The Astrophysical Journal Supplement Series, 262:17 (15pp), 2022 September https://doi.org/10.3847/1538-4365/ac7f31

Binding Energy Evaluation Platform: A Database of Quantum Chemical Binding Energy Distributions for the Astrochemical Community

Giulia M. 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; stvogtgeisse@qcmmlab.com
[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, D-85749 Garching bei München, Germany
Received 2022 April 5; revised 2022 June 17; accepted 2022 July 5; published 2022 August 24

Abstract

The quality of astrochemical models is highly dependent on reliable binding energy (BE) 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 BE distributions of molecules bound to an amorphous solid water (ASW) surface model. BEEP is highly automatized and allows one to sample binding sites on a 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. 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 use of different BE values impacts a typical problem in astrophysics, such as the computation of snow lines in protoplanetary disks. BEEP will be publicly released so that the database can be expanded to other molecules or ice models in a community effort.

Unified Astronomy Thesaurus concepts: Astrochemistry (75); Interstellar dust processes (838); Ice formation (2092); Surface ices (2117); Protoplanetary disks (1300)

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, 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 photodesorption (Minissale et al. 2022). 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 as the BE depends both on the nature of the adsorbed species and 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 built through vapor deposition and exposed to the 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 the multilayer to submonolayer 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 (ML) and multilayer regime. BEs at submonolayer 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 an 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 BEs. Two types of ASW models have been proposed: using a slab of ASW with periodic boundary conditions, or using amorphized water clusters. In the most complete study thus far, using the former approach, Ferrero et al. (2020) computed BEs of 21 molecules and atoms. Their ASW water slab consisted of 60 molecules and they computed the BEs for up to eight binding sites per molecule. The cluster approach consists of one or several water clusters 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 BE distributions of HNCO and H_2CO. More recently Duflot et al. (2021) obtained binding energies of eight different binding sites of several species (H, C, N, O, NH, OH, H_2 O, CH_3, and 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. The other 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 a density functional theory (DFT) level of theory, and only the highest BE values on each water cluster were reported. Based on this set-of-clusters approach, we developed a computational procedure to generate BE distributions and showcased the procedure on the HF molecule adsorbed on a set of 22- and 37-molecule clusters, considering 255 and 126 unique binding sites, respectively (Bovolenta et al. 2020). Recently, Germain et al. (2022) computed a BE distribution for the NH_3 molecule employing a single 200 water icy grain constructed by the semiempirical tight-binding GFN2 method. Finally, the efforts to obtain an extensive BE catalog for small molecules on water surfaces have been limited to DFT computations on small water clusters (up to six molecules, Sil et al. 2017; Das et al. 2018) or interaction with water monomer by linear semiempirical models (Wakelam et al. 2017), which do not capture the complete statistical nature of the interaction on ASW. However, using a full BE distribution reflects a more realistic desorption behavior for molecules adsorbed on ASW ice as suggested in Grassi et al. (2020). Notwithstanding, computing a large set of BEs requires a significant amount of computational resources and data management.

In this work, we present BEEP, a Binding Energy Evaluation Platform meant to offer a straightforward, highly automated and easy-to-use interface for the computation and processing of full BE distributions of molecules. To present the utility of BEEP, we computed BE distributions of 21 astrophysically relevant molecules. The platform is implemented within the QCArchive framework (Smith et al. 2020a) which allows one to transform the database in a fully open-source endeavor, from the data generation to the final user-query of 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). The initial water cluster, consisting of 22 molecules (W_{22}), has been generated by molecular dynamics, using the TIP3P model. We then performed 100 ps of high temperature (300 K) ab initio molecular dynamics (AIMD) simulation at BLYP/def2-SVP (Becke 1988; Lee et al. 1988; Miehlich et al. 1989; Weigend & Ahlrichs 2005) level of theory in order to amorphyze the system. We extracted 100 independent structures ($\tau_{\text{correlation}} \simeq$ 1 ps) from the resulting trajectory, which underwent temperature annealing of 3 ps to reach the target interstellar conditions ($\sim$10 K). We selected the 20 most representative W_{22} clusters, grouping the structures according to geometrical criteria (similarity threshold of root-mean-square deviation of atomic positions (RMSD) $\leqslant$0.40 Å). The surface spanned by these 20 clusters represents our ASW model.

The use of clusters 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 et al. 2001) geometry as a reference (see Appendix E, Table 3). We also conducted an energy benchmark, using the 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 E, Table 3). We used BLYP/def2-SVP as the 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 performed 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 (ZPVE) contribution to the BE, computed at the same level of theory as the geometry optimization.

The BE has been calculated as

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

with ΔE_{CP} being the binding electronic energy corrected for the basis set superposition error (BSSE) and Δ_{ZPVE} the ZPVE correction for the BE. See Appendices C and D and for more details. We consider the BE as a positive quantity, according to convention. 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

Quantum chemistry data has been traditionally 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 data as it is error-prone and the results are 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. 2020a). The core component of BEEP is a central server to which computation results are added in the form of JavaScript Object Notation (JSON) objects that contain the same level of information as a traditional 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 procedure (Section 3.1) and then we will discuss the database results (Section 3.2).

3.1. Computational Procedure

The procedure we developed allows us 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 procedure, we refer the reader to the QCArchive data structures we introduced in Section 2.3.

3.1.1. Sampling Procedure

In order to perform the sampling procedure (Figure 1, blue panel labeled "sampling") within of the QCArchive 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 QCArchive 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 (within 2.5 Å from 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 to geometrical criteria, such that only the structures of RMSD $\geqslant 0.40$ 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 reaching a total of at least 225 ASW–X equilibrium structures, distributed among 12-15 ASW clusters.

3.1.2. Geometry Optimization

In this step (Figure 1, yellow panel labeled "geometry optimization"), the BSCs previously obtained, are further optimized at a more accurate level of theory, such as a hybrid or metahybrid functional with a triple-ζ basis set. According to geometry benchmark results, reported in Appendix E, using a computationally affordable HF-3 c/MINIX (Sure & Grimme 2013) model chemistry can also be a good option for obtaining a refined equilibrium geometry. In the next step, BEs will be calculated carrying out single point energy computations on the BSCs, employing a DFT functional selected through an extensive energy benchmark against a CCSD(T)/CBS reference (see Appendix E).

