In-Depth Exploration of Catalytic Sites on Amorphous Solid Water: I. The Astrosynthesis of Aminomethanol†

Supplementary Material

Giulia M. Bovolenta[a,b], Gabriela Silva-Vera[a], Stefano Bovino[c,d,e], German Molpeceres[f], Johannes Kästner[g] and Stefan Vogt-Geisse*[a]

Contents

[a] *Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, Concepción, Chile*
[b] *Atomistic Simulations, Italian Institute of Technology, 16152 Genova, Italy*
[c] *Chemistry Department, Sapienza University of Rome, P.le A. Moro, 00185 Rome, Italy*
[d] *INAF - Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy*
[e] *Departamento de Astronomía, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Av. Esteban Iturra s/n Barrio Universitario, Concepción, Chile*
[f] *Departamento de Astrofísica Molecular Instituto de Física Fundamental (IFF-CSIC), Madrid, Spain*
[g] *Institute for Theoretical Chemistry, University of Stuttgart, Pfaffenwaldring 55, 70569 Stuttgart, Germany*

S1 DFT geometry and energy benchmark

Reaction 1 is the addition of NH_3 to H_2CO to give NH_2CH_2OH. The mechanism involves the nucleophilic attack of the N atom towards the C-carbonyl atom and the transfer of a proton from NH_3 to the O-carbonyl. We used $NH_3 + H_2CO + W_2$, as reference system, where the water molecules are acting as proton transfer intermediaries in the *proton relay*. The reference system geometry is DF-CCSD(T)-F12/cc-pVDZ-F12 while the reaction energies have been computed at CCSD(T)/CBS extrapolated basis set.

The energy differences are reported in Table S1. The main bond distances involved in the bond breaking/forming processes have been reported in Table S2.

Table S1: Energy barriers ($\Delta E^{\ddagger}$) and reaction energies (ΔE^{o}), using DF-CCSD(T)-F12/cc-pVDZ-F12 geometries without including ZPVE correction. Energies have been computed at CCSD(T)/CBS level of theory. Values in kcal mol^{-1}.

System	$\Delta E^{\ddagger}$	ΔE^{o}
$NH_3 + H_2CO + W_2$	10.27	-13.20

Table S2: Structural parameters involved in Reaction 1 main bond breaking/forming processes for the reference system. Bond distances (B) are in angstrom (Å), bond angles (A) are in degrees (^{o}).

Structure	B(C-N)	B(N-H1)	B(O-H2)	A(O-C-N)
R	2.79	1.01	1.81	112.3
TS	1.55	1.13	1.09	108.7
P	1.48	1.84	0.98	110.2

Around 53 DFT functionals are taken into account for the geometry benchmark belonging to different classes, and two different basis sets: def2-SVP and def2-TZVP. This is due to the fact that a double ζ basis is used to study the reaction on many binding sites on the larger surfaces, hence the need to assess the consistency of a specific DFT functional with the two tiers of method and basis. Dispersion effects are treated using D3BJ and D4 correction factors. Figure S1 shows the RMSD between the reference and the DFT level methods. Geometries of reactant (R), transition state (TS) and product (P) are taken into account. The plot displays only the methods that successfully managed to optimize all the minima. The best method is BHANDHLYP-D4/def2-SVP with a average RMSD error of 0.10 Å. It also provided the best performance using the larger basis set (in average 0.05 Å) tied with MPWB1K-D4/def2-TZVP, ω-B97-D4/def2-TZVP and PW6B95-D4/def2-TZVP. In terms of geometry, R structure has been the most challenging to optimize properly for the majority of the functionals. This is, very likely, due to the fact that R equilibrium geometry is dominated by non-covalent interactions which are more sensitive to the method used to describe them. Regarding the dispersion treatment, the error is comparable employing D4 or D3BJ correction, with slightly better results with D4, for most of the methods.

Regarding the energy benchmark, 258 DFT functionals are considered. Since a high level of accuracy is required for the energy computation, only def2-TZVP basis set has been tested.

Figure S2 shows the mean absolute error (MAE) between the methods, considering both the energy barriers (ΔE_{TS}) and the reaction energies (ΔE), and the reference geometries system. The plots report only the best 20 Meta-Hybrid-GGA and Hybrid-GGA methods that provided the best performance, ordered according to the lower ΔE_{TS} energy. On average, the reaction energies displayed higher MAE error compared to the TS energies. In order to check the consistency of a specific tier of geometry and energy DFT methods in reproducing the reference energy, ΔE_{TS} and ΔE are computed on 7 of the best DFT geometries using the best 4 DFT energies selected from the energy benchmark. The results are reported in Figure S3. The most consistent are the geometries obtained with BHandHLYP-D4 and MPWB1K-D4, as expected from the geometry benchmark results. The choice ultimately fell on BHandHLYP because MPWB1K belongs to the Meta Hybrid-GGA class and, therefore, requires larger computational time. Regarding the energy, the performance of the 4 energy methods is very similar, the best ones being BMK/def2-TZVP and ω-B97M-D3BJ/def2-TZVP with MAE error much below 1 kcal mol^{-1} on both double and triple-ζ geometries.

To summarize, the benchmark allows to identify two suitable tiers of methods for Reaction 1: ω-B97M/def2-TZVP // BHandHLYP-D4/def2-SVP and BMK/def2-TZVP // BHandHLYP-D3BJ/def2-SVP. The former is the best one and is to be used with the Orca software, while the latter is meant to be used with Gaussian software, due to the unavailability of the first tier.

