Draft version March 4, 2022
Typeset using LaTeX **twocolumn** style in AASTeX631

Binding Energy Evaluation Platform: A database of quantum chemical binding energy distributions for the astrochemical community.

Giulia Bovolenta [1], Stefan Vogt-Geisse [1], Stefano Bovino [2] and Tommaso Grassi [3]

[1] *Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, Concepción, Chile*
[2] *Departamento de Astronomía, Facultad Ciencias Físicas y Matemáticas, Universidad de Concepción, Av. Esteban Iturra s/n Barrio Universitario, Casilla 160, Concepción, Chile*
[3] *Centre for Astrochemical Studies,Max-Planck-Institut für extraterrestrische Physik,Giessenbachstrasse 1, 85749 Garching bei München, Germany*

ABSTRACT

The quality of astrochemical models is highly dependent on reliable binding energy values that consider the morphological and energetic variety of binding sites on the surface of ice-grain mantles. Here we present the Binding Energy Evaluation Platform (BEEP) and database that, using quantum chemical methods, produces full binding energy (BE) distributions of molecules bound to an amorphous solid water (ASW) surface model. BEEP is highly automatized and allows to sample binding sites on set of water clusters and to compute accurate BEs. Using our protocol, we computed 21 BE distributions of interstellar molecules and radicals on an amorphized set of 15-18 water clusters of 22 molecules each. The distributions contain between 225 and 250 unique binding sites. We apply a Gaussian fit and report the mean and standard deviation for each distribution. In most cases in which the molecule is bound to the surface through a hydrogen bond, two or more distributions are present and are fitted individually. We compare with existing experimental results and find that the low and high coverage experimental BEs coincide well with the high BE tail and mean value of our distributions, respectively. Previously reported single BE theoretical values are broadly in line with ours, even though in some cases significant differences can be appreciated. We show how the latter impact a typical problem in astrophysics, as the computation of snow lines in protoplanetary discs. BEEP will be publicly released together with the database to allow for expansions to more molecules or ice-models and improvements in a community effort.

1. INTRODUCTION

In dense interstellar clouds, where the temperature is less than 20 K, interstellar dust particles are covered with a layer of ice consisting mostly of H_2O and at a lower proportion with molecules such as CO_2, NH_3 and CH_4 (see e.g. Boogert et al. 2015). In these cold environments, interstellar chemistry can take place on the ice mantles of interstellar dust grains (e.g. Herbst & van Dishoeck 2009). The ice surface is capable of binding different molecules from the gas phase, thus facilitating chemical encounters and promoting the formation of new molecular species, that can be detected once they desorb into the gas phase (e.g. Jorgensen et al. 2020). In that regard, the binding energy (BE) is a crucial parameter when modeling gas-grain chemistry in dense clouds as it determines the desorption rate of the adsorbed species for thermal, chemical and photo desorption. Having knowledge of the BE of molecules on ice mantles allows astrochemical gas-grain models to predict the abundances of molecular and atomic species. The composition, structure and formation of the ice mantles is still a matter of research. However, the broad shape of the water 3.1 μm O-H stretching band observed in different dense cloud regions, suggests, upon comparison with experimental results, that the water component of the ice mantles exists in amorphous form, as layers of amorphous solid water (ASW) (Smith et al. 1989). This is important inasmuch the BE depends both on the nature of the adsorbed species and of the composition and morphology of the ice mantle.

BEs on ice surfaces can be determined experimentally, mainly using Temperature Programmed Desorption (TPD). In TPD experiments, a layer of ASW is build through vapor deposition and exposed to the

Corresponding author: Stefan Vogt-Geisse
stvogtgeisse@qcmmlab.com

species of interest in a constant temperature regime. Once the desired level of coverage is reached, the temperature is increased and the desorbed molecules are collected and analyzed by mass spectrometry. To date, several TPD experiments have been performed using ASW ice as substrate, ranging from multilayer to sub-monolayer regime of adsorbed molecules. One of the first extensive TPD studies, done by Collings et al. (2004), made desorption rate measurements of 16 astrophysically-relevant molecules on an ASW substrate in a monolayer and multilayer regime. Binding energies at sub-monolayer deposition have also been determined using TPD measurements, by inversion of the Polanyi-Wigner equation, which yields a coverage dependent adsorbate BE. The coverage is usually measured as a fraction of a monolayer (ML) and ranges from 1 ML to 10^{-3} ML. Coverage-dependent BE distributions have been obtained for a few astrophysically important molecules such as N_2 (Smith et al. 2016; He et al. 2016), O_2 (Smith et al. 2016; He et al. 2016), CO (Noble et al. 2012; Smith et al. 2016; He et al. 2016), CO_2 (Noble et al. 2012; He et al. 2016), CH_4 (Smith et al. 2016; He et al. 2016) and D_2 (Amiaud et al. 2006; He et al. 2016).

Even though TPD experiments provide valuable BE data, the preparation of the substrate and deposition technique can vary among experiments, which makes it difficult to construct a homogeneous database of experimental BE values. Also, TPD is not suitable to provide BE values for radicals due to the short life-span of these species.

On the other hand, BEs can also be determined using a computational approach by means of *ab initio* quantum chemistry methods and molecular dynamics (MD) simulations. In recent years, there has been important progress in the development of both the construction of ASW models and in the computations of BE. Two types of ASW models have been proposed: using a slab of ASW with periodic boundary conditions, or using amorphized water clusters. Several computational works have been carried out using the former model, for the most relevant species in the interstellar medium. In the most complete study thus far, Ferrero et al. (2020) computed BEs of 21 molecules and atoms. They generated an amorphized water cluster consisting of 60 water molecules and computed the BEs for up to 8 binding sites per molecule, while imposing periodic boundary conditions. The second approach consists of using a cluster model to simulate parts of the ASW surface. Within the cluster approach, two strategies for computing BE have been proposed. First, using a large surface of hundreds of water molecules in a QM/MM embedded regime, in which the bulk is described with a force field and the molecules close to the binding site are computed by means of quantum chemistry methods. Using this approach Song & Kästner (2016, 2017) computed binding energy distributions of HNCO and H_2CO. More recently Duflot et al. (2021) obtained binding energies of 8 different binding sites of several species (H, C, N, O, NH, OH, H_2O, CH_3, NH_3) using a ONIOM QM/QM hybrid method. A similar procedure was used by Sameera et al. (2021) to compute 10 binding sites of the CH_3O radical.

An alternative approach to the cluster model was first introduced by Shimonishi et al. (2018). They used a set of previously annealed 20- molecule water clusters to represent different regions of an ASW surface. This set of water clusters was sampled with different atomic species to compute BEs at Density Functional Theory (DFT) level of theory, and only the highest BE values on each water cluster were reported.

Finally, the efforts to obtain an extensive binding energy catalogue for small molecules on water surfaces have been limited to DFT computations on small water clusters (up to 6 molecules, Sil et al. 2017; Das et al. 2018) or interaction with water monomer by linear semi-empirical models (Wakelam et al. 2017), which do not capture the complete statistical nature of the interaction on ASW.

In this work, we present BEEP, a Binding Energy Evaluation Platform meant to offer a straightforward and easy-to-use interface for the computation and processing of full BE distributions of molecules. To showcase the utility of BEEP, we computed BE distributions of 22 astrophysically-relevant molecules. The platform is implemented within the QCArchive framework(Smith et al. 2020a) which allows to transform the database in a fully open-source endeavour, from the data generation to the final API for querying the BE data.

2. COMPUTATIONAL DETAILS

2.1. *Surface Modeling*

To build an ASW surface serving as an ice mantle model, we adapted the *cluster approach* first introduced by Shimonishi et al. (2018). We generated a surface model consisting of 22 water molecules (W_{22}) and performed a high temperature *ab initio* molecular dynamics simulation followed by temperature annealing in order to amorphyze the system and reach interstellar conditions. Finally, we selected the 20 most representative water structures which form our amorphized set of ASW clusters (see Appendix, G for an example of structure). The use of a cluster of this size allows a good compromise between accuracy and computational time, and has

been validated in our previous work, to which we refer for further details (Bovolenta et al. 2020).

2.2. *Geometry optimization and binding energy calculation*

We performed a DFT geometry benchmark on the W_{1-3}–X systems, with X being the target molecule and W the water cluster, using DF-CCSD(T)-F12/cc-pVDZ-F12 (Bozkaya & Sherrill 2017; Werner et al. 2020; Dunning T.H. et al. 2001) geometry as a reference (see Appendix D, Table 3) .
We also conducted an energy benchmark, using W_4–X system to compare DFT BE values to a CCSD(T)/CBS (Klopper & Kutzelnigg 1986; Feller 1992; Helgaker et al. 1997; Karton & Martin 2006) reference energy (see Appendix, D, Table 4).
We used BLYP/def2-SVP(Becke 1988; Lee et al. 1988; Miehlich et al. 1989; Weigend & Ahlrichs 2005a) as level of theory for the binding site sampling procedure by means of the Terachem software (Ufimtsev & Martinez 2009; Titov et al. 2013), to take advantage of the efficient GPU acceleration. All high level DFT optimizations were preformed together with a def2-TZVP basis set.
We also computed the Hessian matrix for selected structures at the equilibrium geometry to obtain the Zero-Point Vibrational Energy contribution (Δ_{ZPVE}) to the BE, computed at the same level of theory as the geometry optimization.
The binding energy has been calculated as

$$\Delta E_b = \Delta E_{CP} + \Delta_{ZPVE}\,, \tag{1}$$

with ΔE_{CP} being the binding energy corrected for the basis set superposition error. See the Appendix B,C for more details. The binding energy is conventionally assumed to be a positive quantity.
For the single point computations at DFT level of theory, we employed a def2-TZVP basis set. All high level optimization and energy computations were performed using Psi4 (Parrish et al. 2017).

2.3. *QCArchive Framework*

Traditionally, quantum chemistry data has been generated through user-defined individual input files, which are processed by a specific software that stores the results of the computation in output files. These outputs are then parsed either by hand or using custom scripts. This approach has serious limitations when attempting to compute a large volume of quantum chemistry data as it is error-prone and difficult to reproduce, since parsing scripts and output files are usually not available. To overcome these limitations, we build the BEEP platform within the Python-based QCArchive framework. The details about the different components of the QCArchive infrastructure have been described elsewhere (Smith et al. (2021)). The core component of BEEP is a central SQL server to which computation results are added in the form of JSON objects (QCSchema) that contain the same level of information as a standard output file. The access to this database where the user can query existing data and submit additional computations, is controlled by a standard username/password system. Moreover, several data objects can be defined to generate and sort the data. These collections (called *Datasets*) make it possible to extend a procedure, such as a geometry optimization or a BE computation, to a large number of objects in a single operation. Finally, the generated values can be easily accessed from the stored collections.