3.1.3. Binding Energy Calculation

The final part of the procedure (Figure 1, green panel labeled "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 $\geqslant 0.25$ 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 Equations (C1) and (C2)). 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 D). 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 online.[4]

3.2. Binding Energy Distributions

We divided the molecular species into 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 (see Figure 8, Appendix G). 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-3 c/MINIX quality, as we probed it to be a cost-effective alternative to the more expensive DFT methods (see Appendix E). For CO species, the geometry is M05/def2-TZVP (Zhao et al. 2005), as HF-3 c failed to properly describe the binding sites. We computed the ZPVE correction at the HF-3 c/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

[4] www.github.com/QCMM/beep

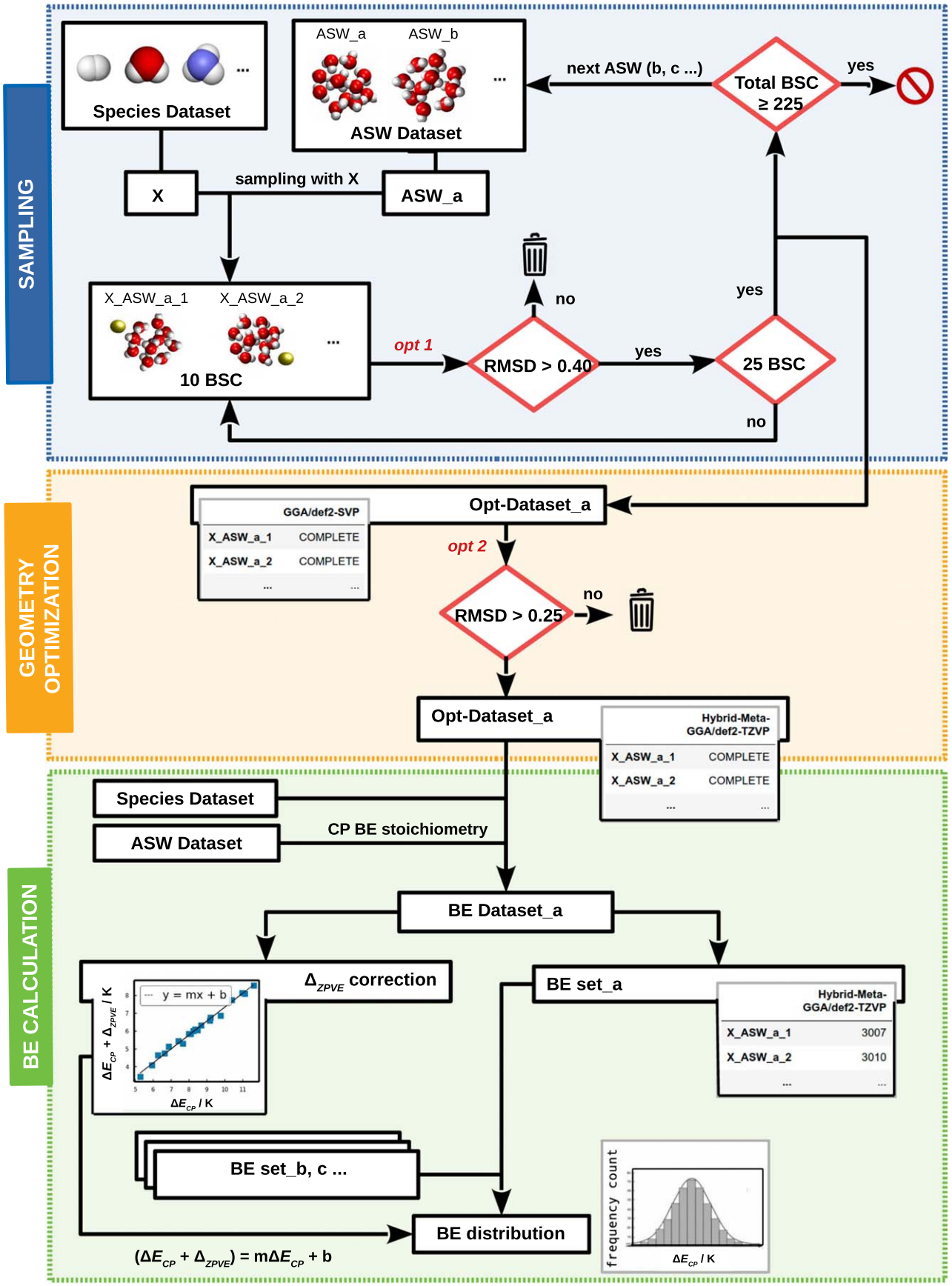

Figure 1. Three-step computational procedure used in this work for building a binding energy distribution. BSC stands for binding site candidate; *opt 1* stands for optimization at gradient generalized approximation (GGA) exchange-correlation DFT functional and *opt 2* for optimization at a higher level 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.

