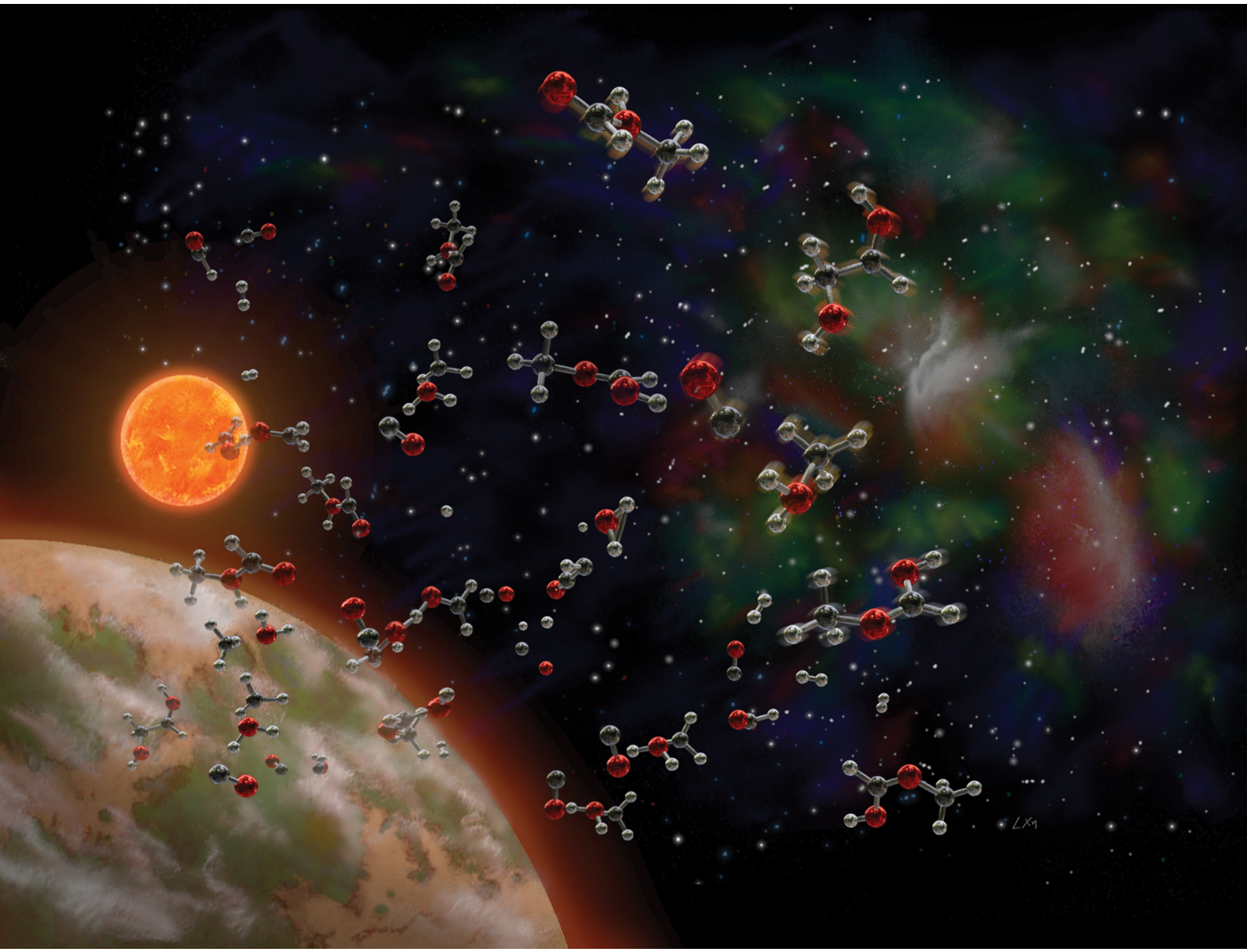
Volume 27
Number 32
28 August 2025
Pages 16695–17084

PCCP

Physical Chemistry Chemical Physics

rsc.li/pccp

25 YEARS ANNIVERSARY



ISSN 1463-9076

ROYAL SOCIETY OF CHEMISTRY

PAPER

Weiqi Wang, Jesús Pérez-Ríos *et al.*
Temperature-driven reaction pathways of gas-phase protonated glycolaldehyde formation and dissociation

