Supporting information for: Formic Acid Synthesis in a Water-Mineral System: Major Role of the Interface

Sara Laporte, *,^{†,‡} Fabio Pietrucci,[†] François Guyot,^{†,¶} and A. Marco Saitta[†]

†Sorbonne Univ, Museum Natl Hist Nat, UMR CNRS 7590, IMPMC, 4 Pl Jussieu, F-75005 Paris, France ‡CNR-IOM DEMOCRITOS, SISSA, Trieste, Italy

> E-mail: slaporte@sissa.it Phone: +39 (0)40 3787 443

Definition of the collective variables The path collective variables we adopted are defined in terms of coordination numbers (see Ref.^{S1}). For each atomic configuration $\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_N$ defining a putative reaction path, we adopt the following distance with respect to the atomic configuration at a given time $\mathbf{R}(t)$:

$$D(\mathbf{R}(t), \mathbf{R}_k) = \sum_{I,S} [C_{I,S}(t) - C_{I,S}^k]^2$$
(1)

with $C_{I,S}(t)$ ($C_{I,S}^k$) referring to the coordination number between atom I and all atoms of type S at time t (for frame k, respectively). $C_{I,S}$ is a smooth switching function, decaying from 1 to 0 mainly in the distance range of typical bond lengths, but with a tail extending up to at least second nearest neighbors in order to allow also the formation of new bonds:

$$C_{I,S} = \sum_{J \in S} c_{I,J} \tag{2}$$

$$c_{I,J} = \frac{1 - \left(\frac{R_{I,J}}{R_{I,J}^0}\right)^n}{1 - \left(\frac{R_{I,J}}{R_{I,J}^0}\right)^m}$$
(3)

The parameters used in this work are n = 6, m = 12, $R_{CO}^0 = 1.8$ Å, $R_{CH}^0 = R_{OH}^0 = 1.5$ Å. In this work, smoothness of the coordination space is important to introduce effective biasing forces and reconstruct smooth free energy landscapes with metadynamics and umbrella sampling.

We employ only two reference configurations in the path-CV definition (N = 2), with frame 1 referring to the CO+H₂O state and frame 2 referring to the HCOOH state. Hence we avoid any prejudice on the reactive pathways. The integer coordination numbers are given in figure S1 to illustrate how this definition of path variables can be applied. However, we employ the real-valued coordination numbers which are shown in tables S1, and which were obtained as averages from equilibrium trajectories of CO + H₂O and cis-HCOOH in bulk water. These values take into account the smooth definition of coordination numbers and leads to a better resolution of different states in the free energy landscape. Using the metric D we can define two path variables s and z:

$$s(t) = \frac{\sum_{k=1}^{N} k e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)}}{\sum_{k'=1}^{N} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_{k'})}}$$
(4)

$$z(t) = -\frac{1}{\lambda} \log \left(\sum_{k=1}^{N} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)} \right)$$
(5)

where s(t) indicates the progress the system has made in the direction of the path, while z(t) indicates the distance strayed from the path. Using the variables s and z, we can project the high dimensional ensemble of atomic configurations $\mathbf{R}(t)$ in two dimensions, for the purpose of applying biasing forces and reconstructing free energy landscapes.



Figure S1: Illustration of the path variables

The coordinates are implemented in a freely-available modified version of the plugin Plumed 1.3, compatible with DFT codes such as CPMD, CP2K and Quantum Espresso. Newer versions Plumed 2.x can be employed without modification, by exploiting suitable input structures. See http://sites.google.com/site/fabiopietrucci/home/download for code downloading and input examples.