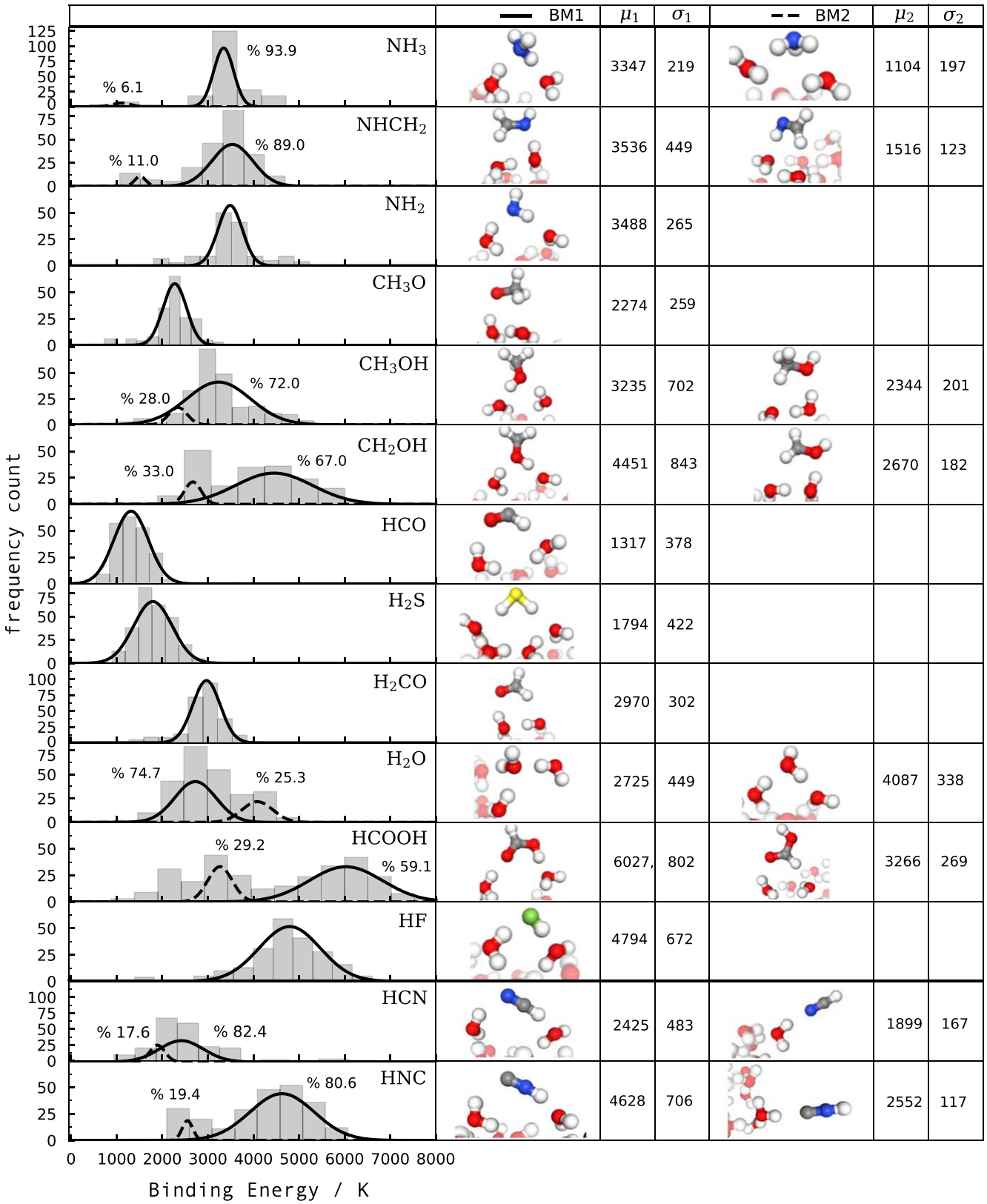

Figure 2. Binding energy distributions for Group H, ASW–X systems, using HF-3 c/MINIX geometries and including ZPVE correction. According to the benchmark results, the energy has been computed at ω − PBE/def2-TZVP level of theory for all species except HNC (B97-2/def2-TZVP), H_2CO (CAM-B3LYP/def2-TZVP), 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 F). Mean (μ) and standard deviation (σ) of the Gaussian fit are reported for the main binding mode (BM1, solid line) and the minor binding mode (BM2, dashed line). The numbers on the plot represent the percentage of minimum energy structures that belong to a specific mode. Columns 2 and 5 report a graphic representation of an example of BM1 and BM2. The atoms in proximity of the binding site have been highlighted. The color scheme for the atoms is red for O, gray for C, white for H, blue for N, yellow for S, and green for F.

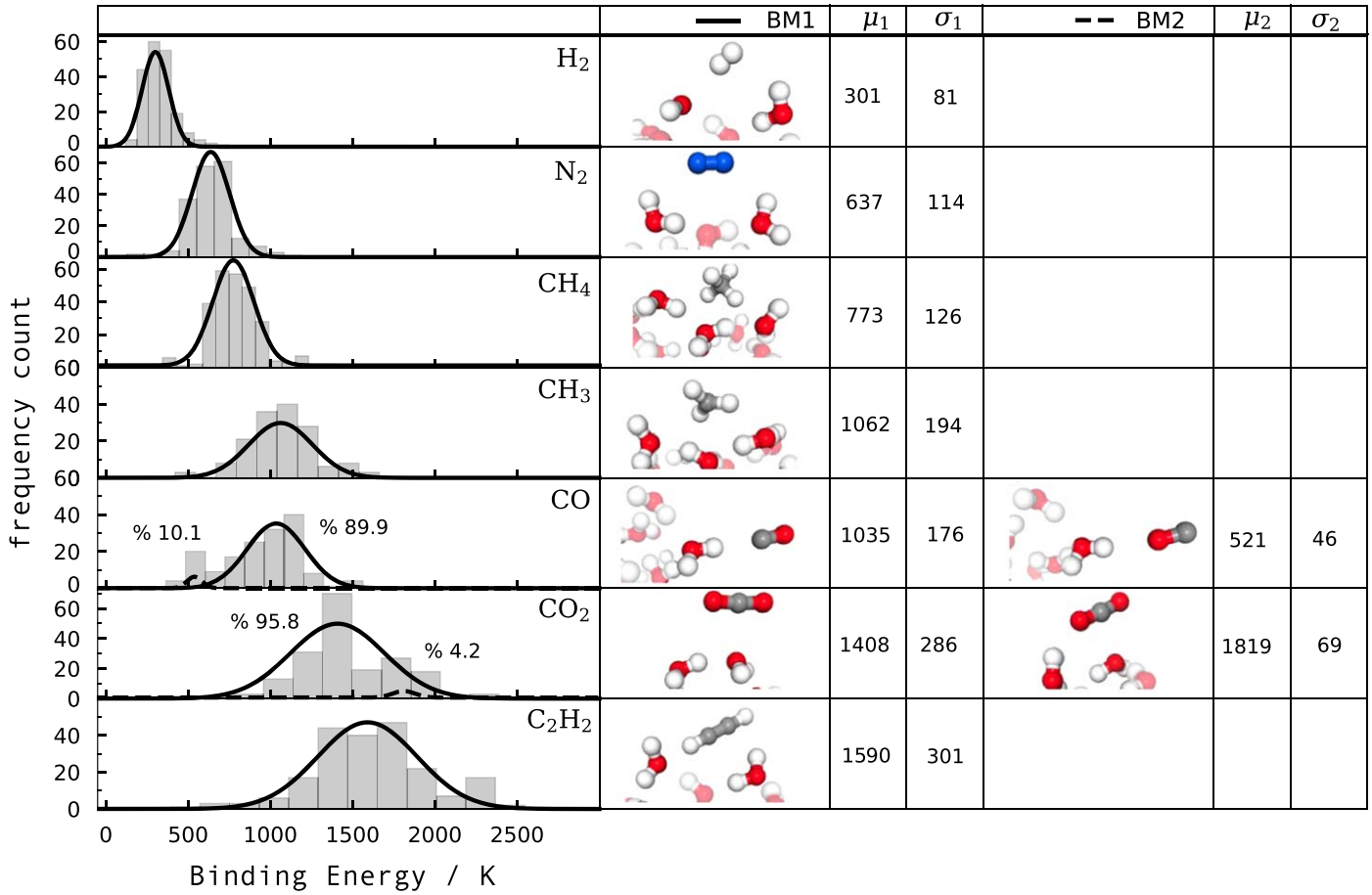

Figure 3. Binding energy distributions for Group D, ASW–X systems, using HF-3 c/MINIX geometries except for CO (M05/def2-TZVP). ZPVE correction has been included only for C_2H_2, see Section 3.2. The energy has been computed at ω-PBE/def2-TZVP level of theory for all species except CH_4 (TPSSH/def2-TZVP). D3BJ dispersion correction has been applied to all DFT energies. See Figure 2's caption for further details.

describe the potential energy well. Notwithstanding, the correction value for the 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 E). If no benchmark value is present, we used the best performing functional for each group.

Finally, while multibinding energy approaches have been recently proposed (Grassi et al. 2020), we decided to also provide a single BE value, representative of the entire distribution, to accommodate the usage of our calculations in standard chemical models. For this purpose, we obtained the mean BE (μ) and standard deviation (σ) by fitting a Gaussian function to the distribution using a bootstrap method (Appendix F). We carried out binding mode analyses in order to identify different binding motifs, which are labeled in the figure along with their percentage and their μ and σ values. An example of binding mode analysis and details about selected bond parameters can be found in a Jupyter Notebook in the BEEP GitHub repository.[5]

3.2.1. Group H: Hydrogen-bonded Structures

Figure 2 shows BE distributions of molecules in Group H. These molecules are mostly bound through electrostatic interactions in the form of hydrogen bonds and therefore present 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 distinct distributions. For NH_3 and $NHCH_2$, there is a main binding mode ($\mu_1 \sim 3400$ K) where the molecule bridges two water molecules via a double hydrogen bond. In the minor binding mode ($\mu_2 \sim 1400$ K), the surface water molecules act solely as hydrogen-bond acceptors, resulting in a lower mean BE. For CH_3OH and its radical species (CH_2OH), the main binding mode is the surface interaction via the OH moiety ($\mu_1 = 3235$ and 4451 K, respectively); while in the minor binding mode the methyl end of the molecule also participates in the interaction. 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 two different components, spanning a range of almost 7000 K. The minor mode 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 ($\mu_1 = 2725$ K) or double ($\mu_2 = 4087$ K) hydrogen bond to the surface. Even though the former occurs more often during the sampling procedure, the majority of the water molecules

[5] https://github.com/QCMM/beep/tree/main/tutorial/data_query

that compose the ASW surface form two hydrogen bonds; therefore, water surface evaporation would mostly fall within the higher BE regime. The halogen (HF) and pseudohalogen (HNC, HCN) molecules have a high standard deviation (σ ~600 K) that 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). Both HCN and HNC species exhibit two binding modes. In the main one ($\mu_1 = 2425$ and 4628 K, respectively), the molecules establish a double hydrogen bonding interaction with the surface. 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 noncorrected 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 G). 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 ($\mu_2 = 521$ K) and a second, which comprises 89.9% of the structures, and involves the C-extremity of the molecule ($\mu_1 = 1035$ 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 ($\mu_2 = 1819$ K).

