

Supporting Information

Synthesis of (D)-erythrose from glycolaldehyde aqueous solutions under electric field

Giuseppe Cassone*^{1a}, Jiri Sponer^{1,2b}, Judit E. Sponer^{1,2c},
Fabio Pietrucci^{3d}, A. Marco Saitta^{3e}, Franz Saija^{4f}

¹ *Institute of Biophysics,*

*Czech Academy of Sciences, Královopolská 135,
61265 Brno, Czech Republic*

² *Regional Centre of Advanced Technologies and Materials,*

*Department of Physical Chemistry,
Faculty of Science, Palacky University,
17. listopadu 12, 77146 Olomouc, Czech Republic*

³ *Sorbonne Universités,*

*Université Pierre et Marie Curie Paris 06,
Institut de Minéralogie,
de Physique des Matériaux et de Cosmochimie,
CNRS, Muséum national d'Histoire naturelle,
Institut de Recherche pour le Développement,
Unité Mixte de Recherche 7590,
F-75005 Paris, France*

⁴ *CNR-IPCF, Viale Ferdinando Stagno d'Alcontres 37,
98158 Messina, Italy*

(Dated: January 3, 2018)

^a Email: cassone@ibp.cz

^b Email: sponer@ncbr.muni.cz

^c Email: judit@ncbr.muni.cz

^d Email: fabio.pietrucci@impmc.upmc.fr

^e Email: marco.saitta@impmc.upmc.fr

^f Email: saija@ipcf.cnr.it

I. *AB INITIO* MOLECULAR DYNAMICS SIMULATIONS

We used the plane-wave/pseudopotentials software package Quantum ESPRESSO [1], based on the Car-Parrinello (CP) approach [2], to perform a series of *ab initio* molecular dynamics (AIMD) simulations of four liquid samples (*i.e.*, neat formaldehyde, a formaldehyde and water mixture, neat glycolaldehyde, and a glycolaldehyde aqueous solution) at ambient temperature and under the action of intense electric fields applied along a given direction (corresponding to the z -axis). The implementation of an external field in numerical codes based on density functional theory (DFT) can be achieved by exploiting the Modern Theory of Polarization and Berry's phases [3–5] (see, *e.g.*, Ref. [6]). The neat formaldehyde sample contained 48 molecules (*i.e.*, 192 atoms) arranged in a cubic cell with side parameter $a = 14.32 \text{ \AA}$ whereas the formaldehyde aqueous mixture simulation box contained 26 formaldehyde and 26 water molecules (*i.e.*, 182 atoms) arranged in an equally-shaped cell with side $a = 12.80 \text{ \AA}$. As far as the glycolaldehyde-containing samples are concerned, the neat one contained 24 molecules (*i.e.*, 192 atoms) placed in a cubic cell with side parameter $a = 13.17 \text{ \AA}$ whereas the glycolaldehyde aqueous solution simulation box contained 12 glycolaldehyde and 32 water molecules (*i.e.*, 192 atoms) arranged in a cubic cell with side $a = 12.55 \text{ \AA}$. In order to check whether the CP technique introduced any kind of artifact to the main simulation here presented (*i.e.*, that one referred to the glycolaldehyde aqueous solution sample), the same calculation has been reproduced also by using the software package CP2K [7], exploiting the Born-Oppenheimer (BO) approach and utilising the same Berry's phases framework for the implementation of the electric field [6]. As expected, we obtained exactly the same result: the synthesis of (D)-erythrose has been recorded for a field intensity of 0.45 V/\AA . As usual, all the structures were replicated in space by means of periodic boundary conditions.

In order to avoid eventual spurious effects stemming from the size limitations of the simulation boxes, the same simulation carried out for the glycolaldehyde aqueous mixture has been performed in a bigger cubic cell having side parameter $a = 14.32 \text{ \AA}$ and composed of 48 water and 18 glycolaldehyde molecules (*i.e.*, 288 atoms). It turned out that the first synthesis of (D)-erythrose has been achieved for the same field strength recorded in the smaller simulation boxes (*i.e.*, 0.45 V/\AA). This latter finding strongly confirms the scalability and the reproducibility of the presented results.

