appendix A Selection of Model Chemistry

We also tested the model chemistry impact on the methods performances comparing the M06-2X results to similarly obtained BE values using an ab initio, namely MP2 (Frisch et al. 1990), model in combination with the same double zeta basis set (Figure 8). We also compare the M06-2X/aug-cc-



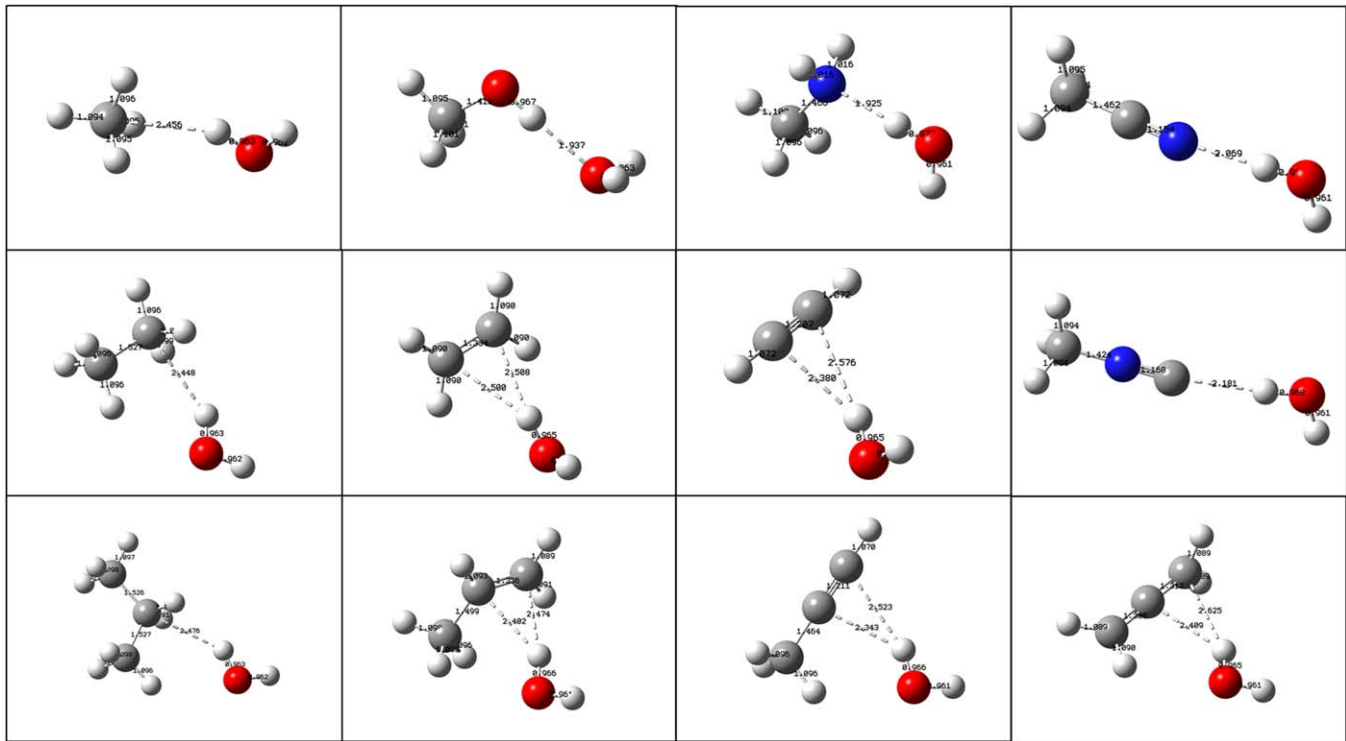
**Figure 8.** Comparison of the deviation from the experimental binding energy predicted by three computational models, M06-2X/aug-cc-pVTZ, M06-2X/aug-cc-pVDZ, and MP2/aug-cc-pVDZ, in combination with the  $1\text{H}_2\text{O}$  and  $2\text{H}_2\text{O}$  representation of the ASW surface. Panel (a) shows the absolute value of the deviation from the experimental values. Panel (b) reports the raw calculated values.