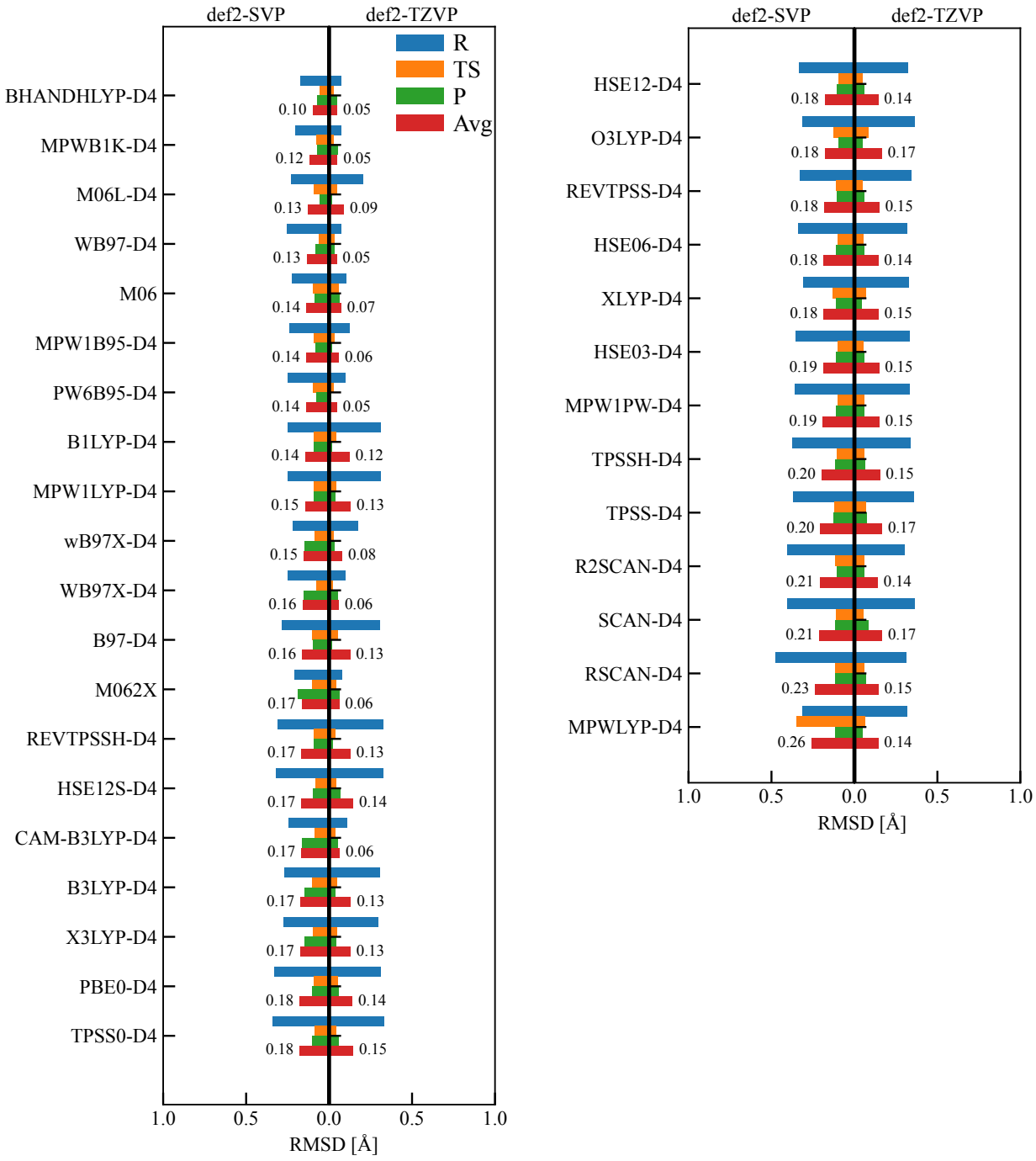

Figure S1: RMSD for different xc-functionals used in the geometry optimization for Reaction 1 on the W_2 ice model. The reference geometry is DF-CCSD(T)-F12/cc-pVDZ-F12.