In each sample, we gradually increased the intensity of the electric field with a step increment of about $0.05 \text{ V}/\text{\AA}$ from zero up to a maximum of $0.45 \text{ V}/\text{\AA}$. In the zero-field cases we executed a dynamics of almost 25 ps whereas, for each other value of the field intensity, we ran the dynamics for at least 2 ps, thus cumulating a total simulation time of about 45 ps for each simulation. Moreover, once the chemical reaction leading to the production of (D)-erythrose occurred in the glycolaldehyde aqueous solution, a rough stability analysis of the formed species was carried out. Indeed, in parallel CP molecular dynamics simulations, the electric field has been switched off and the dynamics of the systems have been followed for times longer than 25 ps. It turned out that the synthesized sugars are stable upon the field shutdown in the simulation boxes where (D)-erythrose has been detected.

The fictitious electronic masses were set to a value of 300 a.u., with a cutoff energy of 40 Rydberg (Ry) and a cutoff energy for the charge density of 320 Ry, which allowed us to adopt a timestep of 0.12 fs. With such cutoff values the samples are well-described since the core electronic interaction is being depicted through ultrasoft pseudopotentials (USPP). As for exchange and correlation effects, we adopted the Perdew-Burke-Ernzerhof (PBE) functional [8]. Moreover, in the simulation performed with the CP2K code we employed the dispersion-corrected version of the PBE functional [9, 10]. The dynamics of ions was simulated classically within a constant number, volume, and temperature (NVT) ensemble, using the Verlet algorithm and a Nosé-Hoover thermostat set at a frequency of 13.5 THz whereas the velocity rescaling method [11] has been employed to fix the temperature during the BO molecular dynamics simulation carried out by means of the CP2K code.

For the sake of completeness, in the following subsections additional remarks that concern some of the performed simulations are presented.

A. Neat formaldehyde

As shown in Fig. 2 of the main text, a system of pure liquid formaldehyde subjected to the electric field action undergoes to an almost complete alignment of the dipole vectors characterizing the formaldehyde molecules even for relatively moderate field strengths (*i.e.*, $0.20 \text{ V}/\text{\AA}$). For higher intensities, however, this molecular alignment results in the formation of longer and longer polymers composed by H_2CO monomers and, for the maximum field strength investigated (*i.e.*, $0.45 \text{ V}/\text{\AA}$), the formation of a paraformaldehyde polymer has

been observed, as shown in Fig. S1. Although the intense field regimes explored, no formyl

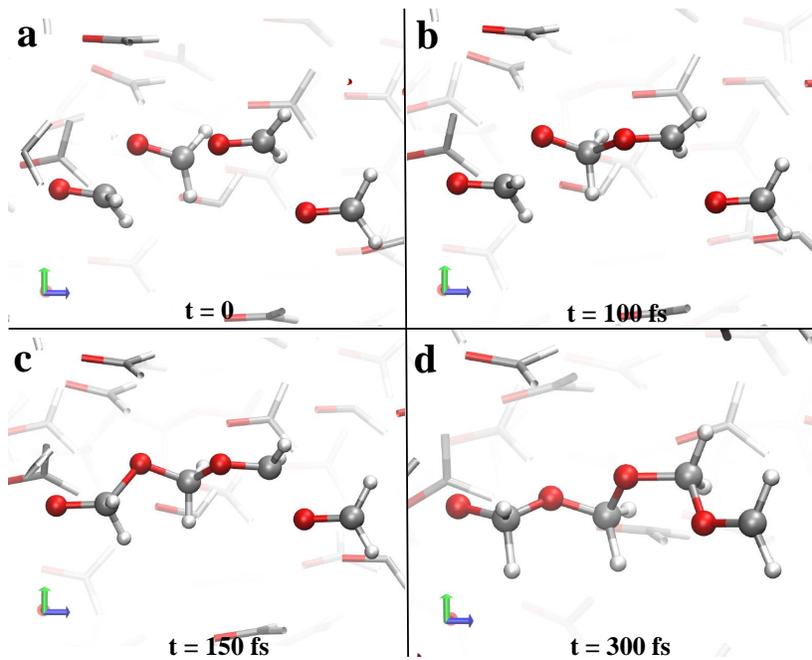


FIG. 1. Paraformaldehyde polymer synthesized at a field strength of 0.45 V/\AA and oriented along the z -axis (*i.e.*, the blue Cartesian axis) in a sample of pure formaldehyde. Please notice the collective orientation of the nucleophilic formaldehyde heads “against” the field orientation.

anion formations have been observed.