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 (see Table 1). 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) and the fitting procedure for obtaining BEs from TPD temperature curves. 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 nonporous ASW (np-ASW) surface at ML and sub-ML 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 the 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 ML regime BEs is shown in Figure 4, where our BE distributions are represented as box plots. 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 us 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. built a database of BE values for W_4–X systems at MP2/aug-cc-pVDZ level of theory without correction for BSSE 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 BSSE correction that has an important effect on the final BE values for this group (BSSE correction ~100–250 K in our BE results).

Recently, Ferrero et al. proposed a new set of BE values, computed at the DFT/A-VTZ* level, including ZPVE and BSSE correction. Their single ASW model slab contains a cavity that allowed them to explore up to eight different binding sites. 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 lower BEs fall within our distribution for most of the systems, but their BEs are on average higher than the ones presented here. A possible reason for higher BE is the shape of the water cluster, which is essentially a nanocavity 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, as they offer more favorable interaction sites for the molecule on the surface. It is still uncertain to what 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. In a recent work, Germain et al. (2022) constructed an ASW cluster containing 200 water molecules and computed the BE distribution for the NH_3 molecule using the GFN2 semiempirical tight-binding method and a GFN-FF force field. Their cluster model contains some nano-cavities and the reported mean BE (4089 K) falls within Ferrero et al.'s (2020) average value (5932 K) and the one reported in this work (3347 K).

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

Table 1
Comparison with Data from the Literature

	BEEP (ASW)		He (np-ASW)[a]		Das[c]	Ferrero (ASW)[d]		KIDA[e]	UMIST[f]
	μ_1, μ_2	Max	$\theta \simeq 1$ ML	$\theta \rightarrow 0$		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	8044			3483	5382	10559	5570[n]	5000
HF	4794	6500			5540			7500	
HCl	[g]	[g]			3924	[g]	[g]	5172	900
HCN	2425, 1899	4252			2352	2496	6337	3700	2050
HNC	4628, 2552	6570			5225			3800	2050

Notes. The first column reports the molecules, columns 2 and 3 our results: the mean of the predominant binding modes identified (μ_1, μ_2) and the highest BE value of each distribution (Max). Columns 4 to 5 report He et al. experimental results at different surface coverage (θ); columns 6 to 8 BEs computed in theoretical studies, and columns 9 and 10 the values present in the astrochemical databases KIDA and UMIST. Units are in kelvin and the references are listed in the notes below.
[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).

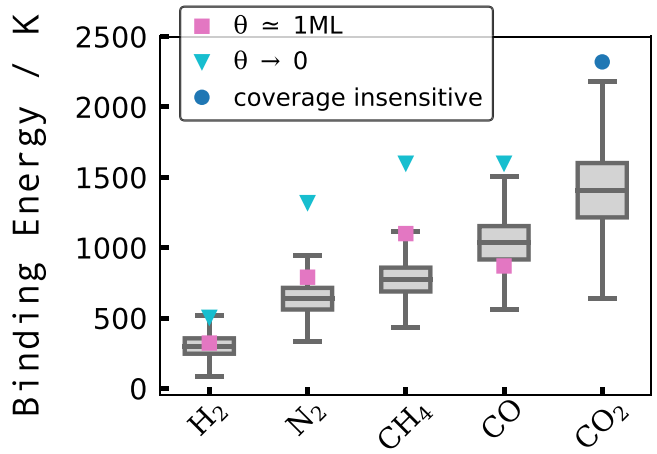

Figure 4. Box plot comparison between BE distributions presented in this work and He et al. (2016) experimental results. The box plot median corresponds to the mean of the main binding mode we identified.

computed using the two-layer ONIOM(QM:MM) approach, at ω-B97XD/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 encompasses the values of our distribution. Finally, in Figure 5, lower panel, we show the comparison of our data with the largely used KIDA and UMIST database 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 semiempirical 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 their model is based on results from different experimental setups, which makes a meaningful comparison difficult.

5. Astrophysical Implications

When comparing results between chemical experiments and quantum chemistry computations, a difference of 0.2–0.3 kcal mol^{-1} (corresponding approximately to 100–150 K) in the final BE is not substantial. However, for astrochemistry modeling, a few tens of kelvin could largely affect the final

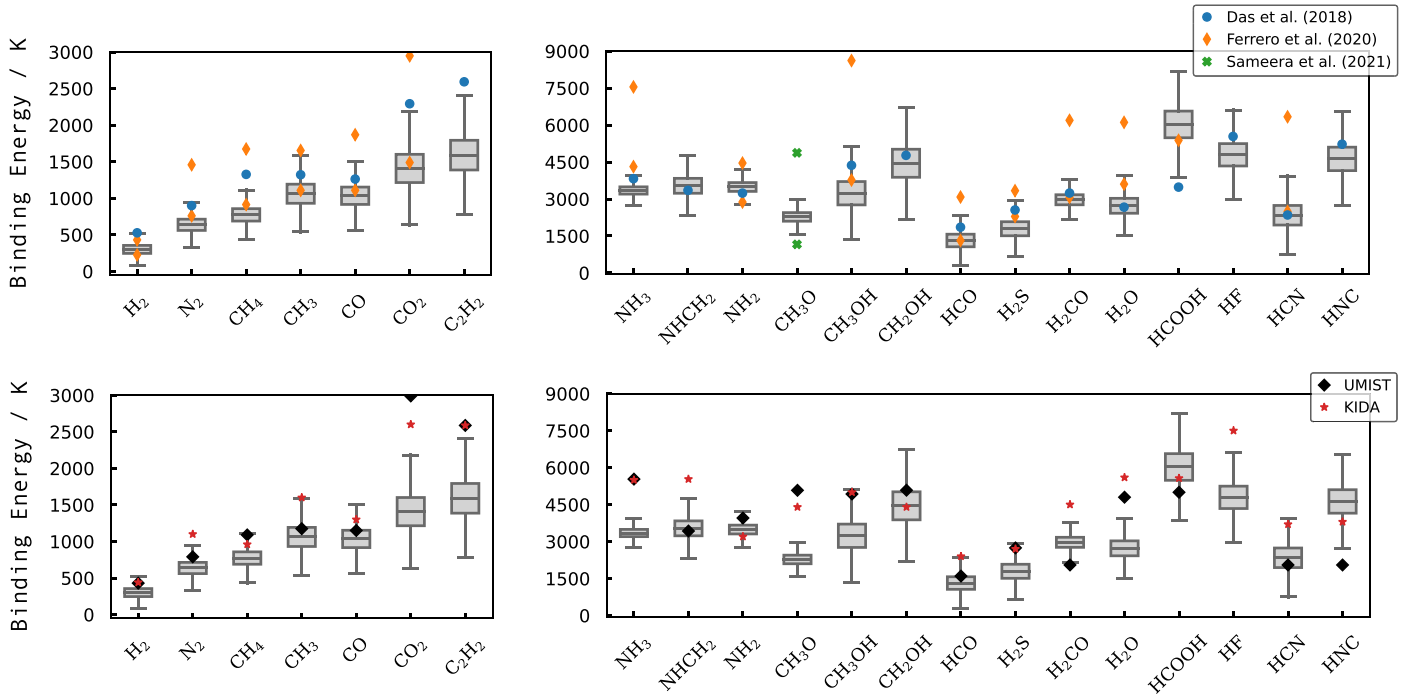

Figure 5. Box plot comparison between BE distributions computed in this work, considering only the main binding mode, and BEs present in previous theoretical studies (upper panels; Das et al. 2018; Ferrero et al. 2020; Sameera et al. 2021) and existing databases (lower panels: KIDA (Wakelam et al. 2017)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