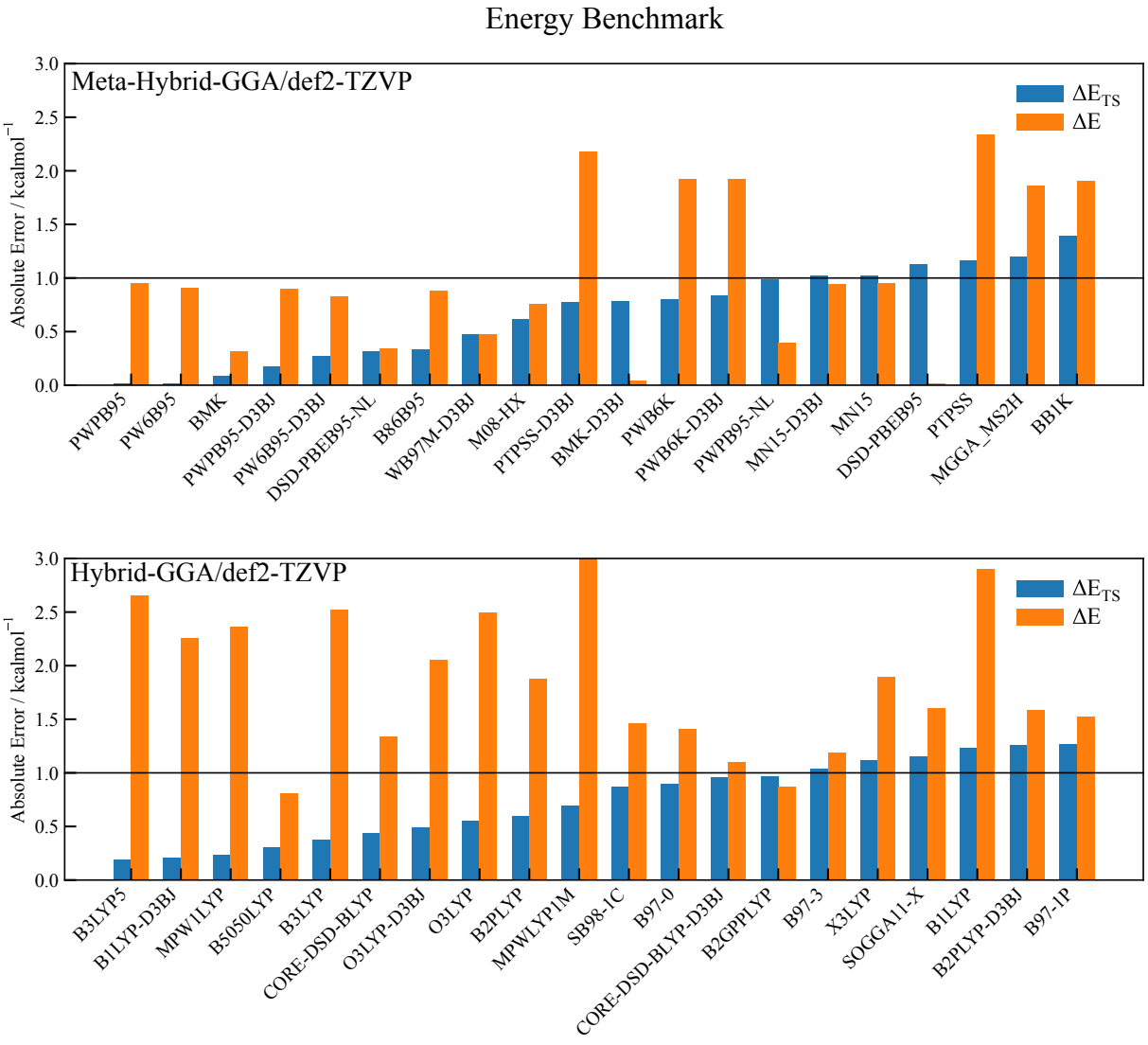

Figure S2: Energy benchmark results for Reaction 1.