B. Formaldehyde aqueous solution

Additionally to polymers similar to that one shown in the inset of Fig. 2 of the main text, also the formation of formaldehyde monohydrate has been detected in a formaldehyde and water mixture. Indeed, the reactivity of such a system is manifest by the fact that not only the formation of the water counterions occurs for field strengths significantly lower than those necessary in neat water (as laid out in the main text), but also because the polymerization reactions shown in Fig. S1 and in the inset of Fig. 2 of the main text take place for field intensities of 0.20 V/\AA , a threshold that in neat water is not even capable to trigger “trivial” water dissociations. However, the hydration of formaldehyde – leading thus to the formaldehyde monohydrate formation – occurs for the first time at 0.35 V/\AA by means of the mechanism shown in Fig. S2. Here, similarly to the case of the (D)-erythrose synthesis, a proton transfer event triggers the subsequent reaction, as depicted in Fig. S2-a-b. In fact, simultaneously to the formation of an hydronium ion (Fig. S2-b) the mutual

approach of the newly formed hydroxide and a formaldehyde molecule takes place (Fig. S2-b-c). Moreover, a concerted release of a proton from another hydronium (Fig. S2-c) leads to the synthesis of a formaldehyde monohydrate species (Fig. S2-d).

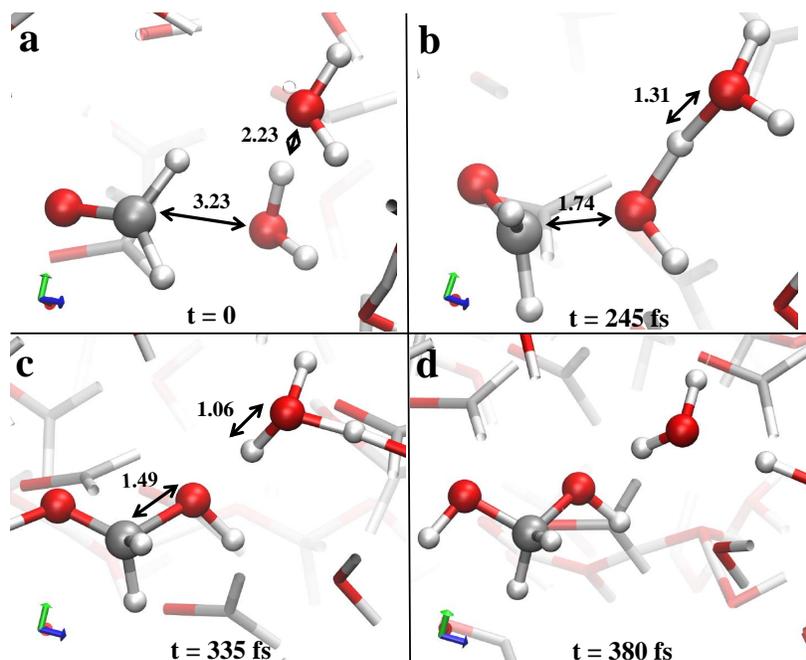


FIG. 2. Formaldehyde monohydrate formation mechanism at a field strength of 0.35 V/\AA from a liquid sample of formaldehyde and water at room temperature. The field orientation coincides with the z -axis (*i.e.*, the blue Cartesian axis). Please notice that all the proton transfers occur prevalently along the field direction whereas the mutual approach of the forming OH^- anion and the formaldehyde molecule (b) is driven by the field. The most negatively charged portion of the anion, as well as the negatively charged head of the formaldehyde species orient in opposition to the field orientation (b-c), leading to the formation of a new C-O bond (c). Subsequently, a proton transfer along the field direction neutralizes the newly formed anion synthesizing a formaldehyde monohydrate species (d).

C. Neat glycolaldehyde

A system composed of neat glycolaldehyde subjected to intense electric fields was not capable to undergo any chemical reaction beyond some feeble proton transfers, between the glycolaldehyde OH groups, occurring for intensities stronger than $0.35 - 0.40 \text{ V/\AA}$.