3. RESULTS

In this section we will first present each step of the computational pipeline (3.1) and then we will discuss the database results (3.2).

3.1. *Computational Pipeline*

The pipeline we developed allows to produce ZPVE corrected BE distributions for closed-shell and open-shell molecules. As shown in Figure 1, it is composed of three main steps: (1) sampling procedure, (2) geometry optimization and (3) BE calculation. In order to go through the pipeline, we recall the reader the QCArchive data structures we introduced in 2.3.

3.1.1. *Sampling procedure*

In order to perform the sampling procedure (Figure 1, blue panel labelled "sampling") within of the QCFractal environment, both the ASW clusters and the target molecules have to be stored in collection objects (ASW Dataset and Species Dataset). The initial molecular geometries contained in the Species Dataset are drawn from the Pubchem library, which can be accessed directly from the QCFractal environment. The sampling procedure is carried out at BLYP/def2-SVP level of theory, and consists of extracting one ASW structure at a time from the ASW Dataset and sample it with the target molecule X. The sampling algorithm places the center of mass of both species on the origin of the system coordinates, and displaces the species X around the surface randomly within a range of distances which maximizes the chance of finding a binding site on the surface. Starting with ice cluster ASW_a, several groups of 10 ASW_a–X binding site candidates (BSC) are generated. These are optimized (*opt1*) and filtered according

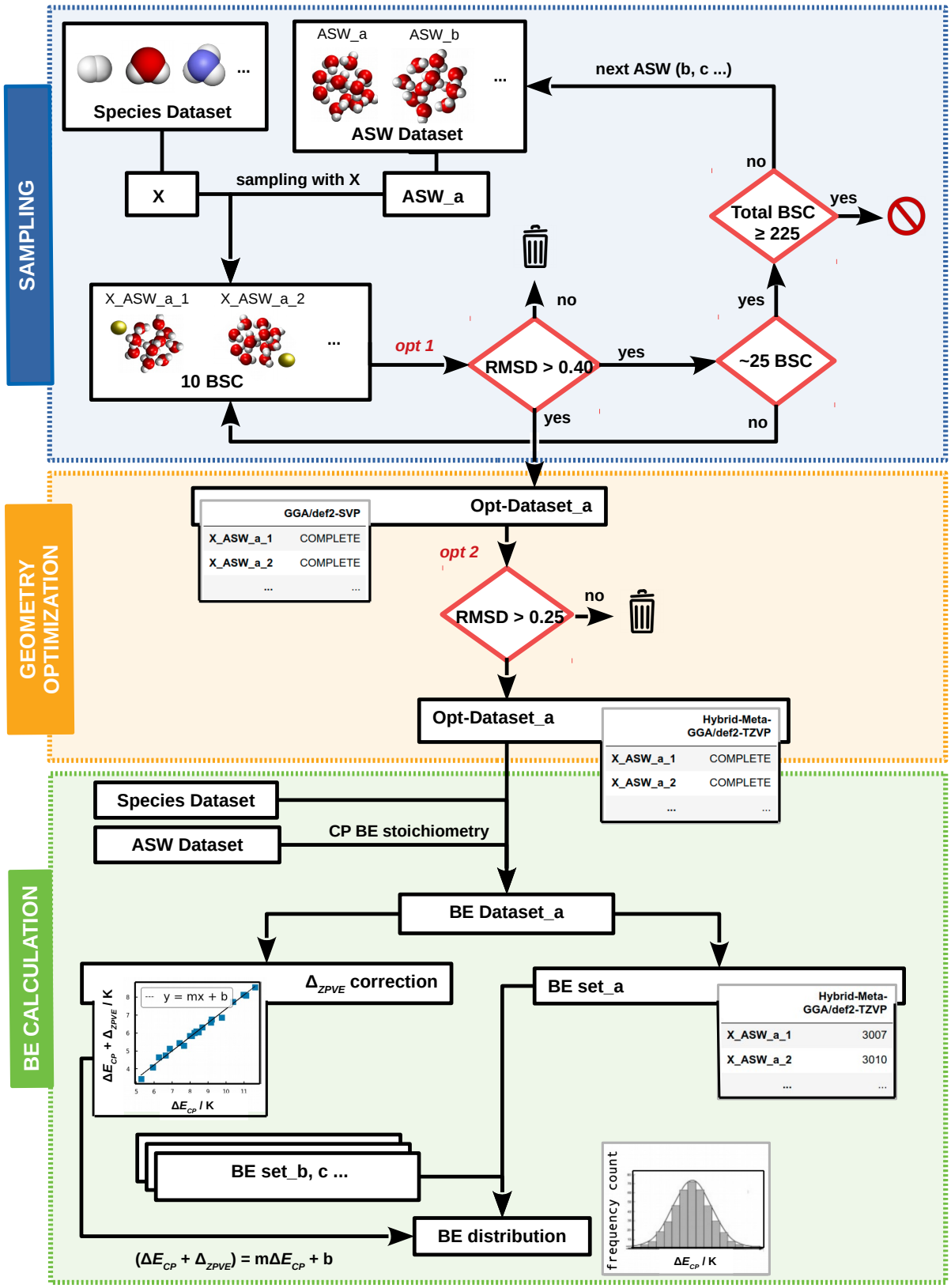

Figure 1: Three-step computational procedure used in this work for building a binding energy distribution. BSC stands for binding site candidate; *opt 1* stands for optimization at gradient generalized approximation (GGA) exchange-correlation DFT functional and *opt 2* for optimization at a higher level of theory that further refines the geometry. The color scheme for the atoms is red for O, white for H, blue for N and yellow for the generic target atom X.

to geometrical criteria such that only the structures of RMSD $\geq$ 0.40 Å with respect to previously found BSC are stored, until 25 BSC is reached or no more new BSC are found. This procedure is repeated on a second cluster ASW_b until a total of at least 225 ASW−X equilibrium structures. The number of ASW clusters that need to be sampled to reach this number of BSC is around 12-15.

3.1.2. *Geometry Optimization*

In this step (Figure 1, yellow panel labelled "geometry optimization"), the BSC candidates obtained in the previous step, are further optimized at a more accurate level of theory, such as a hybrid or meta-hybrid functional with a triple zeta basis set. As shown in the Appendix (D), using a computationally affordable HF-3c/MINIX (Sure & Grimme 2013) model chemistry can also be a good option for obtaining a refined equilibrium geometry of the different binding sites.

3.1.3. *Binding energy calculation*

The final part of the pipeline (Figure 1, green panel labelled "BE calculation") is the computation of BE values and the assembly of a ZPVE corrected BE distribution. To do so, first, the optimized structures are filtered with geometry criteria (RMSD $\geq$ 0.25 Å) to make sure that all binding sites on the ASW cluster are unique. The resulting equilibrium structures are included into a BE *Dataset* collection, together with the optimized target molecule and water cluster to create the stoichiometry of a BE including the counterpoise correction for the BSSE error (see eq. B3). Once a BE *Dataset* for ASW_a−X is generated, it contains all the fragments necessary to compute the BSSE corrected BE values on ASW_a (BE set_a). Analogously, we collect a set of BEs for each of the sampled clusters. Assuming the clusters share common morphological characteristics, as they originate from a single *ab initio* molecular dynamics trajectory and are annealed in the same way, the BEs collected are considered as a single BE distribution of the target molecule on the ice mantle model. We then correct the values by adding Δ_{ZPVE} to the BE. Due to computational cost, we compute the Hessian for the elements of a single BE Dataset (e.g., ASW_a−X corresponding to one sampled water cluster in our set), and use a linear model to correlate ΔE_{CP} and $\Delta E_{CP} + \Delta_{ZPVE}$ (see Appendix, C). Finally, the correction factors are applied to all the computed BEs to obtain a ZPVE corrected BE distribution. The source code of the BEEP protocol and scripts to generate the data can be found in www.github.com/QCMM/beep.

3.2. *Binding energy distributions*

We divided the molecular species in two groups according to the nature of the interaction with the ice surface. Group D accounts for interactions dominated by dispersion, while molecules in Group H predominantly bind through hydrogen bonds. We computed 21 binding energy distributions for closed-shell and open-shell molecules, reported in Figure 2 for Group H and Figure 3 for Group D. The equilibrium geometry is of HF-3c/MINIX quality, as we probed it to be a cost-effective alternative to the more expensive DFT methods (see Appendix, D). For CO species, the geometry is M05/def2-TZVP (Zhao et al. 2005), as HF-3c failed to properly describe the binding sites . We computed the ZPVE correction at the HF-3c/MINIX level of theory for Group H, while for most of the molecules in Group D we could not apply the linear model we used to derive the correction factors, due to poor correlation. This could be attributed to the inadequacy of the harmonic approximation to correctly describe the potential energy well. Notwithstanding, the correction value for Group D molecule is small enough to fall within the accuracy of the method. The BE values were computed using the best performing DFT functional from the energy benchmark for each molecule (see Appendix, D). If no benchmark value is present, we used the best performing functional for each group.

Finally, while multibinding energies approaches have been recently proposed (Grassi et al. 2020), we decided to also provide a single binding energy value, representative of the entire distribution, to accommodate the usage of our calculations in standard chemical models. For this purpose, we obtained the mean binding energy (μ) and standard deviation (σ) by fitting a Gaussian function to the distribution using a bootstrap method (Appendix E). We carried out binding mode analyses in order to identify different binding motifs which are labelled in the figure, along with their percentage and their μ and σ values.