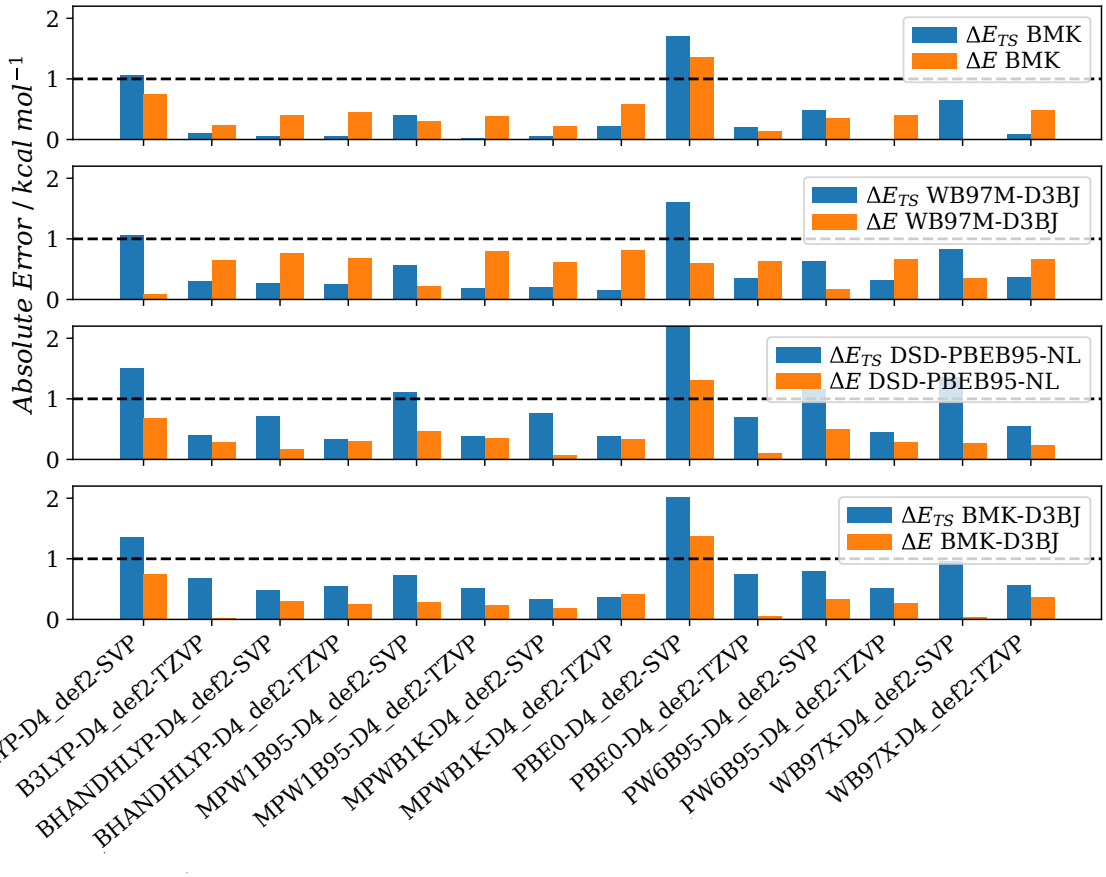

Figure S3: Condensed benchmark results for Reaction 1 on a water dimer. Energies with four different xc-functionals (figure inset) were computed on the geometries that displayed the smallest RMSD values with respect to the DF-CCSD(T)-F12/cc-pVDZ-F12 reference. Several models chemistry fall below the 1 kcal mol^{-1} chemical accuracy threshold.

S2 Energy barriers and reaction energies without ZPVE correction

Table S3: Reaction 1 carried out on different ASW sites and inside of a nano-pore, using BHandHLYP-D4/def2-SVP geometries, computed at ω-B97M-D3BJ/def2-TZVP level of theory for reaction on ASW and BMK/def2-TZVP for the nano-pore. Column one and two indicate the system and the binding regime of the reactants, column three indicates the number X of water molecules involved in the proton relay. The rest of the columns report energy barriers ($\Delta {E_{En}}^{\ddagger}$) and reaction energies ($\Delta {E_{En}}^{o}$) for step n of the reaction, if present, along with the overall TS and reaction energy ($\Delta E^{\ddagger}$ and $\Delta {E_{En}}^{o}$, respectively). Values in kcal mol^{-1}.

AMeOH formation	#	W$_X$	$\Delta E^{\ddagger}_{S1}$	ΔE^{o}_{S1}	$\Delta E^{\ddagger}_{S2a}$	ΔE^{o}_{S2a}	$\Delta E^{\ddagger}_{S2b}$	ΔE^{o}_{S2b}	$\Delta E^{\ddagger}$	ΔE^{o}
ASW clusters:										
Low-BE(NH_3)/Low-BE(H_2CO)	A	2							9.7	-11.2
	B	2							10.3	-11.2
	C	2							10.5	-11.0
	D	2							10.7	-11.5
	E	1							13.7	-17.3
High-BE(NH_3)/Low-BE(H_2CO)	A	2							9.7	-9.2
	B	2							9.9	-11.9
	C	2							10.4	-10.1
Low-BE(NH_3)/High-BE(H_2CO)	A	2	2.9	2.8	5.0	-12.9			7.8	-12.9
	B	1	2.9	2.8	7.6	-11.7			10.4	-11.7
	C	1	3.6	4.0	4.1	-11.5			8.1	-11.5
	D	2	5.5	5.6	8.0	-11.3			13.6	-11.3
High-BE(NH_3)/High-BE(H_2CO)	A	2	0.1	-10.7	12.8	-11.5			2.1	-11.5
	B	1	0.4	-1.1	7.1	-13.0			5.9	-13.0
Porous ASW:	A	2	1.59	-7.32	-1.55	-6.29	-2.48	-12.45	1.59	-12.45

Table S4: Energy results for Reaction 1, reference system (NH_3 + H_2CO + 2 H_2O). Column one indicated the level of theory used for energy and geometry (i.e. energy // geometry); column two indicates the number X of water molecules involved in the proton relay; column The rest of the columns report TS and reaction energy ($\Delta E^{\ddagger}$ and ΔE^{o}, respectively). Values in kcal mol^{-1}.

AMeOH formation	W_X	$\Delta E^{\ddagger}$	ΔE^{o}
CCSD(T)/CBS // DF-CCSD(T)-F12/cc-pVDZ-F12	2	10.3	-13.2
ω-B97M-D3BJ/def2-TZVP // BHandHLYP-D4/def2-SVP	2	10.0	-12.4
BMK/def2-TZVP // BHandHLYP-D3BJ/def2-SVP	2	10.2	-12.8

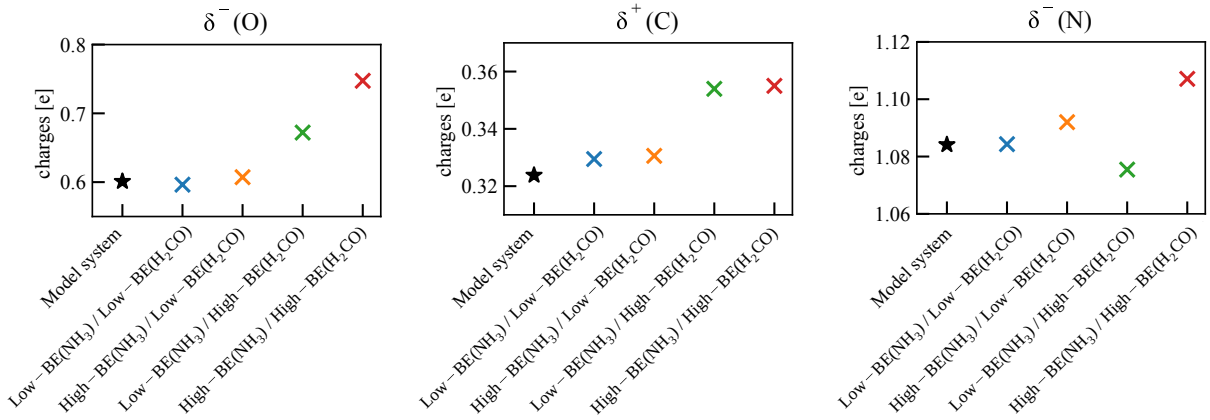

Figure S4: Average NAO partial charges estimated on carbonyl-O (left), carbonyl-C (middle), N (right), for the structures of each case. Analogous partial charges on the model system atoms are reported for reference (black star).