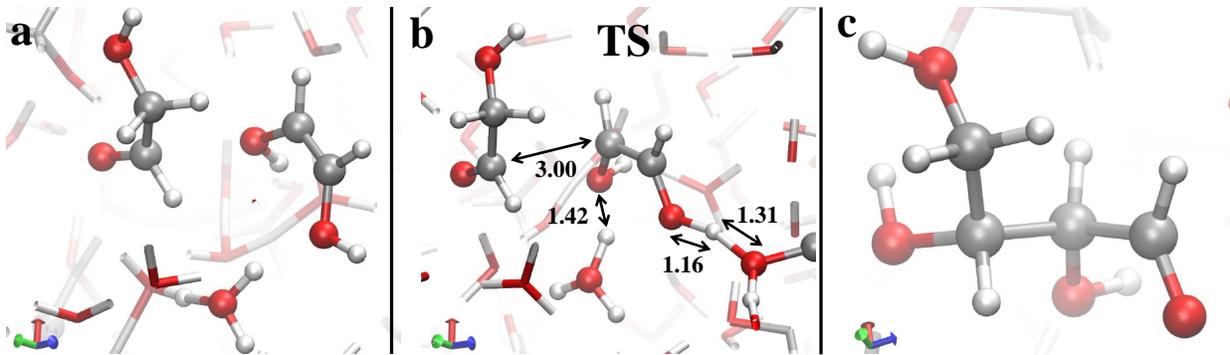


FIG. 3. Transition state (TS) (b) of the reaction leading to the (D)-erythrose synthesis for a field strength of 0.45 V/\AA in a glycolaldehyde aqueous mixture at room temperature. The field is oriented along the z -axis (corresponding to the blue Cartesian one). By launching 50 independent AIMD simulations from the TS structure and differing by the initial atomic velocities, 60% and 40% of them fall in the reactants (a) and products (c) configuration, respectively.

II. TRANSITION STATE DETECTION OF THE (D)-ERYTHROSE SYNTHESIS: COMMITTOR ANALYSIS

With the aim of fully characterizing the transition states (TS) of the reaction leading to the synthesis of (D)-erythrose from a glycolaldehyde aqueous solution at a field strength of 0.45 V/\AA , committor analysis [12] has been performed. By shooting many unbiased trajectories, each one starting from a structure thought to be a plausible TS, the most likely TS associated with the condensed-phase reaction synthesizing (D)-erythrose at finite temperature has been identified. In particular, 12 different structures connecting the reactants and the product from the unbiased AIMD simulation have been considered. It turns out that the configuration marked as TS in Fig. S3 represents the most likely TS. In fact, by launching 50 independent simulations from this latter structure in presence of a field intensity equal to 0.45 V/\AA and with pseudo-random starting nuclear velocities taken from a Maxwell-Boltzmann distribution at 300 K, it has been observed that $50 \pm 10\%$ of the trajectories fell either in the reactants (Fig. S3-a) or in the product (Fig. S3-c) basin. More precisely, 60% of them fell in the reactants whereas 40% of them in the product basin, respectively (Fig. S3-a/c), within 500 fs of dynamics. In Fig. S3 the concerted reaction leading to erythrose, along with several atomic distances (in \AA) topologically defining the TS structure, is shown.

III. DENSITY FUNCTIONAL THEORY CALCULATIONS

The Wannier centres calculation is based on the evaluation of the Maximally Localised Wannier Functions (MLWF) [13, 14] as implemented in Quantum ESPRESSO by means of the Wannier90 [15] code. In particular, after a standard self-consistent field (SCF) calculation conducted at the field strength that render barrierless the reaction producing (D)-erythrose in a glycolaldehyde aqueous solution (*i.e.*, 0.45 V/Å), a band calculation at the Γ point has been carried out in order to obtain the Bloch states. Because of the usage of USPP mimicking the core electronic interaction, such a liquid system is characterized by 272 bands. This value corresponds also to the number of the MLWF and thus to the number of the Wannier centres. The used formalism [15] works by minimizing the total spread of the MLWF in the real space and this is done in the space of unitary matrices that describe rotations of the Bloch states. As a consequence, the kind of calculation implemented in Wannier90 is independent of the basis set used in the underlying band calculation [15].