3.2.1. *Group H: Hydrogen bonded structures*

Figure 2 shows BE distributions of molecules in Group H. These molecules are bound mostly through an electrostatic interaction in the form of hydrogen bonds and therefore exhibit a strong interaction with the ASW surface. This is reflected in the BE values which are in the range of 1000 to 8000 K. It is worth noting that several species exhibit two or more distinct distributions. For NH_3 and $NHCH_2$, there is a main binding mode represented by the formation of a bond with the surface via the N atom (BE $\sim$ 3400 K), keeping a dangling NH bond; while the minor binding mode shows the forma-

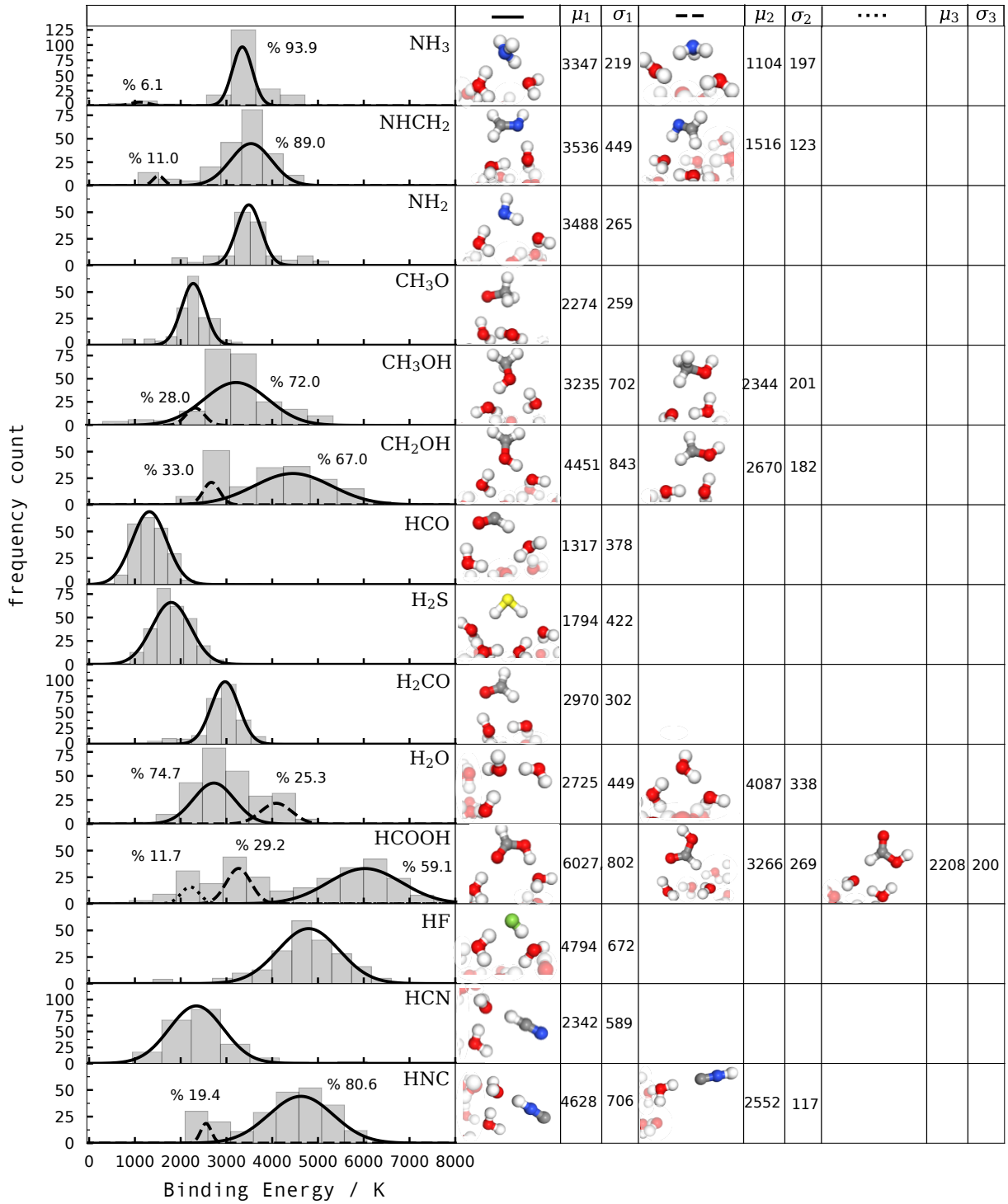

Figure 2: Binding energy distributions for Group H, ASW−X systems, using HF-3c/MINIX geometries and including ZPVE correction. According to the benchmark results, the energy has been computed at ω−PBE/def2-TZVP level of theory for all species except HNC (B97-2/def2-TZVP), H_2CO (CAM-B3LYP/def2-TZVP), CH_3OH (TPSSH/def2-TZVP), HF and HCN (MPWB1K/def2-TZVP). D3BJ dispersion correction has been applied to all DFT energies. Each identified binding mode has been fitted with a Gaussian function, using a bootstrap method (see Appendix, E). Mean (μ) and standard deviation (σ) of the Gaussian fit are reported for the different modes. The numbers on the plot represent the percentage of minimum energy structures that belong to a specific mode. Column 2, 5 and 7 reports a graphic representation of an example minimum of each adsorption type. The atoms in proximity of the binding site have been highlighted. The color scheme for the atoms is red for O, grey for C, white for H, blue for N, yellow for S and green for F.

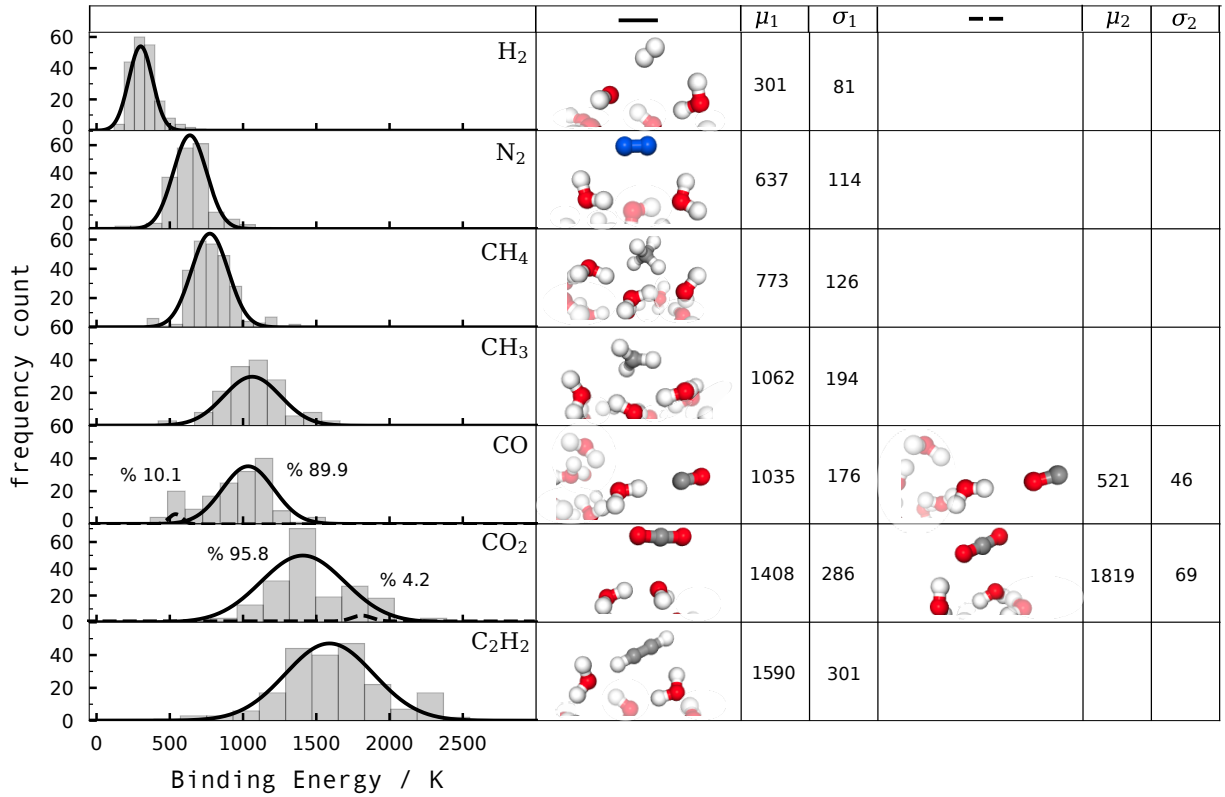

Figure 3: Binding energy distributions for Group D, ASW−X systems, using HF-3c/MINIX geometries. ZPVE correction has been included only for C_2H_2, see text. The energy has been computed at ω-PBE/def2-TZVP level of theory for all species except CH_4 (TPSSH/def2-TZVP). See Figure 2 for further details.

tion of a bond via the H atoms of the molecule (BE $\sim$ 1400 K). On the other hand, for CH_3OH and its radical species (CH_2OH), the main binding mode is the interaction via the sole OH-bond (BE $>$ 4000 K), while in the minor binding mode the methyl end of the molecule participates in the interaction with the surface, and the OH-bond remains dangling (BE $\sim$ 1400 K). In the distribution of the CH_3O radical, we found a single binding mode, corresponding to the less energetically favorable interaction where both the oxygen and the methyl take part. This is consistent with the inability of this radical to form a donor type hydrogen bond. Due to its lack of symmetry, the formic acid presents a rather complex BE distribution with three different components, spanning a range of almost 7000 K. The minor modes present dangling OH-bonds as in the case of the methanol species. Regarding the water molecule, a closer inspection of the binding modes shows a varied scenario where the molecule establishes a single hydrogen bond to the surface (BE $\sim$ 3000 K) or two (BE $\sim$ 4000 K).

Even though the latter occurs more often during the sampling procedure, most water molecules that compose the ASW surface have two hydrogen bonds, and thus the BE of those molecules would fall within the higher BE distribution. The halogen (HF) and pseudo-halogen (HNC, HCN) molecules have a high standard deviation ($\geq$ 590 K) which reflects a high capacity of insertion into the ASW environment. This is especially seen in the HF case, in which the molecule is easily inserted into the hydrogen bond network, forming strong hydrogen bonds with the water surface, as we have shown in a previous work (Bovolenta et al. 2020). While the HNC species present adsorption through both extremities, with a definite preference for CNH–OH bond creation (80.6 %), adsorption through the H atom is largely predominant in HCN species.

We also studied the HCl molecule, but it does not have a BE distribution as it dissociates to its ionic components in the majority of the binding sites, as also pointed out in the recent work of Ferrero et al. (2020). Finally, it is worth noting that the ZPVE correction can significantly reduce the BEs in some cases up to 25% of the non-corrected value.