outcome. The molecular desorption is described by the Polanyi–Wigner equation, where its dependence on the exponential of the BE 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 disk (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 $t_{\rm des}(R) = \nu_0 \exp[\Delta E_b / k_{\rm B} T_{\rm d}(R)]$, with $\nu_0 = 10^{12}$ s^{-1}, $k_{\rm B}$, and $T_{\rm d}$, respectively, the Debye frequency, the Boltzmann constant, and the dust temperature at a given radius R. The viscous time is $t_\nu(R) = R^2 \nu^{-1}(R)$, 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, assuming a temperature radial profile of $T(R) \propto R^{-0.5}$, corresponds to $\varphi_1 \ln(\varphi_2 R) - \Delta E_b \sqrt{R} = 0$, with φ_1 and φ_2 containing all the constant terms (see Appendix H for more details). The results are reported in Figure 6, where the root of the aforementioned transcendental equation is defined sublimation radius, $R_{\rm s}$. 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 molecules involved in the process of planet formation, we obtain $R = 6$ au for the BE computed by Ferrero et al. (2020; the average between their reported maximum and minimum BE values), and 9 au for the ice evaporation binding mode mean value (μ_2) computed in this work. A similar effect is reported for CO, with up to a factor of 3 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. With a binding site distribution, within the framework of this idealized disk model, we expect to observe a smoothed snow line, determined by the interplay between the temperature density profile, and the BE distribution. Increasing the distance from the star, and consequently decreasing the dust temperature, the number of available lower-energy binding sites will grow, depending on the broadening of the distribution. Conversely, a single BE will produce a sharp transition. 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 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 one 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 one to set molecules and obtain all the related BE data. In the GitHub repository[6] 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

[6] www.github.com/QCMM/beep

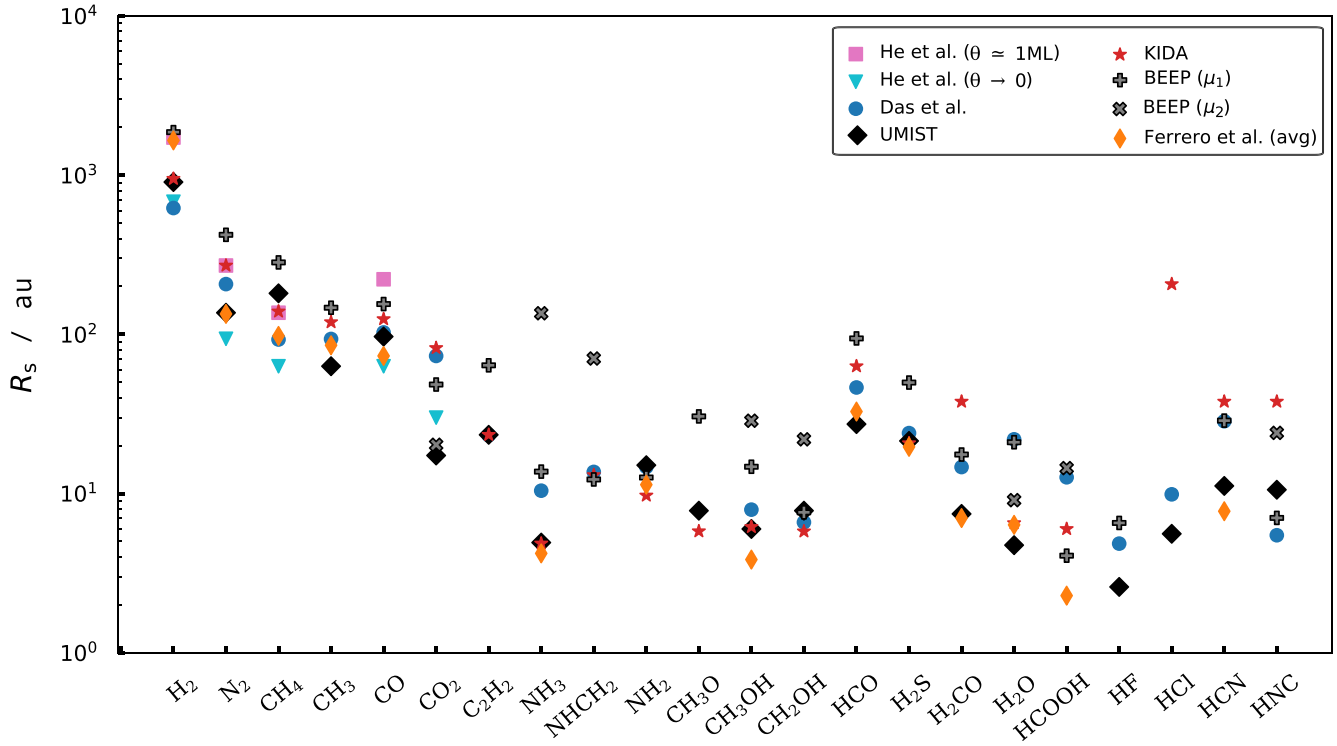

Figure 6. Sublimation radius on the midplane of a typical protoplanetary disk for different species obtained by employing the binding energies obtained in this work and compared with values available in the literature. For the sake of comparison, Ferrero et al. (2020) is the average between their upper and lower limits. See the references in Figure 5.

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 endeavor 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 a fundamental starting point to chemical reactivity studies and diffusion of molecules on the surface of interstellar ices, as having a potential energy map of neighboring 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 one to query the results produced by the protocol. BEEP consists of three highly automated steps: target molecule sampling procedure, geometry optimization of the binding site, and binding energy computation, by means of DFT methods. The binding energy distributions were obtained by sampling the ASW model spanned by a set of 12–15 amorphized water clusters containing 22 molecules each. We categorize the molecules into two groups, based on their type of interaction with the surface: molecules that are bound primarily through hydrogen bonds (Group H) and molecules for which dispersion interactions enable binding to the surface (Group D). We computed 21 binding energy distributions of astrophysically relevant molecules. Each distribution contains between 220–230 binding sites. We report mean values and standard deviation for all distributions, obtained using a Gaussian fit. Most molecules in Group H present two distributions, corresponding to different binding modes, while Group D molecules mainly have one. Group H molecules present stronger interactions with the surface, which is consistent with higher mean binding energies with respect to dispersion bound Group D molecules.