With the aim of visualizing and better interpret the field-induced formation of a new C-C bond in the system, also a Frontier Molecular Orbital (FMO) analysis has been carried out by employing the DFT framework exploited during the AIMD simulation that has led to the synthesis of (D)-erythrose from a glycolaldehyde aqueous mixture. By following the standard FMO theory [16], we have thus calculated the Highest Occupied Molecular Orbitals (HOMO) and the Lowest Unoccupied Molecular Orbitals (LUMO) of the two interacting molecules that for a field strength of 0.45 V/Å give rise to the onset of (D)-erythrose. In particular, the HOMO of the initial enol tautomer (then evolving into a glycolaldehyde enolate) species and the LUMO of the nearby glycolaldehyde molecule have been determined, as shown in Fig. S4. As expected, the most reactive HOMO of the 1,2-ethenediol species is delocalized all along the molecule whereas the first LUMO (*i.e.*, the most likely molecular orbital able to receive an electron from an interacting molecule) of the glycolaldehyde is delocalized between the oxygen of the carbonyl group and the α -Carbon (Fig. S4-a). Albeit this latter atom has, as a first extra-molecular neighbor, one of the oxygen atoms of the 1,2-ethenediol, the confined spatial extent of the LUMO located on this latter oxygen along with the out-of-phase relationship that characterizes the interaction between the closest HOMO and LUMO strongly hampers the formation of a new C-O bond between such molecules (Fig. S4-a). In fact, in a hundred of fs the system naturally explores a configuration having

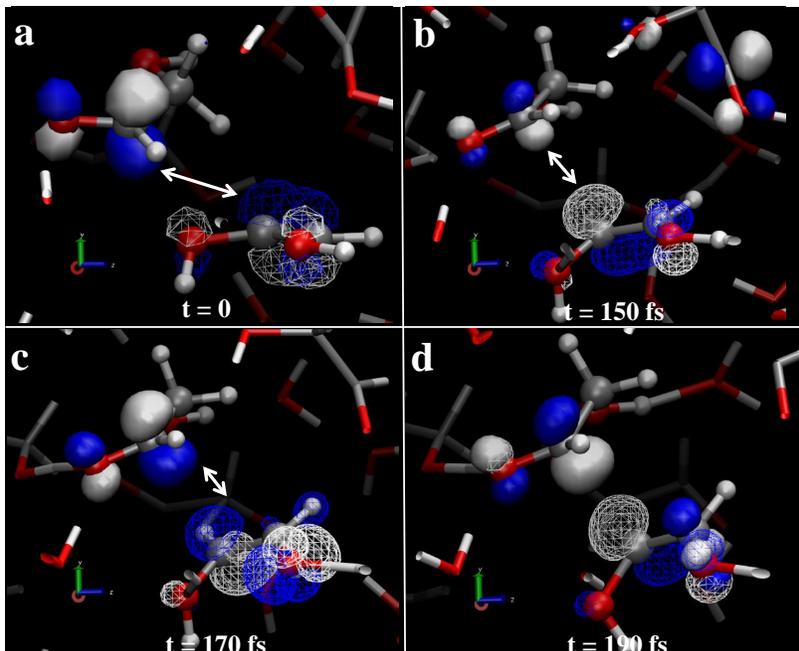


FIG. 4. HOMO of the 1,2-ethenediol (a,b) evolving into a glycolaldehyde enolate (c,d) and LUMO of the glycolaldehyde molecule that, by direct interaction, lead to the (D)-erythrose synthesis *via* an aldol condensation mechanism. Wireframe surfaces mark the first HOMO of 1,2-ethenediol (a,b) subsequently evolving into a glycolaldehyde enolate (c,d). Solid surfaces mark the first LUMO of the glycolaldehyde species. Frontier molecular orbitals have been plotted by choosing an opportune isovalue equal to ± 0.1 ; white orbitals indicate positive electronic phases whereas blue orbitals mark negative electronic phases.

the most expanded HOMO of the 1,2-ethenediol – located on the “future” α -Carbon of the enolate – and the biggest LUMO of the glycolaldehyde – located on its α -Carbon – as first neighbor molecular orbitals which are, additionally, in-phase (Fig. S4-b). The spatial nearness, the energetic nearness (they stem both from Carbon atoms), the symmetric character, and the “electronic resonance” stemming from the perpetuated in-phase relationship that lies between those FMO (Fig. S4-c-d) lead to the formation of the C-C bond marking the synthesis of (D)-erythrose.