PCCP Owner Societies

PCCP

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2025, **27**, 16744

Temperature-driven reaction pathways of gas-phase protonated glycolaldehyde formation and dissociation

Weiqi Wang, *[a] Xiangyue Liu [a] and Jesús Pérez-Ríos *[bc]

Received 8th May 2025,
Accepted 8th July 2025

DOI: 10.1039/d5cp01732g

rsc.li/pccp

Glycolaldehyde, a simple yet crucial organic compound, plays an important role in atmospheric chemistry and prebiotic studies. In this study, we examine the formation and thermal dissociation of protonated glycolaldehyde and its isomers. To achieve this, we develop comprehensive reaction networks using a novel approach based on *ab initio* molecular dynamics simulations, and analyze their behavior across thermal and hyperthermal temperature regimes. This approach offers valuable insights into the free energy landscape, reaction pathways, and temperature-dependent mechanisms of molecular formation. Our results demonstrate that the reaction network is highly temperature-dependent. Above 400 K, the transition to the product predominantly occurs through a direct pathway from reactant to product, primarily driven by transient high-temperature effects. This work highlights the potential of molecular dynamics simulations to enhance our understanding of atmospheric and interstellar chemistry, surpassing the limitations of conventional models.

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1 Introduction

Glycolaldehyde is one of the smallest organic compounds and plays a crucial role in atmospheric chemistry as an oxygenated volatile organic compound. It serves as a precursor to several key atmospheric species, including formaldehyde and formic acid.[1,2] On the other hand, as the simplest monosaccharide with two carbon atoms, glycolaldehyde is significant in the study of life's origins due to its close link to ribose formation. In the study of early Earth atmosphere, photochemical models have been used to explore the formation of glycolaldehyde under varying CO_2 and CH_4 concentrations and ratios.[3–5] Meanwhile, glycolaldehyde and other complex organic molecules closely associated with it, such as its isomers, methyl formate and acetic acid, as well as ethylene glycol, have been detected in interstellar space across regions with varying densities and temperatures.[6–23] The formation of glycolaldehyde in interstellar environments at cold temperatures has been extensively studied, with several potential synthesis pathways proposed, including gas-phase processes and reactions on ice mantles,[17,24,25] and involving protonated species that are further neutralized by electron attachment.[10,26,27]

In theoretical investigations, significant efforts have been devoted to identifying the key chemical species involved in the formation and dissociation of glycolaldehyde under various conditions. In general, approaches for constructing reaction networks and kinetic models are developed by considering stationary points (stable points and saddle points) on the potential energy surface that are either experimentally detected or theoretically predicted to play significant roles in the reaction mechanism. Typically, by establishing the reaction pathways of reactants, intermediates, transition states, and product configurations, the reaction rate constants were estimated using transition state theory, based on barrier energies and equilibrium assumption. However, this conventional approach relies on zero-Kelvin energies calculated at stationary geometries and the harmonic oscillator approximation, which neglects temperature effects relevant under realistic conditions. Furthermore, the accuracy of such models is highly dependent on the complete identification of all relevant intermediates and transition states, which is a challenging task, particularly for complex reaction systems at finite temperatures. Even if all configurations and correct barrier energies are identified, the reaction may still proceed through pathways that do not correspond to the minimum energy path,[28–30] a key assumption underlying transition state theory.

In this study, we explore the formation and thermal dissociation of protonated glycolaldehyde and its isomers, constituting their interconversion with formaldehyde and protonated formaldehyde. Specifically, we investigate the impact of

[a] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany. E-mail: wang@fhi-berlin.mpg.de*
[b] *Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA. E-mail: jesus.perezrios@stonybrook.edu*
[c] *Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY 11794-3800, USA*

 This journal is © the Owner Societies 2025

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

temperature on gas-phase transitions between these species by analyzing gas-phase collisions between formaldehyde and protonated formaldehyde, from which protonated glycolaldehyde may be formed, as suggested in recent experimental work.[31] To this end, we constructed comprehensive reaction networks and examined their behavior as a function of the temperature: thermal and hyper-thermal. In this work, we employ *ab initio* molecular dynamics (AIMD) simulations to explore the configurational space at various temperatures with enhanced sampling methods.[32] This approach allows an efficient sampling of the configuration space relevant to target temperatures, yielding the free energies and, therefore, the relative stability of various species at those temperatures. AIMD simulations also account for anharmonic effects, providing a more precise description of the formation and breaking of chemical bonds. The transition probabilities can be estimated using probabilistic models incorporating all possible intermediates and their transitions observed during the MD simulations. In this formalism, the reaction rate can be expressed as the probability of detecting the product over time. The reaction networks constructed at various temperatures enable us to examine how the reaction mechanism varies with temperature.