Table S1: Real valued coordination tables for the two equilibrium states in bulk water. Left: $CO + H_2O$, right: trans-HCOOH

	С	Ο	Н		С	0	Η
C_1	0.000	1.190	0.196	C_1	0.000	1.942	1.166
O_2	0.940	0.252	0.194	O_2	0.909	0.460	0.533
O_3	0.011	0.433	2.577	O_3	0.858	0.477	1.285

Exploratory metadynamics The first step in finding the mechanism and the free energy profile of a reaction consists in exploring putative reaction pathways. ^{S2} We performed metadynamics using the topological collective variables described above, without any educated guess about the reaction pathway: the system is free to discover a path leading to the designated products, but is also free to explore alternative reactions (see examples in Refs^{S1,S3}). The atomic configurations along the first passage, under mild metadynamics conditions, are likely to pass close to the lowest free energy barrier. The simulation is continued until the system performs the backwards reaction. Once this has occurred we have a first estimate for the free energy landscape. This estimate can be improved using the newly found putative reaction pathway to perform more extensive umbrella sampling simulations.

In practice, well-tempered metadynamics was performed.^{S4} 2D gaussian bias potentials were deposited every 10 fs, the initial height of the gaussians being 0.02 Ry, with the bias factor set to 40. The width of the gaussians were 0.03 along the s coordinate, and 0.025 along the z coordinate.

Umbrella Sampling Umbrella sampling was performed with quadratic potentials in s and z, centered along the reaction pathway found in the exploratory step. Umbrella centers and parameters for the quadratic potentials were chosen ad-hoc, by firstly selecting 10 intermediate values of s and z along the path, then adding extra umbrellas to improve overlap between simulations when needed. For each free energy landscape, between 25 and 35 umbrella sampling windows were needed, as well as unbiased equilibrium runs at the two end states. The bias potential from the metadynamics simulation, albeit imperfectly

converged, was also applied on top of the umbrella potentials to partially flatten the landscape and improve the efficiency of umbrella sampling. Each umbrella sampling simulation in the bulk was run for 5 ps, discarding the first pico-second as equilibration time. Due to a larger computational cost, the MgO/water interface simulations were run for 2.5 ps in each window. The weighted histogram analysis method (WHAM)^{S5} was applied to obtain free energy profiles. The equilibrium trajectories of reactants and products were extended to 14 ps (bulk water and electric field cases) and 6 ps (surface case). The resulting 2D free energy landscapes from umbrella sampling are shown in figure S2.

Uncertainty estimates Statistical uncertainties were estimated by evaluating the ergodicity of the umbrella sampling simulations. For each umbrella trajectory the data (discarding the first ps) was split into two 2 ps blocks and WHAM was applied to each set independently. The value at the free energy minima and height of the barrier was then taken as the average of the two, with an uncertainty taken as the difference between the two 2 ps data sets. Similarly for the surface simulation, the data was split into two 1 ps sets (discarding the initial 0.5 ps). Our conservative analysis indicates an error bar of at most ± 2 kcal/mol.

Validation of transition state configurations To find suitable transition states, configurations were taken from the initial exploratory metadynamics, and left to evolve freely until reaching either reactants or products, with initial random velocities drawn from the Maxwell-Boltzmann distribution at the given T (committor analysis). Ten replicas were initialised to test each configuration, and if 4-6 of those ended up in the $CO+H_2O$ state, and the remaining ones in HCOOH, then the configuration was assumed to be a transition state. Of course a number of transition state configurations may be explored for a same mechanism when the solvent is explicitly simulated, but we employed the first passage occurring in the exploratory metadynamics with moderate bias deposition rate to minimize the risk of extracting improbable transition state configurations.



Figure S2: 2D free energy landscapes reconstructed with umbrella sampling for the targeted reactions, at 300K and 400K.

References

- (S1) Pietrucci, F.; Saitta, A. M. Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios. *Proc. Natl. Acad. Sci. U.S.A* **2015**, *112*, 15030–15035.
- (S2) Pietrucci, F. Strategies for the exploration of free energy landscapes: unity in diversity and challenges ahead. *Rev. Phys.* 2017, 2, 32–45.
- (S3) Pietrucci, F.; Aponte, J. C.; Starr, R. D.; Pérez-Villa, A.; Elsila, J. E.; Dworkin, J. P.; Saitta, A. M. Hydrothermal decomposition of amino acids and origins of prebiotic meteoritic organic compounds. ACS Earth Space Chem. 2018, doi: 10.1021/acsearthspacechem.8b00025.
- (S4) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (S5) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J. Comput. Chem. 1992, 13, 1011–1021.