We studied the effects of our new calculations on the position of the snowline of an idealized protoplanetary disk, finding that our approach might play a relevant role in determining the correct position of the sublimation front. In particular, the sublimation radius changes by a factor of a few, for example, in the case of methane, up to an order of magnitude in the case of, e.g., methanol or ammonia. This suggests that accurate binding energies might have a marked effect on some of the key astrophysical observables.

BEEP is built on an open-source platform and hence any researcher can use it to compute binding energies with a cluster based ice surface model. Finally, we plan to transform BEEP into a widely used tool for standardized ab initio binding energy data for astrochemical modeling and ice-grain surface process studies.

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. G.M.B. gratefully acknowledges support from ANID Beca de Doctorado Nacional 21200180

and Proyecto UCO 1866—Beneficios Movilidad 2021. S.B. gratefully acknowledges support by the ANID BASAL projects ACE210002 and FB210003. S.V.G. is financially supported by ANID grant 11170949.

Appendix A
BEEP Database Access

The BE 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_41tzuDxkTtAfjPu Uq8WK5ssbnmN8QfjsApGXVYk.

Examples of how to use the class with Jupyter Notebook can be found in the tutorial section of our GitHub repository (see footnote 6).

Appendix B
Estimation of Computation Time

We report a detailed estimation of the computation time in order to produce the full BE distribution of an example molecule (constituted by around 225 binding sites), along with the computational resources used in this work.

1. Sampling procedure, carried out at BLYP/def2-SVP level of theory: 1 week, using four Tesla GPUs.
2. Geometry optimization, carried out at hybrid or metahybrid/def2-TZVP level of theory: 3 weeks, using 128 standard high-performance Intel Xeon CPUs.
3. Geometry optimization, carried out at HF-3 c/MINIX level of theory: 1 day, using 128 standard high-performance Intel Xeon CPUs.
4. BE computation, carried out at hybrid or metahybrid/def2-TZVP level of theory: 2 days, using 40 standard high-performance Intel Xeon CPUs.

Appendix C
BSSE Corrected BE Calculation Stoichiometry

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 BE of a molecule X with a basis set χ on a water cluster W with a basis set ω can be calculated as:

$$\Delta E_e = E_{XW}^{XW}(\chi \bigcup \omega) - (E_X^X(\chi) + E_W^W(\omega)). \quad \text{(C1)}$$

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 vice versa, resulting in a lower electronic energy of the supermolecule ($E_{XW}^{XW}(\chi \cup \omega)$) 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), which 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:

$$\Delta_{CP} = E_X^{XW}(\chi \cup \omega) - E_X^{XW}(\chi) + E_W^{XW}(\chi \cup \omega) - E_W^{XW}(\omega). \quad \text{(C2)}$$

Such that the resulting BE is:

$$\Delta E_{CP} = \Delta E_e - \Delta_{CP}. \quad \text{(C3)}$$

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.

Appendix D
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 noncovalently 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-3 c/MINIX), in order to obtain the zero-point vibrational energy contribution (Δ_{ZPVE}) to the BE:

$$\Delta_{ZPVE} = ZPVE_{XW} - (ZPVE_X + ZPVE_W), \quad \text{(D1)}$$

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

$$\Delta E_{CP} + \Delta_{ZPVE} = m\Delta E_{CP} + b, \quad \text{(D2)}$$

with m and b being the ZPVE correction factors. A list of correction factors for each species is reported in Table 2. Finally, the factors are applied to the set of computed BEs for each species in order to derive the ZPVE corrected BE distribution. An example plot with the linear model applied to H_2CO molecule and the code we used in order to process the computed Hessian data can be found online (see footnote 6).

Table 2
Average BE Values without ZPVE Correction and ZPVE Correction Factors

Species	μ_1, μ_2	Δ_{ZPVE1}, Δ_{ZPVE2}	m	b
C_2H_2	1979	−389	0.803	0.000
NH_3	4485, 1522	−1138, −418	0.762	0.142
$NHCH_2$	4358, 1944	822, −428	0.844	0.277
CH_3O	3038	−764	0.814	0.394
CH_3OH	4059, 3729	−823, −1385	0.819	0.170
HCO	2027	−710	0.723	0.299
H_2S	2227	−432	0.806	0.000
H_2CO	3620	−650	0.758	−0.446
H_2O	3789, 5559	−1064, −1475	0.781	0.466
HCOOH	6376, 3334	−350, −68	0.899	−0.508
HF	6006	−1211	0.798	0.000
HCN	2921, 2332	−494, −432	0.826	0.000
HNC	4983, 2748	−355, −196	0.929	0.000

Note. Column 1: species. Column 2: average BE values calculated in this work without 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.

Appendix E
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_{2-3}−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 basis. We also conducted an energy benchmark, using the W_4–X system to compare BSSE 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 metahybrid-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. We also probed the parameterized HF-3 c/MINIX (Sure & Grimme 2013) and PBEh-3 c/def2-mSVP (Grimme et al. 2015) levels of theory. The results are reported in the third 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 metahybrid functionals. 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 quality of the underlying binding site geometries, comparing the BE distribution of the equilibrium structures obtained with HF-3 c/MINIX to the best performing DFT method. Figure 7 reports the comparison for Group D, left panel, and Group H, right panel. For all species, the mean BE (μ) presents a shift passing from DFT to parameterized 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-3 c largely underestimates the BE. In light of these results, we conclude that the HF-3 c/MINIX model chemistry can be used in lieu of a more expensive DFT method, as it shows only a small difference in the position and width of the Gaussian fit of the underlying BE distributions.

Regarding the energy benchmark, Table 3, columns 4–5, 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 online (see footnote 6).

Table 3
Summary of the Results of the Geometry and Energy Benchmarks For W_{2-3}–X (W_2–X For Radicals) and W_4–X (W_3–X for Radicals) Systems, Respectively

	RMSD/Å		BEs/K	MAE/K
Group D	B3LYP(n/N)	HF-3c(n/N)	CCSD(T)/CBS	ω-PBE
H_2	0.11 (3/6)	0.14 (5/6)	320, 116	19
CO	0.12 (7/7)	0.22 (5/7)	950, 870, 791	10
CH_4	0.07 (2/2)	0.14 (2/2)	712	74
CH_3	0.09 (1/2)	0.12 (1/2)	821, 824	45
N_2	0.13 (3/3)	0.26 (3/3)		
Average	0.10 (16/20)	0.18 (16/20)		37
Group H	**PWB6K(n/N)**	**HF-3c(n/N)**	**CCSD(T)/CBS**	**ω-PBE**
NH_3	0.06 (4/7)	0.14 (5/7)	3632, 3516, 3562	79
CH_3OH	0.08 (8/8)	0.13 (6/8)	3922, 4111, 4005	119
HCOOH	0.06 (11/13)	0.17 (10/13)		
H_2CO	0.06 (5/6)	0.15 (4/6)	2600, 1197, 1181	338
HF	0.04 (3/4)	0.06 (2/4)	5956, 5380, 4158	83
HCl	0.07 (6/6)	0.18 (2/6)	3445, 2923, 956	146
HCO	0.05 (3/3)	0.07 (1/3)	2224, 1684	56
HNC	0.08 (4/5)	0.30 (3/5)	4211,3953	305
HCN	0.06 (4/5)	0.20 (3/5)		
Average	0.06 (48/57)	0.15 (36/57)		160

Note. The first column reports the molecules. Columns 2–3 report the performance of the best DFT functional for each group, and of HF-3 c. Only structures that converged (n) to the reference minima (N) were considered for the benchmark. The forth column reports reference energies calculated at CCSD(T)/CBS level of theory. The fifth column reports the mean absolute error (MAE) of the best DFT functional for each group. All DFT geometries and energies were computed using a def2-TZVP basis set and including D3BJ dispersion correction. HF-3 c method is coupled with MINIX basis set.