S3 Partial charges on O, C, N atoms of the reactive complexes

Figure S4 reports the change in average Natural Atomic Orbital (NAO) partial charges for each case with respect to the model system. Substantially, the Low-BE/Low-BE regime (blue) behaves as the model system (black star). The presence of a third H-Bond established by NH_3 (yellow), results in the increase of the magnitude of the charge on N-atom. When a third HB is established by H_2CO (green), it results in the increase of the magnitude of C and O charges. The last case, High-BE/High-BE, red, is constituted by structures which present the combination of the characteristics of the previous two cases, reaching $N_{H-Bond} = 4$ (two H-Bonds established by each reactant). Those structures present the highest intramolecular polarization on both the reactants.

S4 Example of ASW H-bond network anomaly

Figure S6, shows systems A and D belonging to Low-BE(NH_3)/High-BE(H_2CO) regime. The systems are used to illustrate an example of H-bond network coordination anomaly in the ASW. Figure S6a and S6c report reactants and TS_{S2} transition states for system A (uppper panel) and D (middle panel), respectively. It can be seen that system D lacks of a H-bond donor group (labelled as D) acting on the assisting water (W_B), with respect to system A. Thereby, the proton donor character of W_B in system D is worsened, with consequent shifting in the order of S2 events. In fact, the protonation of the assisting water W_B, usually the intermediate step in the proton relay, takes place first with respect to O−H bond formation. The result is confirmed by the bond order derivative analysis in Figure S6b and S6d. The red solid line representing the assisting water protonation ($W_B - H_B$ bond formation, in the figure), which in system A (usual order) is the second event to take place, is the first to happen in system D, instead. As a consequence, the TS_{S2} for the two systems differ considerably in terms of structural parameters and electrostatic polarization. In fact, the coordination defect in system D alters its geometry, breaking the planarity of the 4 heavy atoms establishing the proton relay ($N - O_W - O_W O_{H2CO}$). Moreover, the oxygen atom of W_B in system D presents a lone pair not engaged in any H-bond, which results in a extraordinary electron-rich region, as it can be appreciated by the maps of the electrostatic potential (MEPs), reported in Figure S6e and S6f, for system A and D, respectively. The MEPs use color to indicate electron-rich (red) and electron-poor (blue) regions.