3.2.2. *Group D: structures bound by dispersion*

Figure 3 shows the BE distributions of Group D. In order to identify the molecules that belong to this group, we compared the BE distributions obtained with and

without including D3BJ dispersion correction to the energy computation. For molecules in Group D, the dispersion interaction is fundamental in order to achieve an attractive interaction with the water surface (see Appendix, F). They are mainly homonuclear or highly symmetric molecules. The mean BE values range between 300 and 1800 K and are significantly lower than in the Group H molecules. Furthermore, the standard deviation is also less than in Group H molecules, which is consistent with a smaller capacity of the molecule to deform the binding site environment. Most molecules therefore present a single binding motif. An outlier is CO since its BE distribution reveals two distinct binding modes: a weak interaction where the CO molecule is bound to the surface via an electrostatically unfavorable CO–H interaction (BE $\sim$ 500 K) and a second, more dominant, binding mode which embarks 89.9 % of the structures. In the latter, the CO molecule is bound through its C- extremity (BE $\sim$ 1000 K). The other molecule that presents more than one binding mode is CO_2. In the highest BE motif the CO_2 interacts with the surface through both the C and one of the O atoms of the molecule.

4. COMPARISON WITH EXPERIMENTAL RESULTS AND PREVIOUS THEORETICAL STUDIES

We compared our BE values with available experimental results, previous theoretical studies and existing astrochemical databases.

Making a meaningful comparison of calculated BEs with experimental data is challenging, due to the variety of conditions under which the experiments are performed. In addition, the experimental data strongly depend on the pre-exponential factor used in the Polanyi-Wigner equation employed to derive the BEs (see Minissale et al. 2022). We decided to take into account the work of He et al. (2016) where they presented TPD measurement of BEs of relatively simple molecules (N_2 , H_2, CO, CH_4, and CO_2) on a non-porous ASW (np-ASW) surface at monolayer (ML) and submonolayer coverage. In He et al. experiments, it is possible to distinguish between two situations in terms of the coverage (θ) of the target molecule on the surface. The low coverage limit ($\theta \rightarrow 0$), represents a situation in which mostly the binding sites of high BE would be occupied, corresponding to the high energy tail of BE distribution. On the other hand, BE values obtained at the monolayer regime ($\theta \simeq 1$ ML) can be related with the mean of our BE distribution, where a variety of adsorption sites with different energies are occupied. The comparison between our results and their low coverage and monolayer regime BEs is shown in Figure 4. Overall, the experimental results of these limiting coverage cases coincide well with the computational values obtained in this work. The comparison is particularly good for H_2, N_2 and CO, (a difference of $<$ 155 K in the low coverage regime and $<$ 170 K in the ML regime) while the error for CH_4 is larger (a difference in low coverage regime of 207 K, and a difference in ML regime of 337 K). In light of these results, we conclude that our approach of sampling a number of independent ASW clusters of a limited size (22 water molecules) allows to reproduce the statistical nature of the interaction of those molecules with an actual ice surface.

Regarding the comparison with previously reported theoretical values, we took into account the works of Das et al. (2018) and Ferrero et al. (2020) (Figure 5, upper panel). Das et al. computed values for W_4–X systems at MP2/aug-cc-pVDZ level of theory without correction for counterpoise and nor for ZPVE. Also, the existence of multiple binding sites is not considered. Regarding Group H, in most cases Das' values fall within the range of energies we found for the same systems. For molecules in Group D, Das' values mostly overestimate ours. This is consistent with the lack of CP correction that has an important effect on the final BE values for this group (CP correction $\sim$ 100-250 K in our BE results).

Recently, Ferrero et al. proposed a new set of BE values, computed at DFT/A-VTZ* level, including ZPVE and CP correction. They simulated a periodic amorphous slab model that presents a concave region on the upper surface, and identified different binding sites per molecule. The aim of their work was different than ours inasmuch as they tried to obtain a range of possible BE values and not a full distribution. Their single ASW model slab contains a cavity that allowed them to explore up to 8 different binding sites. Their lower BE fall within our distribution for most of the systems, but their BEs are in average higher than the ones presented here. An important contribution to this discrepancy is the different approach to the ZPVE correction, which can account to up to a 25% of the total BE value. Within our pipeline, we compute a ZPVE correction for each individual molecule, which reduces the BE by factors ranging between 0.72 to 0.92 (see Appendix, C). Meanwhile, in Ferrero et al. they used a single correction factor of 0.854, which was computed based on ZPVE values on a crystalline water surface. Another possible reason for higher BE is the shape of the water cluster as it contains a nano-cavity and, as recently pointed out (see Rimola et al. 2018; Enrique-Romero et al. 2019; Bovolenta et al. 2020), the presence of cavities notably increases the BEs, offering more favourable interaction

Table 1: Comparison with data from the literature. The first column reports the molecules, column 2 and 3 our results: the mean of the predominant binding modes identified (μ_1, μ_2, μ_3) and the highest BE value of each distribution (Max). Column 4 to 5 reports experimental results, columns 6 to 8 BEs computed in theoretical studies, columns 9 and 10 the values present in the astrochemical databases KIDA and UMIST. Units are in K and the references are listed in the notes below.

	BEEP (ASW)		**He** (np-ASW)[a]		**Das**[c]	**Ferrero** (ASW)[d]		**KIDA**[e]	**UMIST**[f]
	$\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$	**Max**	$\boldsymbol{\theta \simeq 1ML}$	$\boldsymbol{\theta \rightarrow 0}$		**Min**	**Max**		
H_2	310	660	322	505	528	226	431	440	430
N_2	637	1189	790	1320	900	760	1458	1100	790
CH_4	773	1393	1100	1600	1327	914	1674	960	1090
CH_3	1062	1662			1322	1109	1654	1600	1175
CO	1035	1561	870	1600	1263	1109	1869	1300	1150
CO_2	1408, 1819	2389	2320[h]		2293	1489	2948	2600	2990
C_2H_2	1590	2547			2593			2587	2587
NH_3	3347, 1104	4715			3825	4314	7549	5500	5534
$NHCH_2$	3536, 1516	4695			3354			5534[m]	3428
NH_2	3488	5235			3240	2876	4459	3200	3956
CH_3O	2274	3343						4400	5080
CH_3OH	3235, 2344	5331			4368	3770	8618	5000	4930
CH_2OH	4451, 2670	6594			4772			4400	5084
HCO	1317	3764			1857	1315	3081	2400	1600
H_2S	1794	2940			2556	2291	3338	2700	2743
H_2CO	2970	3800			3242	3071	6194	4500	2050
H_2O	2725, 4087	4885			2670	3605	6111	5600	4800
HCOOH	6027, 3266, 2208	8044			3483	5382	10559	5570[n]	5000
HF	4794	6500			5540			7500	
HCl	[g]	[g]			3924	[g]	[g]	5172	900
HCN	2342	4252			2352	2496	6337	3700	2050
HNC	4628, 2552	6570			5225			3800	2050

[a] He et al. (2016); [c] Das et al. (2018); [d] Ferrero et al. (2020); [e] Wakelam et al. (2017); [f] McElroy et al. (2013); [g] HCl molecules dissociate; [h] coverage insensitive; [m] Ruaud et al. (2015); [n] Collings et al. (2004).

sites for the molecule on the surface. It is still uncertain to which extent the real ASW ice-mantle surface contains such defects, and therefore how statistically relevant they are for our aim of obtaining a full distribution of BE. A recent chemical kinetics simulations of ice-mantle formation has shed some insights into these questions as they conclude that the surface is relatively uniform after a sufficient amount of ice mantle as been build up (Christianson & Garrod 2021). Therefore, the presence of surface defects such as nano-cavities will most likely affect the high energy tail of the distribution, while the mean value will be mostly determined by the interaction of the molecule with a more uniform ASW surface.

Regarding the CH_3O radical, we took into account for comparison the recent work of Sameera et al. (2021). They used 10 molecular-dynamic generated ASW structural models composed of 162 water molecules, which have been sampled with the target CH_3O. The resulting 10 BEs have been computed using the two-layer ONIOM(QM:MM) approach, at wB97XD/def2-TZVP (Chai & Head-Gordon 2008) level of theory including ZPVE correction; we reported their minimum and maximum values in Figure 5, upper panel. They identified a wide range of energy (1160 - 4874 K), that contains the values of our distribution, nevertheless their average BE is greater than ours by about 1300 K.

Finally, in Fig. 5, lower panel, we show the comparison of our data with the largely used KIDA and UMIST databases values. They mostly fall in the range of our BE distributions, except for some specific cases, where the agreement is poor (CO_2 and NH_3 among them). These KIDA values are mostly based on the BE calculated in Wakelam et al. (2017) using a semi-empirical model consisting of a linear fit between the BEs on water monomers and experimental values on ASW surfaces. The BEs calculated using this model tend to overestimate our average values for both Group H and

D. It is important to consider that the comparison is highly related to the experimental values available that have been used to build the linear model.

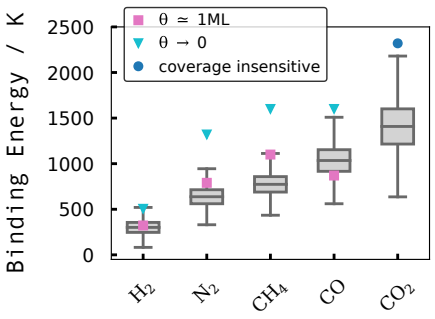

Figure 4: Comparison between He et al. (2016) experimental results and the BE distributions presented in this work.