pVDZ and M06-2X/aug-cc-pVTZ (Kendall et al. 1992) performances in evaluating the BE. As only a minimal variation in the BE calculated using the direct method at the M06-2X/aug-cc-pVTZ, M06-2X/aug-cc-pVDZ and MP2/aug-cc-pVDZ is observed (Figure 8), we chose, for clarity, to limit our study to solely the M06-2X/aug-cc-pVDZ model chemistry. We found that the BE estimations depend only slightly on the model chemistry used, but the variation is larger when the cluster size is increased from 1 to 2H<sub>2</sub>O regardless of the model used. In a few cases—N<sub>2</sub>, CO, HNC, and SO<sub>2</sub>—there are real differences for different model chemistries

within the same cluster size, which we speculate are due to a molecular peculiarity that we have not further investigated. In either case, even for these molecules the differences are within the reported uncertainties ( $\leq 20\%$ ), justifying the use of a single model chemistry in the main section of the paper.

## Appendix B

Binding geometries of small hydrocarbons are shown in Figure 9.



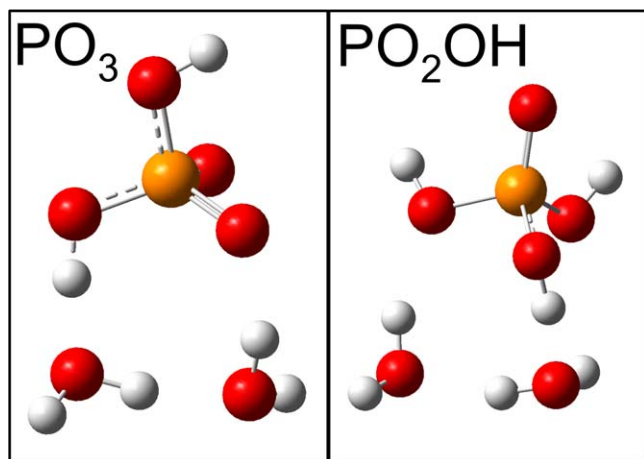
**Figure 9.** Optimized binding geometries (M06-2X/aug-cc-pVDZ) of small hydrocarbons with the 1 H<sub>2</sub>O representation of the ASW surface. Bond lengths are in angstroms.



## Appendix C

### PO<sub>3</sub> Optimization

Figure 10 shows the optimized geometries of PO<sub>3</sub> and PO<sub>2</sub>OH with the three water molecules cluster resulting in the formation of the PO<sub>4</sub> moiety.



**Figure 10.** Optimized binding geometries (M06-2X/aug-cc-pVDZ) of PO<sub>3</sub> and PO<sub>2</sub>OH with the 3H<sub>2</sub>O cluster. These local minima show the formation of additional coordination between the P species and one of the waters of the cluster. This does not exclude the existence of physisorption configurations and it does not necessarily imply reactivity. Additional investigation that are beyond the scope of this work are needed to better understand this behavior.

### ORCID iDs

Elettra L. Piacentino <https://orcid.org/0000-0001-6947-7411>

Karin I. Öberg <https://orcid.org/0000-0001-8798-1347>

### References

Agúndez, M., Cernicharo, J., Decin, L., Encrenaz, P., & Teyssier, D. 2014, *ApJL*, **790**, L27

Agúndez, M., Fonfría, J. P., Cernicharo, J., et al. 2012, *A&A*, **543**, A48

Altwegg, K., Balsiger, H., Bar-Nun, A., et al. 2016, *SciA*, **2**, e1600285

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481

Bahr, S., Toubin, C., & Kempter, V. 2008, *JChPh*, **128**, 134712

Behrard, A., Fayolle, E. C., Graninger, D. M., et al. 2019, *ApJ*, **875**, 73

Bergin, E. A., Blake, G. A., Ciesla, F., Hirschmann, M. M., & Li, J. 2015, *PNAS*, **112**, 8965