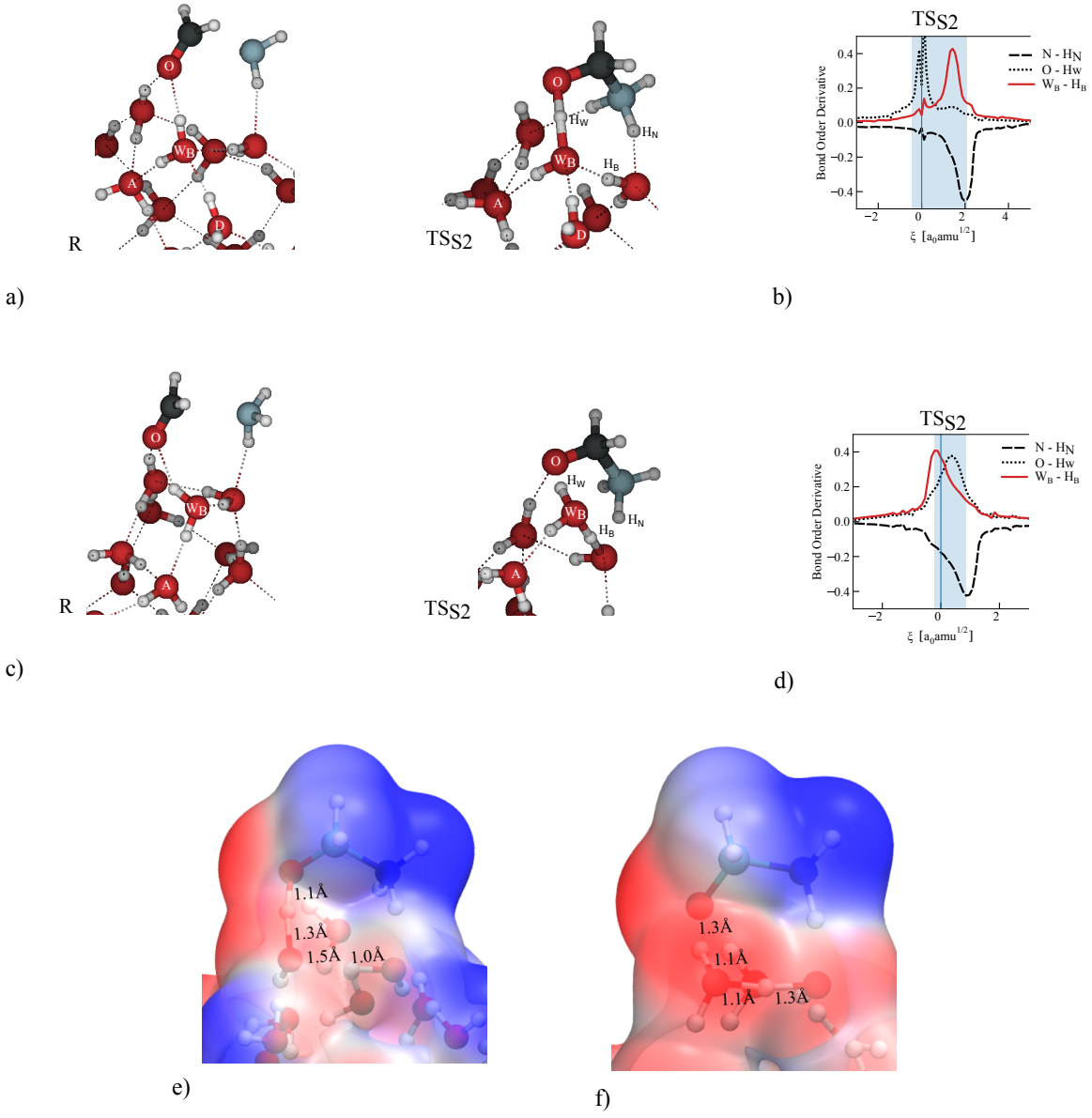

Figure S5: Upper panel: a) Geometries relative to system A reactants (R, left) and second step TS (TS_{S2}, right). H-bond donor and acceptor groups are labelled as A, D. W_B stands for bridging water, assisting in the proton relay. b) Bond order derivative for TS_{S2} events for the system A. Atoms involved in bond forming/breaking events have been labelled on TS_{S2} structure. Middle panel: c) and d) analogous of a) and b), respectively, for system D. Lower panel: e) and f) comparison between MEP of two TS_{S2} for system A and D respectively. MEP color code ranging from blue (electron-poor regions) to red (electron-rich). The color scheme for the atoms is red for O, black for C, blue for N, and white for H in the upper and middle panel, while C is in light green in the lower panel.

S5 Machine-learned interatomic potential

We trained a *ad hoc* machine-learned interatomic potential (MLIP) for the astrophysically relevant system: formyl radical (HCO) adsorbed on ASW, to be used in a subsequent work. We employed a Gaussian-Moment Neural Network (GMNN) potential[10, 11] trained on energies and atomic forces obtained from DFT. The GMNN source code is available in GitLab[a].

Training set preparation: The training data set consist of 9226 configurations, a summary can be found in Table S5. We used a variegate set of systems, such as water clusters of difference sizes (W_X, X = 30, 50, 70), as well as structures where the species HCO interacts with the water clusters (HCO+W_X, X = 22, 37, 60). The initial $W_{30,50,70}$ were generated by random placement of water molecules and subsequent geometry optimization at the GFN2-xTB[1] level of theory. HCO + W_{22} initial structures have been selected from the binding sites identified in a previous study[3], while initial HCO + $W_{37,60}$ structures are generated by random sampling of HCO on previously amorphyzed clusters, i.e. the systems underwent high temperature (300 K) *ab initio* molecular dynamics (AIMD) at BLYP/def2-SVP[2, 7, 8, 9] level of theory, followed by temperature annealing to reach $\sim$ 10 K.

The set of initial systems is then subjected to molecular dynamics (MD) simulations at various temperatures in order to perform a quick PES exploration from which MD snapshots are extracted to be refined. We employed GFN2-xTB and GFNFF[4] as propagation methods, depending on the system size (Table S5, second column), and applied a spherical boundary potential to confine the simulation, in order to prevent evaporation. The parameters of the spherical potential are chosen according to the cluster size. This procedure provides 1700 W_X points, and 7526 HCO+W_X points, for a total of 9226 configurations.

Refinement of energies and forces: Energies and forces of DFT quality are subsequently computed for the total set of configurations extracted from the MD trajectories. The DFT method used to calculate reference energy and atomic force values, is MPWB1K-D3BJ/def2-TZVP[12], selected based on the results of geometry and energy benchmarks carried out in a previous binding energy study[3]. Basis Set Superposition Error (BSSE) has been accounted for using the gCP factor, which incorporates an approximation of the counterpoise correction to the energy and forces calculation. D3BJ dispersion correction has been applied to all DFT energies and forces.

GMNN model Training: We randomly selected 2/3 of the structures from the data set in Table S5 to be used as training data, while 1/6 of the structures were used for validation and the remaining 1/6 structures were used as test data. We trained an ensemble of three MLIP models for 1000 epochs, using the same training data but with different randomly initialized parameters. The cut-off radius is set to 5.5 Å. Other relevant hyper-parameters used for the training procedure can be found in the original work[11].

In order to assess the quality of the GMNN potential, we analysed the errors obtained by the ensemble on 3 models on the test data. We obtained a mean absolute error (MAE) of 0.77 kcal/mol and 0.44 kcal/mol/Å for predicted energies and forces, respectively. The respective root-mean-square errors (RMSEs) are 1.02 kcal/mol and 0.64 kcal/mol/Å.