5. ASTROPHYSICAL IMPLICATIONS

While to compare the results between chemical experiments and/or calculations, a difference of 0.2-0.3 kcal mol^{-1} (corresponding approximately to 100-150 K) in the final BE is not substantial, for astrochemistry modelling a few tens of Kelvins could largely affect the final outcome. The molecular desorption is described by the Polanyi-Wigner equation, where its dependence on the exponential of the binding energy plays a crucial role in determining the efficiency of the process. To show this effect on a realistic, yet idealized, astrophysical case, we have calculated the sublimation radius in a protoplanetary disc (i.e., the so-called snow line) by equating the desorption and the viscous time, and finding the corresponding radius (see e.g. Grassi et al. 2020). The evaporation time is defined as $t_{\rm des}(R) \propto \exp[{\rm BE}/k_{\rm B}T_{\rm d}(R)]$, with $k_{\rm B}$ and $T_{\rm d}$ respectively the Boltzmann constant and the dust temperature at a given radius R. The viscous time is $t_\nu(R) \propto R^2/\nu(R)$, where $\nu(R) = \alpha\, c_{\rm s}^2(R)\Omega_{\rm K}^{-1}(R)$ is the viscosity, assuming an α-viscous prescription with $\alpha = 10^{-2}$, and $c_{\rm s}$ the speed of sound and $\Omega_{\rm K}$ the Keplerian angular frequency. By means of the bisection method, we solve $t_{\rm des}(R) = t_\nu(R)$ for R, that corresponds to $\varphi \log R - {\rm BE}\sqrt{R} = 0$, with φ containing all the constant terms (see Appendix H for more details). The results are reported in Fig. 6. As expected, the position of the snow lines is affected by the assumed BE up to approximately an order of magnitude in the worst cases. For water, one of the most important molecule involved in the process of planet formation, we obtain $R = 4$ au for the BE computed by Ferrero et al. (the highest, driven by the presence of a nanocavity), and 20 au for our mean value. A similar effect is reported for CO, with up to a factor of three in the final radius. Other species, like molecular hydrogen, show larger differences; However we do not expect them to form observable snow lines, since they are involved in other chemical processes that are not captured by our simplified disk model, but we report them anyway for the sake of completeness. An accurate determination of the BE is then fundamental to quantitatively assess quantities like snow line positions in planet-forming regions and evaporation fronts during star-formation.

6. DATABASE FEATURES, ACCESSIBILITY AND USE-CASES

Due to the nature of QCArchive Databases, BEEP is extendable to an increasingly large number of molecules. Moreover, different cluster surface models of different sizes and composition (e.g., different ice mixtures) can be easily added to the platform environment and used to produce new BE distribution data.

At the moment, the BEEP platform can be accessed with a username and password, which are provided in the Appendix A. This allows the user to query the database for BE, binding site structures and many other properties. To make the access to the database a user-friendly experience, we included a Python module that allows to query the data without having to know the QCArchive syntax. The core of this Python module is the *BindingParadise* class that is initialized with the user's credentials and allows to set molecules and obtain all the related BE data. In the GitHub repository (www.github.com/qcmm/beep) we included an example jupyter-notebook to showcase the different query options. The libraries to compute and store a BE distribution are also contained in the Python module. In principle, any researcher can install the module to run the software and spin up a QCFractal server to store its own BE data. However, our idea is to make this a collaborative endeavour in which different researchers use the proposed protocol to generate new BE data and store it in our open BEEP database. This allows us to expand the database in terms of new ice models and a more extensive BE catalog with more computed molecules. The database will be able to produce input files in the standard astrochemical software format, both in a single BE fashion and in more complex multibinding approaches. A database of reproducible and accurate BEs is also fundamental starting point to chemical reactivity studies and diffusion of molecules on the surface of interstellar ices as having a potential energy map of neighbouring

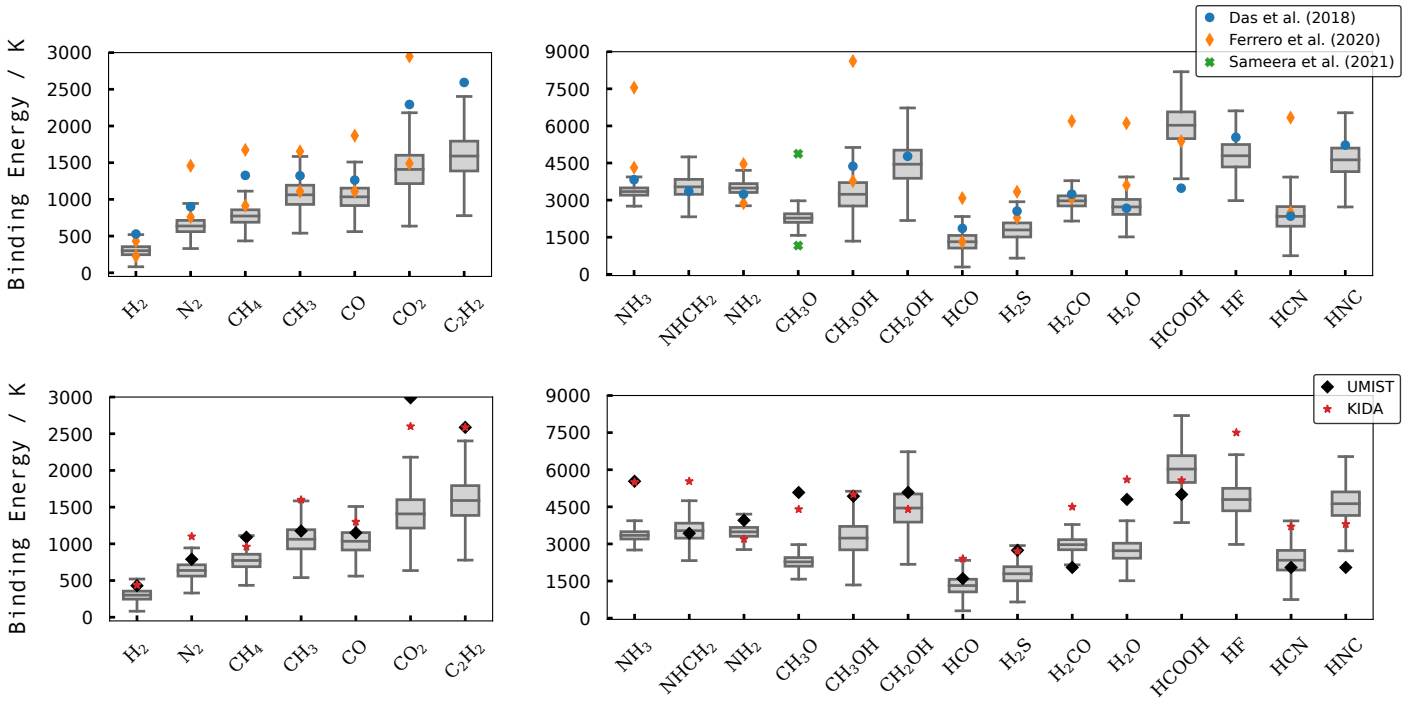

Figure 5: Comparison between BE distributions computed in this work, considering only the main binding mode, and BEs present in previous theoretical studies (Das et al. (2018); Ferrero et al. (2020); Sameera et al. (2021)) upper panels, and existing databases BE data: KIDA (Wakelam et al. 2017), UMIST (McElroy et al. 2013), lower panels.

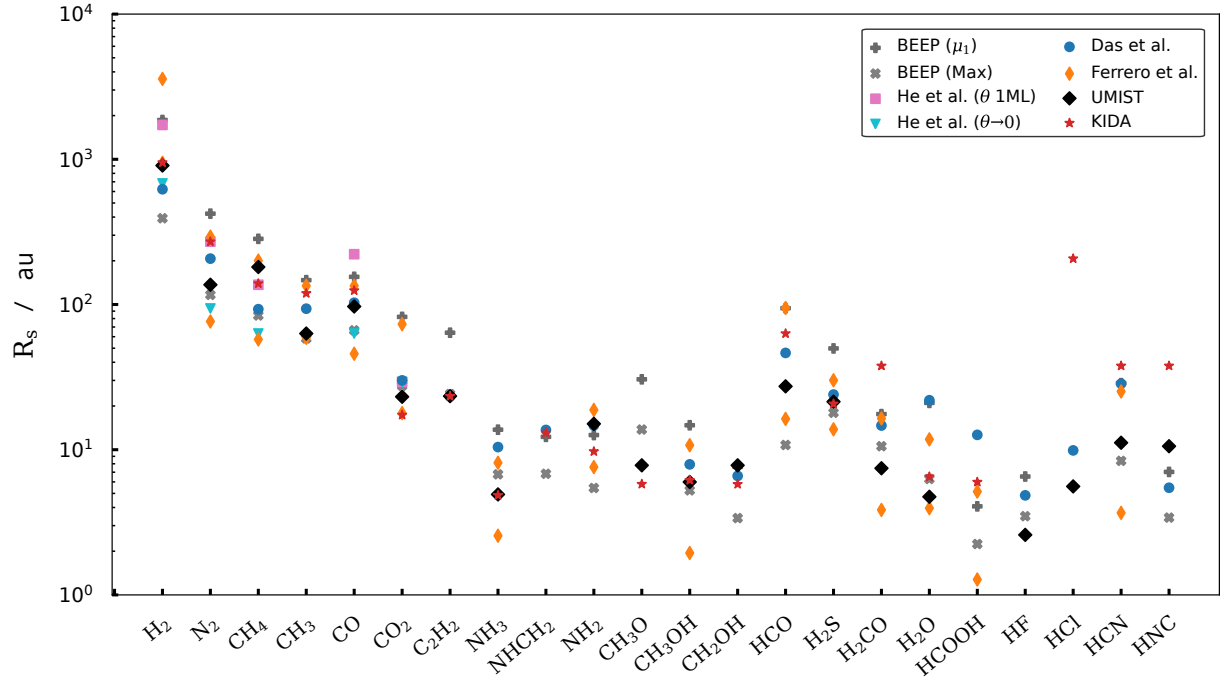

Figure 6: Sublimation radius for different species obtained by employing the binding energies obtained in this work and compared with values available in literature.

binding sites will be paramount in finding diffusive transition states and computing diffusion energy barriers.

7. CONCLUSION

In this work, we present a Binding Energy Evaluation Platform (BEEP) that implements a protocol to compute binding energies on ASW cluster models. It also contains a database that allows to query the results produced by the protocol. The pipeline consists of three highly automated steps that are a target molecule sampling procedure, geometry optimization of the binding site and binding energy computation. The binding energy distributions were obtained by sampling the ASW model spanned by a set of 12-17 amorphized water clusters containing 22 molecules each. We categorized the molecules into two groups, one group with molecules that bind to the surface primarily through hydrogen-bonds (group H) and a second group for which dispersion interactions enable binding to the surface (group D). A DFT equilibrium geometry and energy benchmark on small water clusters using a high-level coupled cluster wavefunction reference, allowed us to establish that the HF-3c/MINIX level of theory is an acceptable model chemistry for equilibrium geometries of binding sites and that ω-PBE/def2-TZVP with counterpoise correction yields excellent binding energies at a relatively low computational cost. However, a hybrid DFT functional such a B3LYP (group D) or meta-hybrid such as PWB6K (group H) should be employed for a more refined binding site geometry, albeit at a much higher computational cost. Using the level of theory with the highest fidelity with respect to the benchmark reference of the respective group or molecule, we computed 21 binding energy distributions on the ASW model surface. Each BE distribution contains between 220-230 structures. Most molecules in group D contain only one BE distribution while for molecules in group H, more than one distribution, corresponding to different binding modes of the target molecule to the surface, were identified and fitted. BEEP is build on an open-source platform and hence any researcher can use it to compute binding energies with a cluster based ASW ice surface model. Finally, we plan to transform BEEP into a widely-used tool for standardized *ab initio* BE energy data for astrochemical modeling and ice-grain surface processes studies.