Bergner, J. B., Öberg, K. I., Walker, S., et al. 2019, *ApJL*, **884**, L36

Bertin, M., Doronin, M., Michaut, X., et al. 2017, *A&A*, **608**, A50

Boogert, A. C. A., Gerakines, P. A., & Whittet, D. C. B. 2015, *ARA&A*, **53**, 541

Chaabouni, H., Diana, S., & Nguyen, T. 2020, in *Proc. IAU 350, Laboratory Astrophysics: From Observations to Interpretation*, ed. F. Salama & H. Linnartz (Cambridge: Cambridge Univ. Press), 370

Collings, M. P., Anderson, M. A., Chen, R., et al. 2004, *MNRAS*, **354**, 1133

Collings, M. P., Dever, J. W., Fraser, H. J., & McCoustra, M. R. S. 2003, *Ap&SS*, **285**, 633

Cooper, G. W., Onwo, W. M., & Cronin, J. R. 1992, *Geochim. Cosmochim. Acta*, **56**, 4109

Das, A., Sil, M., Gorai, P., Chakrabarti, S. K., & Loison, J. C. 2018, *ApJS*, **237**, 9

Dunning, T. H. J. 1989, *JChPh*, **90**, 1007

Fagerbakke, K. M., Heldal, M., & Norland, S. 1996, *Aquat. Microb. Ecol.*, **10**, 15

Fayolle, E. C., Balfé, J., Loomis, R., et al. 2016, *ApJL*, **816**, L28

Ferrero, S., Zamirri, L., Ceccarelli, C., et al. 2020, *ApJ*, **904**, 11

Fontani, F., Rivilla, V. M., Caselli, P., Vasyunin, A., & Palau, A. 2016, *ApJL*, **822**, L30

Fraser, H. J., Collings, M. P., McCoustra, M. R. S., & Williams, D. A. 2001, *MNRAS*, **327**, 1165

Frisch, M. J., Head-Gordon, M., & Pople, J. A. 1990, *CPL*, **166**, 275

Frisch, M. J., Trucks, G. W., Schlegel, H. B., et al. 2016, *Gaussian 16 Revision C.01* (Wallingford, CT: Gaussian Inc.)

Gálvez, O., Ortega, I. K., Maté, B., et al. 2007, *A&A*, **472**, 691

Gardner, E., Lehto, H. J., Lehto, K., et al. 2020, *MNRAS*, **499**, 1870

Germain, A., Tinacci, L., Pantaleone, S., Ceccarelli, C., & Ugliengo, P. 2022, *ESC*, **6**, 1286

Hama, T., & Watanabe, N. 2013, *ChRv*, **113**, 8783

He, J., Acharyya, K., & Vidali, G. 2016, *ApJ*, **825**, 89

Hollenbach, D., Kaufman, M. J., Bergin, E. A., & Melnick, G. J. 2009, *ApJ*, **690**, 1497

Jones, A. P., Köhler, M., Ysard, N., Bocchio, M., & Verstraete, L. 2017, *A&A*, **602**, A46

Jura, M., & York, D. G. 1978, *ApJ*, **219**, 861

Karssemeijer, L. J., & Cuppen, H. M. 2014, *A&A*, **569**, A107

Karssemeijer, L. J., de Wijs, G. A., & Cuppen, H. M. 2014, *PCCP*, **16**, 15630

Kendall, R. A., Dunning, T. H. J., & Harrison, R. J. 1992, *JChPh*, **96**, 6796

Landmesser, H., Kosslick, H., Storek, W., & Fricke, R. 1997, *Solid State Ionics*, **101**, 271

Le Guillou, C., Bernard, S., Brearley, A. J., & Remusat, L. 2014, *GeoCoA*, **131**, 368

Lebouteiller, V., Kuassivi, & Ferlet, R. 2006, in *ASP Conf. Ser.*, **348**, *Astrophysics in the Far Ultraviolet: Five Years of Discovery with FUSE*, ed. G. Sonneborn, H. W. Moos, & B. G. Andersson (San Francisco, CA: ASP), 480