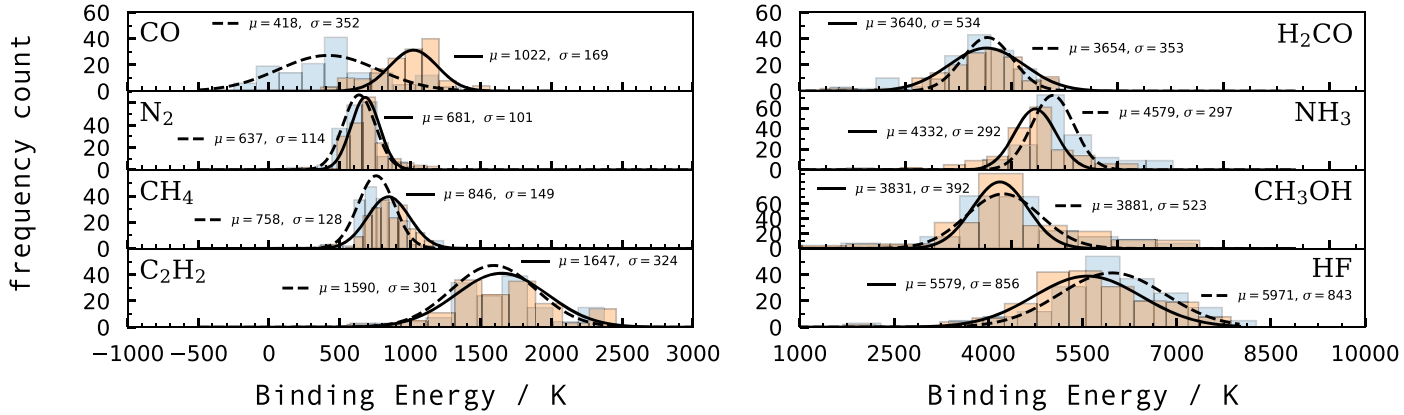

Figure 7. Comparison between BE distributions obtained using metahybrid GGA geometries (orange histogram, Gaussian function represented with solid line) and HF-3 c geometries (blue, dashed line). Left panel: Group D; right panel: Group H. BE values are shown without ZPVE correction.

Appendix F
Gaussian Fitting Procedure

To fit the BE distribution data with a Gaussian function, we employed a bootstrap method. We first divide our sample in equally spaced bins, so that each bin contains N_i samples, with a Poisson error $\sqrt{N_i}$. We then produce 10^4 distributions analog to the original data, randomizing the points assuming a Gaussian error of $\sqrt{N_i}$ around the mean N_i and we fit each distribution with

$$f(x) = a \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{F1}$$

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.

Appendix G
Dispersion Correction

Figure 8 reports the comparison between the histograms of the BE distributions computed in this work, with and without including dispersion correction (D3BJ). The low impact of this contribution on the BE is reflected in a small shift of the distributions for most of the molecules in Group H, lower panel. On the other hand, the D3BJ correction is essential for Group D molecules, upper panel, as it shifts the BE distributions into the bound regime.

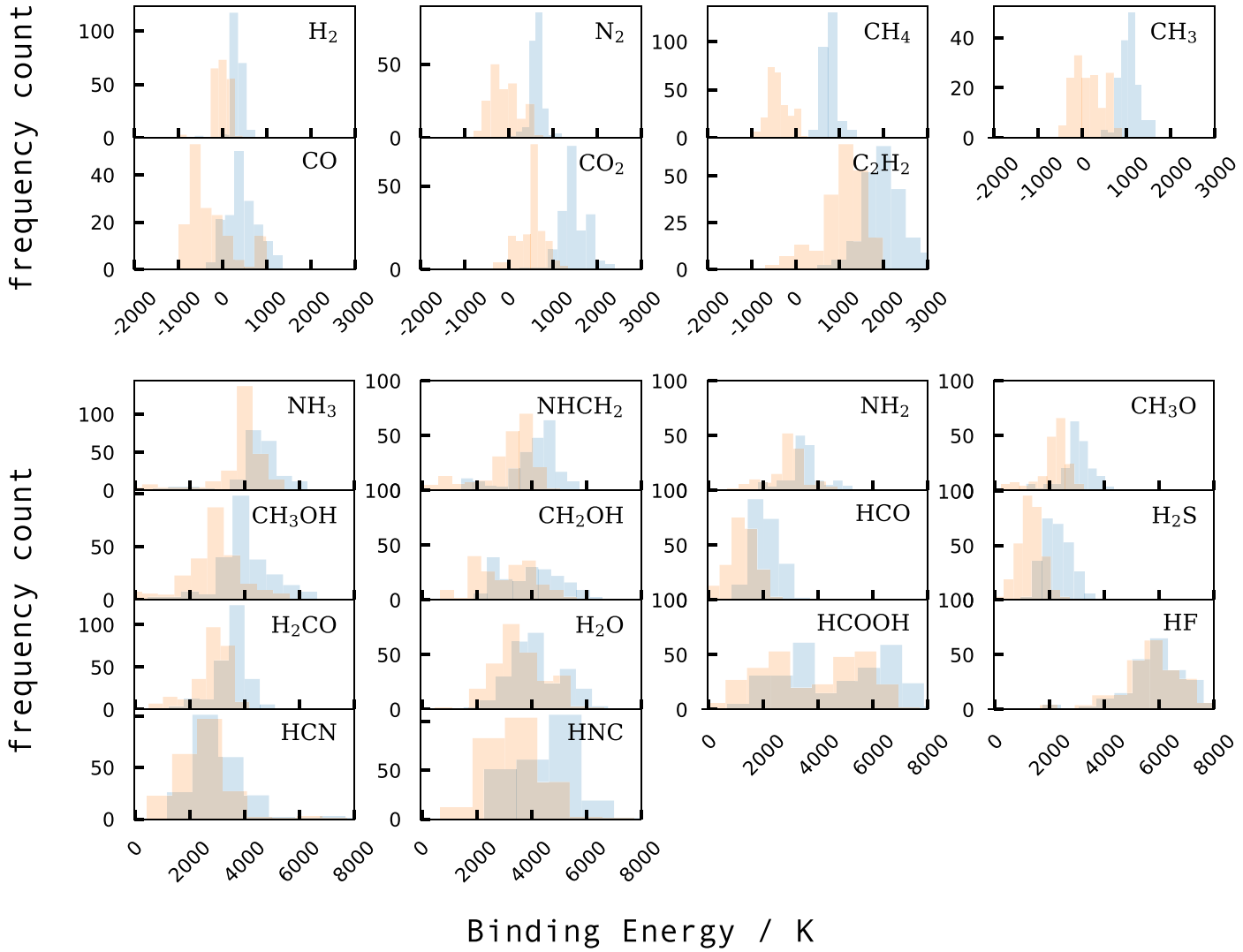

Figure 8. Binding energy distributions computed using the best performing DFT functional from the energy benchmark for each molecule, with (blue) or without (orange) including D3BJ correction. Upper panel: Group D; lower panel: Group H.