ACKNOWLEDGMENTS

The computations were performed with resources provided by the Kultrun Astronomy Hybrid Cluster hosted at the Astronomy Department, Universidad de Concepción. We would like to thank Benjamin Pritchard for his guidance on the QCFractal platform. GB gratefully acknowledges support from ANID Beca de Doctorado Nacional 21200180 and Proyecto UCO 1866 - Beneficios Movilidad 2021. SB gratefully acknowledges support by the ANID BASAL projects ACE210002 and FB210003.

REFERENCES

Amiaud, L., Fillion, J. H., Baouche, S., et al. 2006, The Journal of Chemical Physics, 124, 094702, doi: 10.1063/1.2168446

Becke, A. D. 1988, Phys. Rev. A, 38, 3098, doi: 10.1103/PhysRevA.38.3098

—. 1993, The Journal of Chemical Physics, 98, 5648, doi: 10.1063/1.464913

Boogert, A., Gerakines, P., & Whittet, D. 2015, Annual Review of Astronomy and Astrophysics, 53, 541, doi: 10/ggrs2p

Bovolenta, G., Bovino, S., Vöhringer-Martinez, E., et al. 2020, Molecular Astrophysics, 100095, doi: 10.1016/j.molap.2020.100095

Boys, S. F., & Bernardi, F. 1970, Molecular Physics, 19, 553, doi: 10.1080/00268977000101561

Bozkaya, U., & Sherrill, C. D. 2017, J. Chem. Phys., 147, 044104, doi: 10.1063/1.4994918

Chai, J.-D., & Head-Gordon, M. 2008, Phys. Chem. Chem. Phys., 10, 6615, doi: 10.1039/B810189B

Christianson, D. A., & Garrod, R. T. 2021, Frontiers in Astronomy and Space Sciences, 8

Collings, M. P., Anderson, M. A., Chen, R., et al. 2004, Monthly Notices of the Royal Astronomical Society, 354, 1133, doi: 10.1111/j.1365-2966.2004.08272.x

Das, A., Sil, M., Gorai, P., Chakrabarti, S. K., & Loison, J.-C. 2018, The Astrophysical Journal Supplement Series, 237, 9, doi: 10.3847/1538-4365/aac886

Duflot, D., Toubin, C., & Monnerville, M. 2021, Frontiers in Astronomy and Space Sciences, 8, doi: 10.3389/fspas.2021.645243

Dunning T.H., J., Peterson, K. A., & Wilson, A. K. 2001, Journal of Chemical Physics, 114, 9244, doi: 10.1063/1.1367373

Enrique-Romero, J., Rimola, A., Ceccarelli, C., et al. 2019, ACS Earth and Space Chemistry, 3, 2158, doi: 10/ggtpwm

Feller, D. 1992, J. Chem. Phys., 96, 6104, doi: 10.1063/1.462652

Ferrero, S., Zamirri, L., Ceccarelli, C., et al. 2020, The Astrophysical Journal, 904, 11, doi: 10.3847/1538-4357/abb953

Grassi, T., Bovino, S., Caselli, P., et al. 2020, Astronomy & Astrophysics, 643, A155, doi: 10.1051/0004-6361/202039087

Grimme, S., Brandenburg, J. G., Bannwarth, C., & Hansen, A. 2015, J Chem Phys, 143, 054107, doi: 10.1063/1.4927476

He, J., Acharyya, K., & Vidali, G. 2016, The Astrophysical Journal, 825, 89, doi: 10.3847/0004-637X/825/2/89

Helgaker, T., Klopper, W., Koch, H., & Noga, J. 1997, J. Chem. Phys., 106, 9639, doi: 10.1063/1.473863

Herbst, E., & van Dishoeck, E. F. 2009, Annual Review of Astronomy and Astrophysics, 47, 427, doi: 10.1146/annurev-astro-082708-101654

Jorgensen, J. K., Belloche, A., & Garrod, R. T. 2020, arXiv:2006.07071 [astro-ph]. https://arxiv.org/abs/2006.07071

Karton, A., & Martin, J. M. L. 2006, Theor Chem Acc, 115, 330, doi: 10.1007/s00214-005-0028-6

Klopper, W., & Kutzelnigg, W. 1986, Journal of Molecular Structure: THEOCHEM, 135, 339, doi: 10.1016/0166-1280(86)80068-9

Lee, C., Yang, W., & Parr, R. G. 1988, Phys. Rev. B, 37, 785, doi: 10.1103/PhysRevB.37.785

McElroy, D., Walsh, C., Markwick, A. J., et al. 2013, Astronomy & Astrophysics, 550, A36, doi: 10.1051/0004-6361/201220465

Miehlich, B., Savin, A., Stoll, H., & Preuss, H. 1989, Chemical Physics Letters, 157, 200, doi: 10/d8qp3r

Minissale, M., Aikawa, Y., Bergin, E., et al. 2022, arXiv e-prints, arXiv:2201.07512. https://arxiv.org/abs/2201.07512

Noble, J. A., Congiu, E., Dulieu, F., & Fraser, H. J. 2012, Monthly Notices of the Royal Astronomical Society, 421, 768, doi: 10.1111/j.1365-2966.2011.20351.x

Parrish, R. M., Burns, L. A., Smith, D. G. A., et al. 2017, J. Chem. Theory Comput., 13, 3185, doi: 10/gcz64j

Rimola, A., Skouteris, D., Balucani, N., et al. 2018, ACS Earth and Space Chemistry, 2, 720, doi: 10/gdzs9h

Ruaud, M., Loison, J. C., Hickson, K. M., et al. 2015, Monthly Notices of the Royal Astronomical Society, 447, 4004, doi: 10.1093/mnras/stu2709

Sameera, W. M. C., Senevirathne, B., Andersson, S., et al. 2021, The Journal of Physical Chemistry A, 125, 387, doi: 10.1021/acs.jpca.0c09111

Shimonishi, T., Nakatani, N., Furuya, K., & Hama, T. 2018, The Astrophysical Journal, 855, 27, doi: 10.3847/1538-4357/aaaa6a

Sil, M., Gorai, P., Das, A., Sahu, D., & Chakrabarti, S. K. 2017, The European Physical Journal D, 71, 45, doi: 10/gg3w7q

Smith, D. G. A., Altarawy, D., Burns, L. A., et al. 2020a, WIREs Computational Molecular Science, n/a, e1491, doi: 10.1002/wcms.1491

Smith, D. G. A., Burns, L. A., Simmonett, A. C., et al. 2020b, J. Chem. Phys., 152, 184108, doi: 10/gg8w5c

Smith, D. G. A., Altarawy, D., Burns, L. A., et al. 2021, WIREs Comput Mol Sci, 11, doi: 10.1002/wcms.1491

Smith, R. G., Sellgren, K., & Tokunaga, A. T. 1989, The Astrophysical Journal, 344, 413, doi: 10.1086/167809
Smith, R. S., May, R. A., & Kay, B. D. 2016, The Journal of Physical Chemistry B, 120, 1979, doi: 10.1021/acs.jpcb.5b10033
Song, L., & Kästner, J. 2016, Physical Chemistry Chemical Physics, 18, 29278, doi: 10/ggsr6k
—. 2017, The Astrophysical Journal, 850, 118, doi: 10.3847/1538-4357/aa943e
Sure, R., & Grimme, S. 2013, Journal of Computational Chemistry, 34, 1672, doi: 10.1002/jcc.23317
Titov, A. V., Ufimtsev, I. S., Luehr, N., & Martinez, T. J. 2013, J. Chem. Theory Comput., 9, 213, doi: 10/f4m53f
Ufimtsev, I. S., & Martinez, T. J. 2009, J. Chem. Theory Comput., 5, 2619, doi: 10/bf3qt3
Vydrov, O. A., Heyd, J., Krukau, A. V., & Scuseria, G. E. 2006, J. Chem. Phys., 125, 074106, doi: 10.1063/1.2244560
Vydrov, O. A., & Scuseria, G. E. 2006, J. Chem. Phys., 125, 234109, doi: 10.1063/1.2409292
Vydrov, O. A., Scuseria, G. E., & Perdew, J. P. 2007, J. Chem. Phys., 126, 154109, doi: 10.1063/1.2723119
Wakelam, V., Loison, J. C., Mereau, R., & Ruaud, M. 2017, Molecular Astrophysics, 6, 22, doi: 10.1016/j.molap.2017.01.002
Wang, L.-P., & Song, C. 2016, J. Chem. Phys., 144, 214108, doi: 10/gf3hv9
Weigend, F., & Ahlrichs, R. 2005a, Phys. Chem. Chem. Phys., 7, 3297, doi: 10.1039/B508541A
—. 2005b, Phys. Chem. Chem. Phys., 7, doi: 10.1039/b508541a
Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R., & Schütz, M. 2012, WIREs Computational Molecular Science, 2, 242, doi: 10/dkf9dz
Werner, H.-J., Knowles, P. J., Manby, F. R., et al. 2020, J. Chem. Phys., 152, 144107, doi: 10.1063/5.0005081
Zhao, Y., Schultz, N. E., & Truhlar, D. G. 2005, J. Chem. Phys., 123, 161103, doi: 10.1063/1.2126975
Zhao, Y., & Truhlar, D. G. 2005, J Phys Chem A, 109, 5656, doi: 10.1021/jp050536c

APPENDIX

A. BEEP DATABASE ACCESS

The binding energy and binding site data generated using BEEP can be accessed using the python *BindingParadise* class. Refer to the GitHub repositories for installation instructions. To initialize a class object and access the data you can use the following credentials:

username : guest password: pOg_41tzuDxkTtAfjPuUq8WK5ssbnmN8QfjsApGXVYk

Examples of how to use the class in the form of jupyter-notebooks can be found in our GitHub repository at www.github.com/QCMM/beep.