ASW surface modeling: We built a set of 5 periodic ice models composed of 500-water

[a] `https://gitlab.com/zaverkin_v/gmnn`

molecule each. The choice of such model size allows to account for the diverse morphological and energetic characteristics of a realistic ice. The initial 3D cell has volume (X × X × X/2) and cells dimensions (X) according to the molecular density value: 0.8 g cm^{-3}. After minimization, the system was equilibrated in conditions of constant number of molecules, volume and temperature (NVT), for 100 ps at 300 K. 5 structures are extracted from the resulting trajectories ($\tau_{correlation} \simeq 20$ ps), and underwent temperature annealing of 10 ps to reach ~ 10 K. Periodic boundary conditions are applied in two directions along the surface and we imposed a constraints to the position of the molecules in proximity of the cell's lower face (covering 1/3 of the height), as to reproduce the *bulk* of the ice models. The applied constraint ensures that the specific structural diversity of the surfaces is preserved, and it is maintained through all the following calculations.

All the simulations are carried out using a Langevin thermostat, with friction coefficients of 0.02 ps^{-1} and timestep of 0.5 fs. To run molecular dynamics calculations, GMNN is interfaced to the ASE package[5].

Table S5: Structures included in the training data. The last column refers to the number of structures extracted for refinement for each system. The propagation method is only used for sampling geometries, while energy and gradients are computed at MPWB1K-D3BJ/def2-TZVP level of theory.

System	Propagation Method (NVT)	T(K)	Number of points
W_{30}	GFN2-xTB	50/300/500	200/400/400
W_{50}	GFN2-xTB	50/300/500	167/167/167
W_{70}	GFN2-xTB	50/300/500	50/50/100
			sub-total: 1700
HCO + W_{22}	GFN-FF	100/300	150/150
	metadynamics - GFN-FF		150
HCO + W_{37}	GMNN potential[a]	100	500
	GFN-FF	300	500
	metadynamics - GFN-FF		500
HCO + W_{60}	GMNN potential[a]	100	100
	GFN-FF	300	100
	metadynamics - GFN-FF		100
			sub-total: 7526
			Total: 9226

[a] We used a previously trained (intermediate) GMNN.

S6 Computational Details

S6.1 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 analogue 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

$$f(x) = a \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{1}$$

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.

S6.2 Tri-Surface plots generation

A Triangular 3D surface (Tri-Surface) Plot is a type of surface plot, created by triangulation of compact surfaces using finite number of triangles in a manner that each point on the surface is a triangle. The intersection of any two triangles results in void or a common edge or vertex. The triangulation took into account solely atoms that belongs to the surface of the periodic ice models, based on their Z value. All surface plots are created using $ax.plot\_trisurf()$ function of matplotlib library[6].