IV. METADYNAMICS CALCULATIONS

To the aim of investigating reaction (1) of the main text, a series of Car-Parrinello simulations have been carried out in the liquid phase composed of 26 formaldehyde and 26 water (replicated in space by means of periodic boundary conditions and at 300 K) in conjunction with a recently developed path-Collective Variables (path-CV) metadynamics (MetD) approach [17], as implemented in the PLUMED-1.3 software [18] and available also in the

PLUMED-2.x versions [19]. This method, not requiring an educated guess about transition states and intermediates, represents a powerful tool to discover and sample possible chemical reaction pathways, taking into account a large number of degrees of freedom of the solution. In particular, we performed the MetD calculation with the aim of quantitatively determine the free-energy barrier heights characterizing the formation of glycolaldehyde from a formaldehyde aqueous solution. Because of the limited simulation time, the free-energy estimates related to the depths of the (meta)stable basins are not accurate. However, the approximated evaluation of the barriers separating those basins are, by construction, reliable.

During the MetD calculations the following parameters have been adopted for the coordination function (see Eq. 3 in Ref. [17]): $N = 6$, $M = 12$, $R_{SS'}^0 = 1.8 \text{ \AA}$ for $S, S' = \text{C}, \text{O}$, 1.5 \AA for $S = \text{C}, \text{O}$, $S' = \text{H}$, and 1.4 \AA for $S = S' = \text{H}$. We chose the parameter λ by following the empirical rule $\lambda D(R_k, R_{k,k'}) \sim 2.3$; this way, we set $\lambda = 0.58$. The MetD potential was composed of Gaussians with widths $\sigma_S = 0.02$ and $\sigma_Z = 0.1$ and height $3 \text{ kcal}\cdot\text{mol}^{-1}$ deposited every 35 fs. Finally, we restricted the maximum Z values explored by employing a semi-parabolic wall potential placed at $Z = 2.0$.

The contact matrix defining the reactants of the CV space have been initially evaluated by performing short equilibration dynamics within the reactants basin and selecting two nearby formaldehyde molecules. As far as the matrix defining the reaction product is concerned, a short simulation run of a single glycolaldehyde molecule in the gas-phase has been employed to define each matrix element. From these preliminary calculations, the matrices displayed in Table I and Table II – defining the CV space shown in Fig. 1 of the main text – have been carried out.

Atom	O	C	H
O ₁	0.223	0.980	0.569
O ₂	0.142	1.092	0.718
C ₁	1.143	0.062	1.871
C ₂	1.046	0.089	1.818

TABLE I. Contact matrix defining two formaldehyde molecules (in a formaldehyde aqueous solution) by means of the atomic coordination numbers of their oxygens (*i.e.*, O₁ and O₂) and of their carbons (*i.e.*, C₁ and C₂). This represents the reactants state of the MetD calculation shown in Fig. 1 of the main text.

Atom	O	C	H
O ₁	0.017	0.956	1.229
O ₂	0.017	1.054	0.194
C ₁	0.953	0.740	1.998
C ₂	1.058	0.740	1.085

TABLE II. Contact matrix defining a glycolaldehyde molecule (in a formaldehyde aqueous solution) by means of the atomic coordination numbers of its own oxygens (*i.e.*, O₁ and O₂) and of its own carbons (*i.e.*, C₁ and C₂). This represents the product state of the MetD calculation shown in Fig. 1 of the main text.