B. BINDING ENERGY CALCULATION STOICHOIMETRY

In the following, we define the electronic energy of a molecule M in the geometry G computed with the basis γ as $E_M^G(\gamma)$. Considering this notation, the binding energy of a molecule X with a basis set χ on a water cluster W with a basis set ω can be calculated as:

$$\Delta E_e = E_{XW}^{XW}(\chi \cup \omega) - (E_X^X(\chi) + E_W^W(\omega)) \tag{B1}$$

However, when using this expression, one does not consider that the basis function centered at W assists in lowering the energy of fragment X and viceversa, resulting in a lower $E_{XW}^{XW}(\chi \cup \omega)$ and hence an overestimation of the BE. This effect is commonly known as basis set superposition error (BSSE). A way to correct for this error is the so-called counterpoise method (CP) (Boys & Bernardi 1970), that considers the energy of the fragments in the geometry of the supermolecule with the basis of the respective partner. Thus the correction is calculated as:

$$\begin{aligned}\Delta_{CP} = E_X^{XW}(\chi \cup \omega) - E_X^{XW}(\chi) \\ + E_W^{XW}(\chi \cup \omega) - E_W^{XW}(\omega)\end{aligned} \tag{B2}$$

Such that the resulting BE is:

$$\Delta E_{CP} = \Delta E_e - \Delta_{CP} \tag{B3}$$

It is important to notice that at the CBS limit, the correction term is zero since, χ, ω and $\chi \cup \omega$ are the same.

C. GEOMETRY OPTIMIZATION AND Δ_{ZPVE} CORRECTION

The optimization algorithm for all equilibrium structures presented in this work is geomeTRIC (Wang & Song 2016), which uses a coordinate system especially suitable for optimizations of non-covalently bound systems. Due to computational cost, we computed the Hessian matrix for the binding sites of a single ASW cluster, at the level of theory of the optimization (HF-3c/MINIX), in order to obtain the Zero-Point Vibrational Energy contribution (Δ_{ZPVE}) to the BE:

$$\Delta_{ZPVE} = ZPVE_{XW} - (ZPVE_X + ZPVE_W) \tag{C4}$$

with X being the target molecule, W the water cluster and XW the supermolecule.
The linear model we used to correct the BEs (ΔE_{CP}) is an equation in the form:

$$\Delta E_{CP} + \Delta_{ZPVE} = m\Delta E_{CP} + b, \tag{C5}$$

with m and b being the ZPVE correction factors. A list of correction factors for each species is reported below. Finally, the factors are applied to the set of computed BEs for each species in order to derive the ZPVE corrected BE distribution.
The code we used in order to process the computed Hessian data makes use of Psi4 functions and can be found at www.github.com/QCMM/beep.

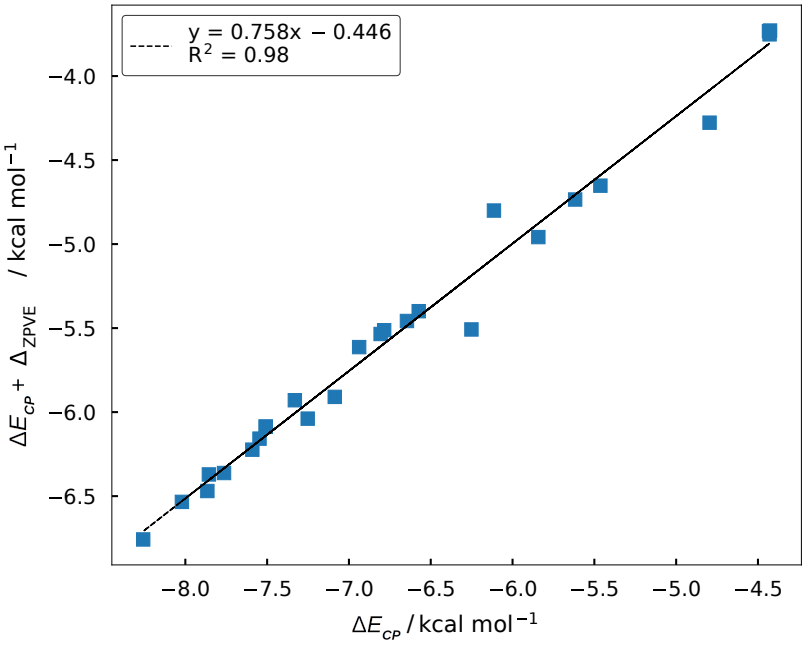

Figure 7: Linear model applied to H_2CO molecule.

Table 2: Column 1: species; column 2: average BE values calculated in this work with ZPVE correction. Columns 3-5: Δ_{ZPVE} computed at HF-3c/MINIX and correction factors (m and b) obtained using a linear model. All the energies are in kelvin.

	μ_1, μ_2, μ_3	Δ_{ZPVE}	m	b
C_2H_2	1590	-389	0.803	0.000
NH_3	3347, 1104	-951	0.762	0.142
$NHCH_2$	3536, 1516	-491	0.844	0.277
CH_3O	2274	-277	0.814	0.394
CH_3OH	3235, 2344	-613	0.819	0.170
HCO	1317	-297	0.723	0.299
H_2S	1794	-432	0.806	0.000
H_2CO	2970	-650	0.758	-0.446
H_2O	2725, 4087	-465	0.781	0.466
HCOOH	6027, 3266, 2208	-960	0.899	-0.508
HF	4794	-1211	0.798	0.000
HCN	2342	-494	0.826	0.000
HNC	4628, 2552	-355	0.929	0.000

D. GEOMETRY AND ENERGY BENCHMARKS

In order to obtain the best possible equilibrium geometry at a reasonable computational cost, we performed a geometry benchmark on the W_{1-3}–X systems, with X being the target molecule and W the water cluster. The benchmark has been conducted for 13 selected molecules. A DF-CCSD(T)-F12/cc-pVDZ-F12 geometry was used as a reference, and we probed 24 gradient generalized approximation (GGA) exchange-correlation density functionals including functionals with exact exchange (hybrid functionals), the Laplacian of the electron density (meta functionals) and long-range correction, paired with a def2-TZVP (Weigend & Ahlrichs 2005b) basis. We also probed the parametrized HF-3c/MINIX (Sure & Grimme 2013) and PBEh-3c/def2-mSVP(Grimme et al. 2015) levels of theory. We also conducted an energy benchmark, using the W_4–X system to compare basis set superposition error corrected DFT BE values to a CCSD(T)/CBS reference energy. The Molpro (Werner et al. 2012) program was used for reference geometries and Psi4 (Smith et al. 2020b) software package was used for all energy computations.

Table 3 reports geometry benchmark results. Generally, the meta-hybrid-GGA methods have a very good performance across the groups. The most dependable functionals are B3LYP(Becke 1993; Lee et al. 1988) for Group D and PWB6K(Zhao & Truhlar 2005) for Group H, as both show an average RMSD value that is below 0.1 Å with respect to the reference geometry. Additionally, we probed the HF-3c method coupled with MINIX basis set to gauge the accuracy of this very cost-effective level of theory. The results are reported in the last column of Table 3 and show an average RMSD that is below 0.2 Å for both groups, which is in line with the RMSD values of hybrid and meta-hybrid functionals probed in this benchmark. This makes it a cost-effective alternative to the computationally more expensive DFT methods.

Furthermore, we evaluated the dependence of the BE distributions on the underlying binding site geometries, comparing the BE distribution of the equilibrium structures obtained with HF-3c/MINIX and the best performing DFT method. Figure 8 reports the comparison for Group D, left panel and Group H, right panel. We fitted a Gaussian function to the distributions using a bootstrap method (see Appendix, E). For all species, the mean BE (μ) presents a shift passing from DFT to parametrized methods, while the standard deviation (σ) is mostly unchanged. The shift in the position of μ is below 400 K for all the species except CO ($\Delta\mu$ of 604 K), for which HF-3c largely underestimates the BE, predicting a repulsive tail in the distribution. Interestingly, HF-3c distributions are slightly shifted to lower values for Group D and to higher values for Group H. In light of these results, we conclude that the HF-3c/MINIX model chemistry can be used in lieu of a more expensive DFT method, as it shows only small difference in the position and width of the Gaussian fit of the underlying BE distributions.

Regarding the energy benchmark, Table 4, for both groups the best DFT functional is the ω-PBE (Vydrov & Scuseria 2006; Vydrov et al. 2006, 2007) with BSSE and D3BJ dispersion corrections, coupled with def2-TZVP basis set. The average mean absolute error (MAE) is 37 and 160 K for Group D and H respectively. Full benchmark results can be found at www.github.com/QCMM/beep.

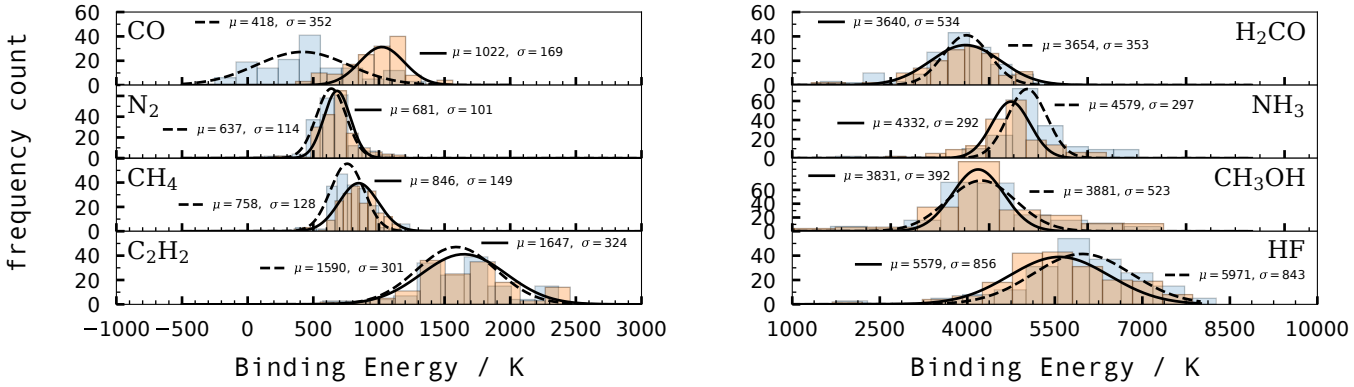

Figure 8: Comparison between binding energy distributions obtained using Meta-hybrid GGA geometries (yellow histogram, Gaussian function represented with solid line) and HF-3c geometries (blue, dashed line). Left panel: Group D; right panel: Group H. BE values are shown without ZPVE correction.