S7 Reactive complexes categorization

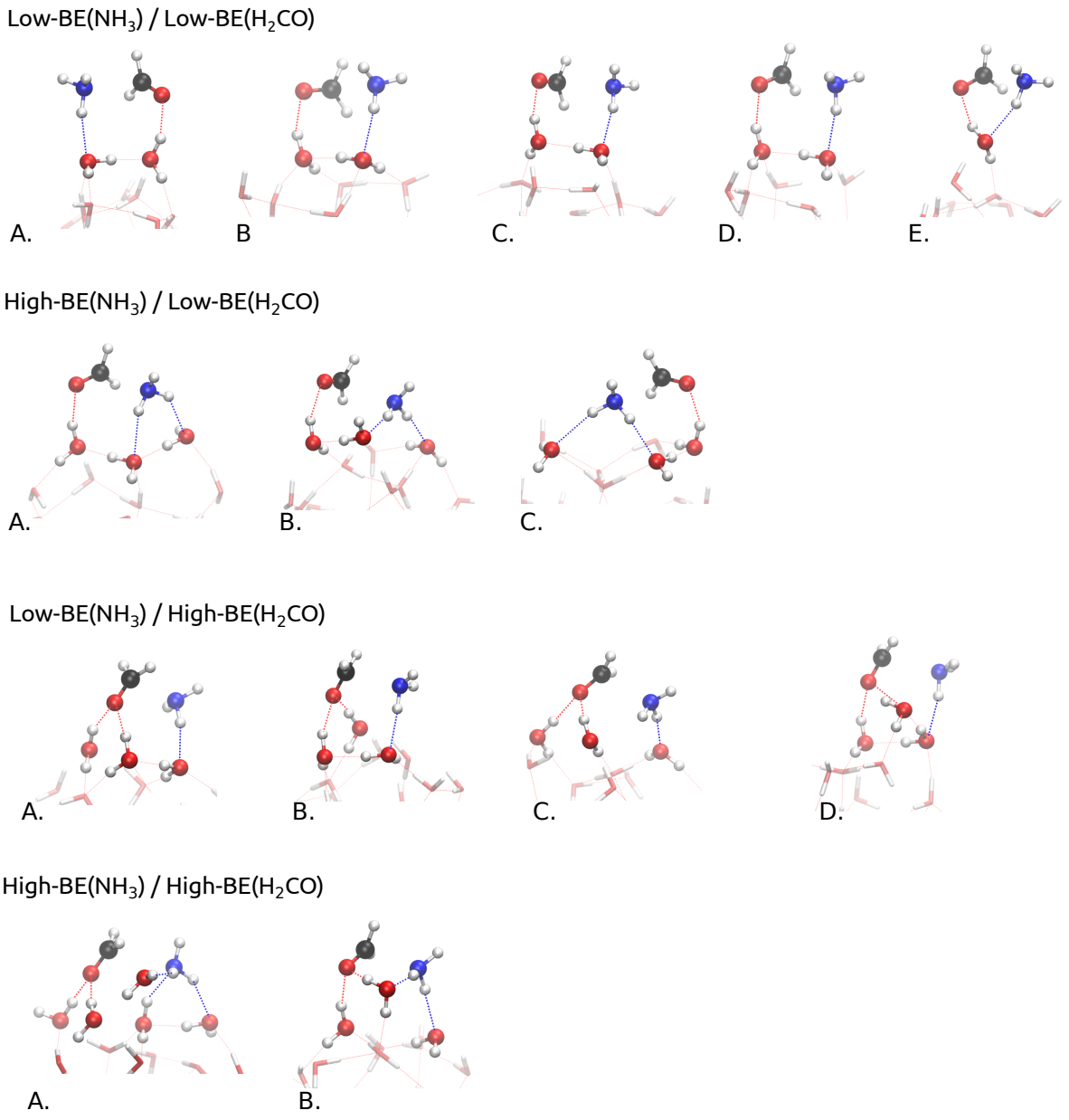

Figure S6: Reactant states listed in Table 1 and 2. Geometries obtained using BHandDHLYP-D4/def2-SVP level of theory. The color scheme for the atoms is red for O, black for C, blue for N, and white for H.

S8 XYZ coordinates

XYZ coordinates of all the states can be found in a associated zip file. The naming convention is LL for Low-BE(NH_3)/Low-BE(H_2CO), LH for Low-BE(NH_3)/High-BE(H_2CO), HL for High-BE(NH_3)/Low-BE(H_2CO) and HH for High-BE(NH_3)/High-BE(H_2CO), followed by the labelling and the state specification (R for reactant etc.). NP stands for nano-pore.

References

[1] C. Bannwarth, S. Ehlert, and S. Grimme. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation*, 15(3):1652–1671, Mar. 2019. ISSN 1549-9618. doi: 10.1021/acs.jctc.8b01176. URL https://doi.org/10.1021/acs.jctc.8b01176. Publisher: American Chemical Society.

[2] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, Sept. 1988. doi: 10/d8f7wp. URL https://link.aps.org/doi/10.1103/PhysRevA.38.3098. Publisher: American Physical Society.

[3] G. M. Bovolenta, S. Vogt-Geisse, S. Bovino, and T. Grassi. Binding Energy Evaluation Platform: A Database of Quantum Chemical Binding Energy Distributions for the Astrochemical Community. *The Astrophysical Journal Supplement Series*, 262(1):17, Aug. 2022. ISSN 0067-0049. doi: 10.3847/1538-4365/ac7f31. URL https://dx.doi.org/10.3847/1538-4365/ac7f31. Publisher: The American Astronomical Society.

[4] J. D. Gale, L. M. LeBlanc, P. R. Spackman, A. Silvestri, and P. Raiteri. A Universal Force Field for Materials, Periodic GFN-FF: Implementation and Examination. *Journal of Chemical Theory and Computation*, 17(12):7827–7849, Dec. 2021. ISSN 1549-9618, 1549-9626. doi: 10.1021/acs.jctc.1c00832. URL https://pubs.acs.org/doi/10.1021/acs.jctc.1c00832.

[5] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, July 2017. ISSN 0953-8984, 1361-648X. doi: 10.1088/1361-648X/aa680e. URL https://iopscience.iop.org/article/10.1088/1361-648X/aa680e.

[6] J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55. URL http://ieeexplore.ieee.org/document/4160265/.

[7] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, Jan. 1988. ISSN 0163-1829. doi: 10.1103/PhysRevB.37.785. URL https://link.aps.org/doi/10.1103/PhysRevB.37.785.

[8] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chemical Physics Letters*, 157(3):200–206, May 1989. ISSN 0009-2614. doi: 10/d8qp3r. URL http://www.sciencedirect.com/science/article/pii/0009261489872343.

[9] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297, 2005. ISSN 1463-9076, 1463-9084. doi: 10.1039/b508541a. URL http://xlink.rsc.org/?DOI=b508541a.

[10] V. Zaverkin and J. Kästner. Gaussian Moments as Physically Inspired Molecular Descriptors for Accurate and Scalable Machine Learning Potentials. *Journal of Chemical Theory and Computation*, 16(8):5410–5421, Aug. 2020. ISSN 1549-9618. doi: 10.1021/acs.jctc.0c00347. URL https://doi.org/10.1021/acs.jctc.0c00347. Publisher: American Chemical Society.

[11] V. Zaverkin, D. Holzmüller, I. Steinwart, and J. Kästner. Fast and Sample-Efficient Interatomic Neural Network Potentials for Molecules and Materials Based on Gaussian Moments. *Journal of Chemical Theory and Computation*, 17(10):6658–6670, Oct. 2021. ISSN 1549-9618. doi: 10.1021/acs.jctc.1c00527. URL https://doi.org/10.1021/acs.jctc.1c00527. Publisher: American Chemical Society.

[12] Y. Zhao and D. G. Truhlar. Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions. *The Journal of Physical Chemistry A*, 108(33): 6908–6918, Aug. 2004. ISSN 1089-5639, 1520-5215. doi: 10.1021/jp048147q. URL https://pubs.acs.org/doi/10.1021/jp048147q.