Lefloch, B., Vastel, C., Viti, S., et al. 2016, *MNRAS*, **462**, 3937

Lodders, K. 2003, *ApJ*, **591**, 1220

Marboeuf, U., Thiabaud, A., Alibert, Y., Cabral, N., & Benz, W. 2014, *A&A*, **570**, A36

Mardirossian, N., & Head-Gordon, M. 2017, *MolPh*, **115**, 2315

Milam, S. N., Halfen, D. T., Tenenbaum, E. D., et al. 2008, *ApJ*, **684**, 618

Molpeceres, G., & Kästner, J. 2021, *ApJ*, **910**, 55

Nguyen, T., Oba, Y., Sameera, W. M. C., Kouchi, A., & Watanabe, N. 2021, *ApJ*, **918**, 73

Noble, J. A., Congiu, E., Dulieu, F., & Fraser, H. J. 2012a, *MNRAS*, **421**, 768

Noble, J. A., Theule, P., Mispelaer, F., et al. 2012b, *A&A*, **543**, A5

Öberg, K. I., & Bergin, E. A. 2021, *PhR*, **893**, 1

Öberg, K. I., Fayolle, E. C., Cuppen, H. M., van Dishoeck, E. F., & Linnartz, H. 2009, *A&A*, **505**, 183

Öberg, K. I., Murray-Clay, R., & Bergin, E. A. 2011, *ApJL*, **743**, L16

Olanrewaju, B. O., Herring, J. C., Grieves, G. A., Aleksandrov, A., & Orlando, T. M. 2011, *JPCA*, **115**, 5936

Pasek, M. A., & Lauretta, D. S. 2005, *ASBio*, **5**, 515

Pasek, M. A., Smith, V. D., & Lauretta, D. S. 2004, in *35th Lunar and Planetary Science Conference*, ed. S. Mackwell & E. Stansbery (Houton, TX: LPI), 1703

Penteado, E. M., Walsh, C., & Cuppen, H. M. 2017, *ApJ*, **844**, 71

Rivilla, V. M., Drozdovskaya, M. N., Altwegg, K., et al. 2020, *MNRAS*, **492**, 1180

Rivilla, V. M., Fontani, F., Beltrán, M., et al. 2018, *IAUS*, **332**, 409

Rubin, M., Altwegg, K., Balsiger, H., et al. 2019, *MNRAS*, **489**, 594

Schäuble, M. J., & Baragiola, R. A. 2014, *JGRE*, **119**, 2017

Shimonishi, T., Nakatani, N., Furuya, K., & Hama, T. 2018, *ApJ*, **855**, 27

Smith, R. S., May, R. A., & Kay, B. D. 2016, *JPCB*, **120**, 1979

Suhasaria, T., Thrower, J. D., & Zacharias, H. 2017, *MNRAS*, **472**, 389

Tenenbaum, E. D., Woolf, N. J., & Ziurys, L. M. 2007, *ApJL*, **666**, L29

Turner, A. M., Abplanalp, M. J., Chen, S. Y., et al. 2015, *PCCP*, **17**, 27281

Viana, R. B., & da Silva, A. B. 2015, *Comput. Theor. Chem.*, **1059**, 35

Wakelam, V., Loison, J.-C., Mereau, R., & Ruaud, M. 2017, *MolAs*, **6**, 22

Williams, J. P., & Cieza, L. A. 2011, *ARA&A*, **49**, 67

Yamaguchi, T., Takano, S., Sakai, N., et al. 2011, *PASJ*, **63**, L37

Zamirri, L., Ugliengo, P., Ceccarelli, C., & Rimola, A. 2019, *ESC*, **3**, 1499

Zhao, Y., & Truhlar, D. G. 2008, *Theo. Chem. Acc.*, **120**, 215

Ziurys, L. M., Milam, S. N., Apponi, A. J., & Woolf, N. J. 2007, *Natur*, **447**, 1094