Table 3: Summary of the results of the geometry benchmark for W_{2-3}–X (W_2–X for radicals) systems. The first column reports the molecules. Columns 2-3 report the performance of the best DFT functional for each group, and of HF-3c. Only structures that converged (n) to the reference minima (N) were considered for the benchmark. All DFT geometries were computed using a def2-TZVP basis set, while HF-3c is coupled with MINIX basis set.

	RMSD / Å	
Group D	**B3LYP(n/N)**	**HF-3c(n/N)**
H_2	0.11 (3/6)	0.14 (5/6)
CO	0.12 (7/7)	0.22 (5/7)
CH_4	0.07 (2/2)	0.14 (2/2)
CH_3	0.09 (1/2)	0.12 (1/2)
N_2	0.13 (3/3)	0.26 (3/3)
Average	0.10 (16/20)	0.18 (16/20)
Group H	**PWB6K(n/N)**	**HF-3c(n/N)**
NH_3	0.06 (4/7)	0.14 (5/7)
CH_3OH	0.08 (8/8)	0.13 (6/8)
HCOOH	0.06 (11/13)	0.17 (10/13)
H_2CO	0.06 (5/6)	0.15 (4/6)
HF	0.04 (3/4)	0.06 (2/4)
HCl	0.07 (6/6)	0.18 (2/6)
HCO	0.05 (3/3)	0.07 (1/3)
HNC	0.08 (4/5)	0.30 (3/5)
HCN	0.06 (4/5)	0.20 (3/5)
Average	0.06 (48/57)	0.15 (36/57)

Table 4: Summary of the results of the energy benchmark for W_4-X (W_3-X for radicals) systems. The first column report the molecules. The second column reports reference energies calculated at CCSD(T)/CBS level of theory. The third column reports the Mean Absolute Error (MAE) of the best DFT functional for each group. All DFT energies were computed using a def2-TZVP basis set and including D3BJ dispersion correction.

	BEs / K	**MAE / K**
Group D	**CCSD(T)/CBS**	**ω-PBE**
H_2	320, 116	19
CO	950, 870, 791	10
CH_4	712	74
CH_3	821, 824	45
Average		37
Group H	**CCSD(T)/CBS**	**ω-PBE**
NH_3	3632, 3516, 3562	79
CH_3OH	3922, 4111, 4005	119
H_2CO	2600, 1197, 1181	338
HNC	4211,3953	305
HF	5956, 5380, 4158	83
HCl	3445, 2923, 956	146
HCO	2224, 1684	56
Average		160

E. GAUSSIAN FITTING PROCEDURE

To fit the binding energy distribution data with a Gaussian function, we employed a bootstrap method. We first divide our sample in 30 equally-spaced bins, so that each bin contains N_i samples, with a Poisson error $\sqrt{N_i}$. We then produce 10^4 distributions analogue to the original data, but where the points are randomized assuming a Gaussian error $\sqrt{N_i}$ around the mean N_i and we fit each distribution with

$$f(x) = a \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{E6}$$

where a, μ, and σ are free parameters. The binned distribution of each parameter after the 10^4 iterations is also a Gaussian, where the average is the value we assume for the given parameter and the dispersion is the associated error.

F. DISPERSION CORRECTION

The following figure reports the comparison between the histograms of the BE distributions computed in this work, with and without including the dispersion correction (D3BJ).
The overlapping is almost complete for most of the molecule in Group H, lower panel, while there is a significant difference for Group D, upper panel. This reflects the importance of the dispersion contribution in ensuring an attractive interaction with the surface.

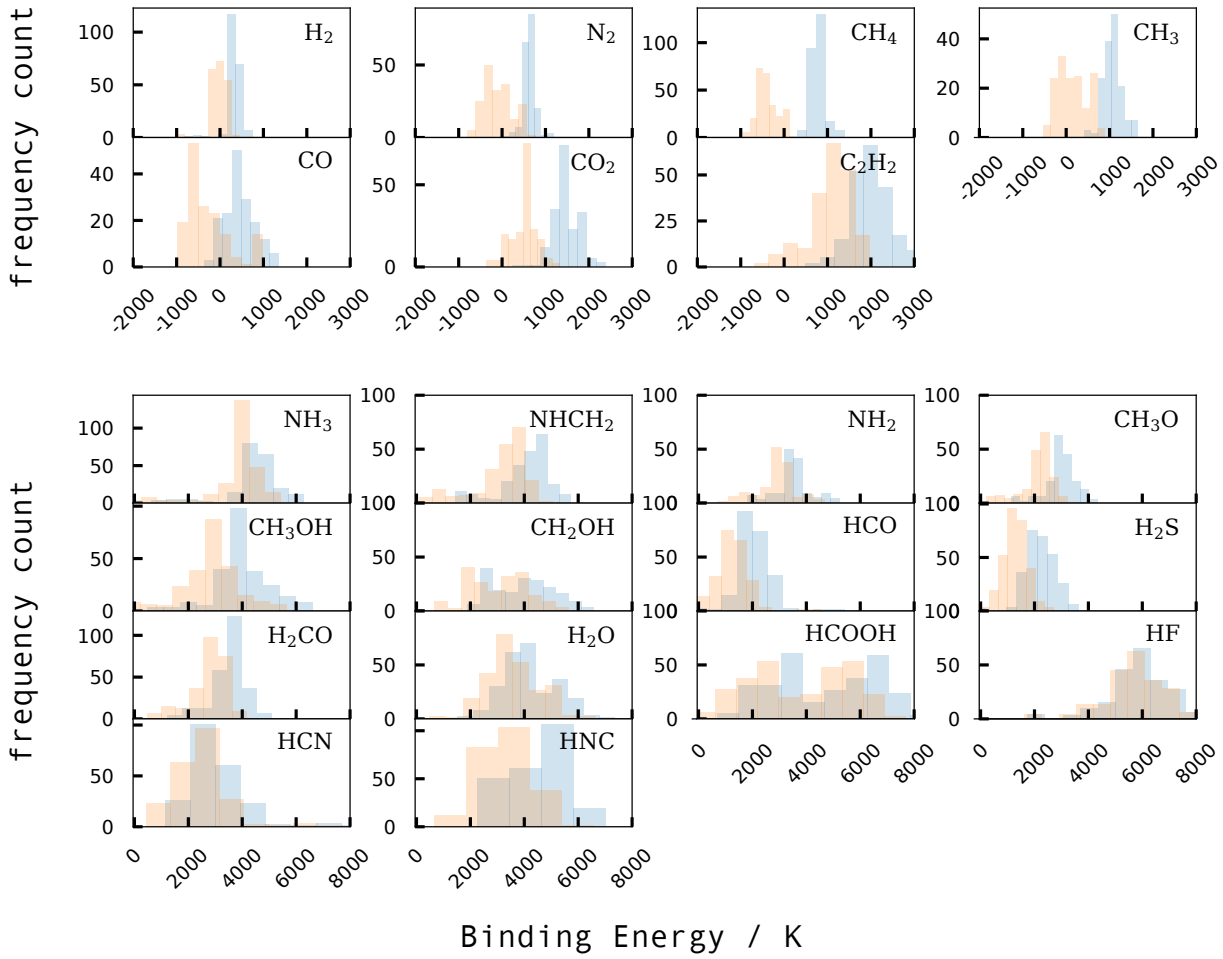

Figure 9: BE distributions computed using the best performing DFT functional from the energy benchmark for each molecule, with (blue) or without (yellow) including D3BJ correction. Upper panel: Group D; lower panel: Group H.

G. ASW CLUSTERS

In order to generate the ASW models used in this work, we employed *ab initio* annealing molecular dynamics, followed by optimization at BLYP/def2-SVP method and basis. All the steps have been conducted using Terachem software.

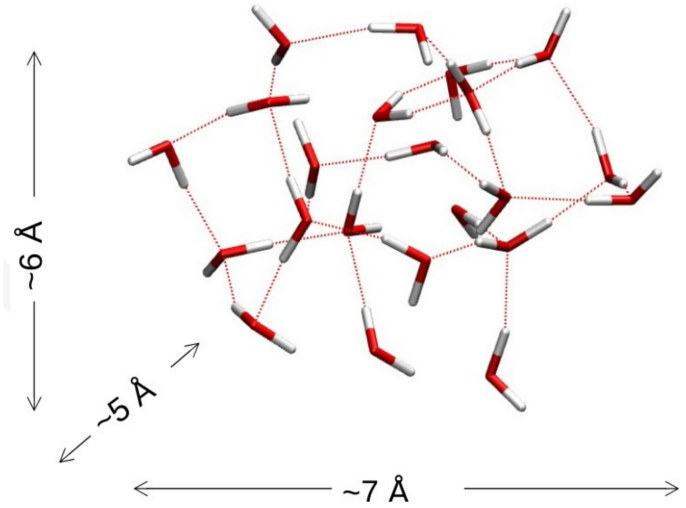

Figure 10: One of the W_{22} clusters after geometry optimization.

H. ASTROPHYSICAL FRAMEWORK

We assume a protoplanetary disk with a gas and dust temperature radial profile on the midplane $T_d(R) = T(R) = T_0(R/1\,\mathrm{au})^{-0.5}$, with $T_0 = 200\,\mathrm{K}$. The α-viscosity $\nu(R) = \alpha c_\mathrm{s}^2(R)\Omega_\mathrm{K}^{-1}(R)$ depends on the thermal speed of sound $c_\mathrm{s}(R) = \sqrt{k_\mathrm{B}T(R)\mu^{-1}m_\mathrm{p}^{-1}}$, where the mean molecular weight is $\mu = 2.34$, and m_p is the mass of the proton, and on the Keplerian angular frequency $\Omega_\mathrm{K} = \sqrt{GM_*R^{-3}}$, where G is the gravitational constant, and $M_* = 1\,\mathrm{M}_\odot$ is the mass of the central star.