Appendix 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 $T_0 = 200$ K. The α-viscosity $\nu(R) = \alpha c_s^2(R)\Omega_K^{-1}(R)$ depends on the thermal speed of sound $c_s(R) = \sqrt{k_B T(R)\mu^{-1}m_p^{-1}}$, where the mean molecular weight is $\mu = 2.34$, and m_p is the mass of the proton, and on the Keplerian angular frequency $\Omega_K = \sqrt{GM_*R^{-3}}$, where G is the gravitational constant, and $M_* = 1\,M_\odot$ is the mass of the central star. With these definitions, the constants in the main text are

$$\varphi_1 = k_B T_0 \sqrt{R_0}\,,$$
$$\varphi_2 = \frac{\mu m_p}{\alpha k_B T_0}\nu_0\sqrt{\frac{GM_*}{R_0}}\,, \tag{H1}$$

where $R_0 = 1$ au is the position where $T(R) = T_0$, and $\nu_0 = 10^{12}$ s^{-1} is the Debye frequency.

ORCID iDs

Giulia M. Bovolenta https://orcid.org/0000-0001-8908-9109
Stefan Vogt-Geisse https://orcid.org/0000-0002-3102-1774
Stefano Bovino https://orcid.org/0000-0003-2814-6688
Tommaso Grassi https://orcid.org/0000-0002-3019-1077

References

Amiaud, L., Fillion, J. H., Baouche, S., et al. 2006, JChPh, 124, 094702
Becke, A. D. 1988, PhRvA, 38, 3098
Becke, A. D. 1993, JChPh, 98, 5648
Boogert, A., Gerakines, P., & Whittet, D. 2015, ARA&A, 53, 541
Bovolenta, G., Bovino, S., Vöhringer-Martinez, E., et al. 2020, MolAs, 21, 100095
Boys, S. F., & Bernardi, F. 1970, MolPh, 19, 553
Bozkaya, U., & Sherrill, C. D. 2017, JChPh, 147, 044104
Chai, J.-D., & Head-Gordon, M. 2008, PCCP, 10, 6615
Collings, M. P., Anderson, M. A., Chen, R., et al. 2004, MNRAS, 354, 1133
Das, A., Sil, M., Gorai, P., Chakrabarti, S. K., & Loison, J.-C. 2018, ApJS, 237, 9
Duflot, D., Toubin, C., & Monnerville, M. 2021, FrASS, 8, 24
Dunning, T. H. J., Peterson, K. A., & Wilson, A. K. 2001, JChPh, 114, 9244
Enrique-Romero, J., Rimola, A., Ceccarelli, C., et al. 2019, ESC, 3, 2158

Feller, D. 1992, JChPh, 96, 6104
Ferrero, S., Zamirri, L., Ceccarelli, C., et al. 2020, ApJ, 904, 11
Germain, A., Tinacci, L., Pantaleone, S., Ceccarelli, C., & Ugliengo, P. 2022, ESC, 6, 1286
Grassi, T., Bovino, S., Caselli, P., et al. 2020, A&A, 643, A155
Grimme, S., Brandenburg, J. G., Bannwarth, C., & Hansen, A. 2015, JChPh, 143, 054107
He, J., Acharyya, K., & Vidali, G. 2016, ApJ, 825, 89
Helgaker, T., Klopper, W., Koch, H., & Noga, J. 1997, JChPh, 106, 9639
Herbst, E., & van Dishoeck, E. F. 2009, ARA&A, 47, 427
Jorgensen, J. K., Belloche, A., & Garrod, R. T. 2020, ARA&A, 58, 727
Karton, A., & Martin, J. M. L. 2006, Theor. Chem. Acc., 115, 330
Klopper, W., & Kutzelnigg, W. 1986, J. Mol. Struct.: THEOCHEM, 135, 339
Lee, C., Yang, W., & Parr, R. G. 1988, PhRvB, 37, 785
McElroy, D., Walsh, C., Markwick, A. J., et al. 2013, A&A, 550, A36
Miehlich, B., Savin, A., Stoll, H., & Preuss, H. 1989, CPL, 157, 200
Minissale, M., Aikawa, Y., Bergin, E., et al. 2022, ESC, 6, 597
Noble, J. A., Congiu, E., Dulieu, F., & Fraser, H. J. 2012, MNRAS, 421, 768
Parrish, R. M., Burns, L. A., Smith, D. G. A., et al. 2017, J. Chem. Theory Comput., 13, 3185
Rimola, A., Skouteris, D., Balucani, N., et al. 2018, ESC, 2, 720
Ruaud, M., Loison, J. C., Hickson, K. M., et al. 2015, MNRAS, 447, 4004
Sameera, W. M. C., Senevirathne, B., Andersson, S., et al. 2021, JPCA, 125, 387
Shimonishi, T., Nakatani, N., Furuya, K., & Hama, T. 2018, ApJ, 855, 27
Sil, M., Gorai, P., Das, A., Sahu, D., & Chakrabarti, S. K. 2017, EPJD, 71, 45
Smith, D. G. A., Altarawy, D., Burns, L. A., et al. 2020a, WIREs Comput. Mol. Sci., 11, e1491
Smith, D. G. A., Burns, L. A., Simmonett, A. C., et al. 2020b, JChPh, 152, 184108
Smith, R. G., Sellgren, K., & Tokunaga, A. T. 1989, ApJ, 344, 413
Smith, R. S., May, R. A., & Kay, B. D. 2016, JPCB, 120, 1979
Song, L., & Kästner, J. 2016, PCCP, 18, 29278
Song, L., & Kästner, J. 2017, ApJ, 850, 118
Sure, R., & Grimme, S. 2013, JCoCh, 34, 1672
Titov, A. V., Ufimtsev, I. S., Luehr, N., & Martinez, T. J. 2013, J. Chem. Theory Comput., 9, 213
Ufimtsev, I. S., & Martinez, T. J. 2009, J. Chem. Theory Comput., 5, 2619
Vydrov, O. A., Heyd, J., Krukau, A. V., & Scuseria, G. E. 2006, JChPh, 125, 074106
Vydrov, O. A., & Scuseria, G. E. 2006, JChPh, 125, 234109
Vydrov, O. A., Scuseria, G. E., & Perdew, J. P. 2007, JChPh, 126, 154109
Wakelam, V., Loison, J. C., Mereau, R., & Ruaud, M. 2017, MolAs, 6, 22
Wang, L.-P., & Song, C. 2016, JChPh, 144, 214108
Weigend, F., & Ahlrichs, R. 2005, PCCP, 7, 3297
Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R., & Schütz, M. 2012, WIREs Comput. Mol. Sci., 2, 242
Werner, H.-J., Knowles, P. J., Manby, F. R., et al. 2020, JChPh, 152, 144107
Zhao, Y., Schultz, N. E., & Truhlar, D. G. 2005, JChPh, 123, 161103
Zhao, Y., & Truhlar, D. G. 2005, JPCA, 109, 5656