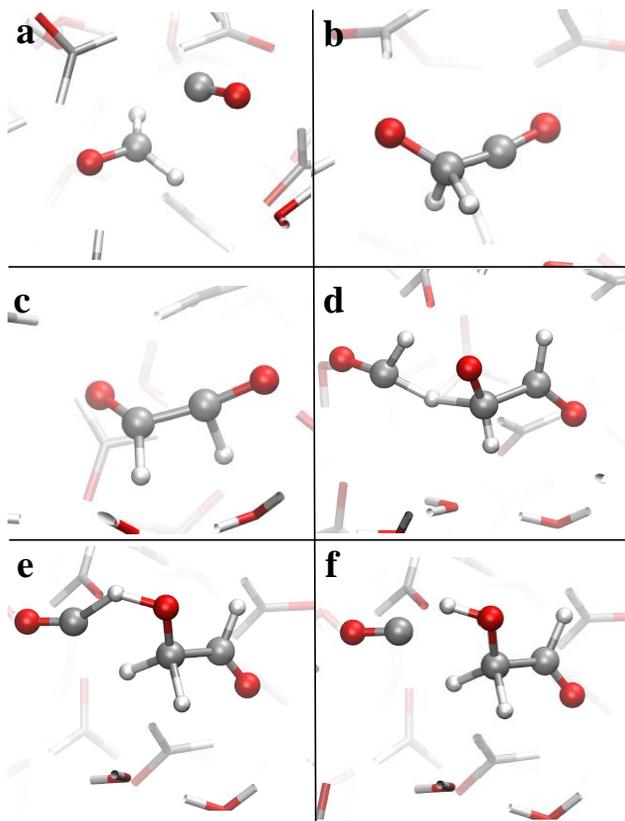


FIG. 5. Reaction mechanism leading to the biased synthesis of glycolaldehyde from a formaldehyde aqueous solution at room temperature revealed by state-of-the-art MetD. This configuration change corresponds to the reaction pathway A₂ shown in Fig. 1 of the main text. As it is straightforward to observe, it is definitely unlikely, as confirmed by the up-hill free-energy change associated with it (see Fig. 1 of the main text).

The very unfavorable free-energy changes associated with the chemical transformation investigated (*i.e.*, reaction (1) of the main text) is further highlighted by the mechanism revealed by our MetD simulation, as shown in Fig. S5.

-
- [1] Giannozzi, P. *et al.* QUANTUM ESPRESSO: a Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys.: Condens. Matter* **2009**, 21, 395502-395537.
- [2] Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, 55, 2471.
- [3] King-Smith, R. D.; Vanderbilt, D. Theory of Polarization of Crystalline Solids. *Phys. Rev. B.* **1993**, 47, 1651-1654.
- [4] Resta, R. Macroscopic Polarization in Crystalline Dielectrics: the Geometric Phase Approach. *Rev. Mod. Phys.* **1994**, 66, 899-915.
- [5] Berry, M. V. Quantal Phase Factors Accompanying Adiabatic Changes. *Proc. R. Soc. Lond. A* **1984**, 392, 45.
- [6] Umari, P.; Pasquarello, A. Ab initio Molecular Dynamics in a Finite Homogeneous Electric Field. *Phys. Rev. Lett.* **2002**, 89, 157602.
- [7] Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: Atomistic Simulations of Condensed Matter Systems. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2014**, 4, 15-25.
- [8] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, 77, 3865 and *Phys. Rev. Lett.* **1997**, 78, 1396.
- [9] Grimme, S.; Anthony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, 132, 154104.
- [10] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comp. Chem.* **2011**, 32, 1456.
- [11] Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, 126, 014101.
- [12] Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Sark. *Annual Rev. of Phys. Chem.* **2002**, 53, 291-318.
- [13] Marzari, N.; Vanderbilt, D. Maximally Localized Generalized Wannier Functions for Composite Energy Bands. *Phys. Rev. B* **1997**, 56, 12847-12865.

- [14] Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally Localized Wannier Functions: Theory and Applications. *Rev. Mod. Phys.* **2012**, 84, 1419-1475.
- [15] Mostofi, A. A.; Yates, J. R.; Lee, Y.-S.; Souza, I.; Vanderbilt, D.; Marzari, N. Wannier 90: a Tool for Obtaining Maximally-Localised Wannier Functions. *Comp. Phys. Comm.* **2008**, 178, 685-699.
- [16] Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **1952**, 20, 722.
- [17] 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. USA* **2015**, 112, 15030.
- [18] Bonomi, M. *et al.* PLUMED: a Portable Plugin for Free Energy Calculations with Molecular Dynamics. *Comp. Phys. Comm.* **2009**, 180, 1961-1972.
- [19] Tiribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED2: New Feathers for an Old Bird. *Comp. Phys. Comm.* **2014**, 185, 604.