2 Method

We simulate the formation and thermal dissociation of glycolaldehyde across a broad temperature range to investigate the influence of temperature on the reaction mechanism. High-temperature simulations facilitate faster convergence in the sampling of configurational space within the framework of enhanced sampling methods. By integrating these simulations with various post-processing techniques, including clustering, dimensionality reduction, and reweighting methods, we analyze the free energy to assess the relative stability of key intermediates. The reaction mechanism is elucidated through reaction networks constructed using the transition probabilities between these intermediate species. Fig. 1 illustrates the workflow of the methodology employed in this work. The details of each of the steps are elaborated in this section.

2.1 Configurational space sampling

The configurational space is sampled under specific conditions to identify the most relevant configurations, including the intermediate species, their stability, and the evolution between different species and configurations. Here, we perform the sampling by AIMD, where the interaction between atoms is calculated using the MN15 density functional[33] with the def2-TZVPD basis set[34] implemented in the Gaussian 16 package.[35] The accuracy of the MN15 density functional has been examined by comparing the potential energy curves for the formation of the formaldehyde–protonated formaldehyde complex with those predicted by coupled-cluster singles, doubles, and perturbative triples [CCSD(T)] calculations. The MD simulations rely on the velocity Verlet algorithm[36] to integrate Newton's equation of motion with a time step Δt of 0.5 fs. The stochastic velocity rescaling thermostat[37] is implemented to control the temperature. In this approach, the system is coupled to a heat bath characterized by a thermostat parameter of 10 fs. To overcome the energy barriers in the reactions, replica-exchange molecular dynamics (REMD)[38,39] has been performed in the simulation to accelerate the convergence of the configurational space sampling. In the end, replicas at 11 temperatures (400 K, 514 K, 662 K, 853 K, 1098 K, 1414 K, 1820 K, 2343 K, 3017 K, 3883 K, and 5000 K) are used in the simulation with a total simulation time of 1.7 ns, with 20 trajectories run in parallel for each replica. The initial condition used in this study consists solely of H_2CO and H_2COH^+. While employing multiple initial conditions can indeed accelerate the convergence of REMD sampling, we chose not to do so explicitly. In each replica, the first 220 ps of sampling are excluded from the analysis, and after 11 ps of thermalization, each replica evolves into a distinct configuration. As a result, the production phase of the REMD simulation effectively begins from multiple initial conditions, even though only one was used initially.

To reproduce the constant volume at most of the experiments that are carried out, a repulsive spherical wall potential (power wall)[40] is applied in this system with 6 Å radius.

2.2 Structural representation

The Cartesian coordinates of the sampled configurations cannot be directly used for clustering because they are not invariant to translation and rotation. In this work, configurations are represented by an adjacency matrix to identify structures within the configurational space. The cutoff parameters are determined based on the typical chemical bond length between two elements. A few unclassified structures have been manually reviewed and integrated into the classified group. Once incorporated, they can be utilized for clustering as updated landmarks, as explained in the next section.

2.3 Clustering of the configurational space

During sampling, numerous structures are generated. While these structures are not identical, they share structural similarities. Therefore, it is beneficial to discretize the continuous configurational space based on structural similarities. This allows for the calculation of the relative stability of species within each group (''cluster'') of structures, providing more compact and insightful information about the reaction pathways.[41] Moreover, choosing suitable landmarks in the configurational space is essential for developing accurate reaction networks and studying reaction pathways. This choice affects the accuracy of the reaction network and guarantees reliable models of intermediate states and their transitions.

We have developed a ''local minima similarity'' (L-similarity) clustering algorithm. The algorithm emphasizes the transitions between stable and metastable species. The workflow of the L-similarity clustering algorithm is shown below.

1. Select a cutoff value to identify the high-dimensional metric between structures.

2. Add the local minimum points of the configurational space to the structure list and designate these local minimum points as the current ''landmarks''.

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

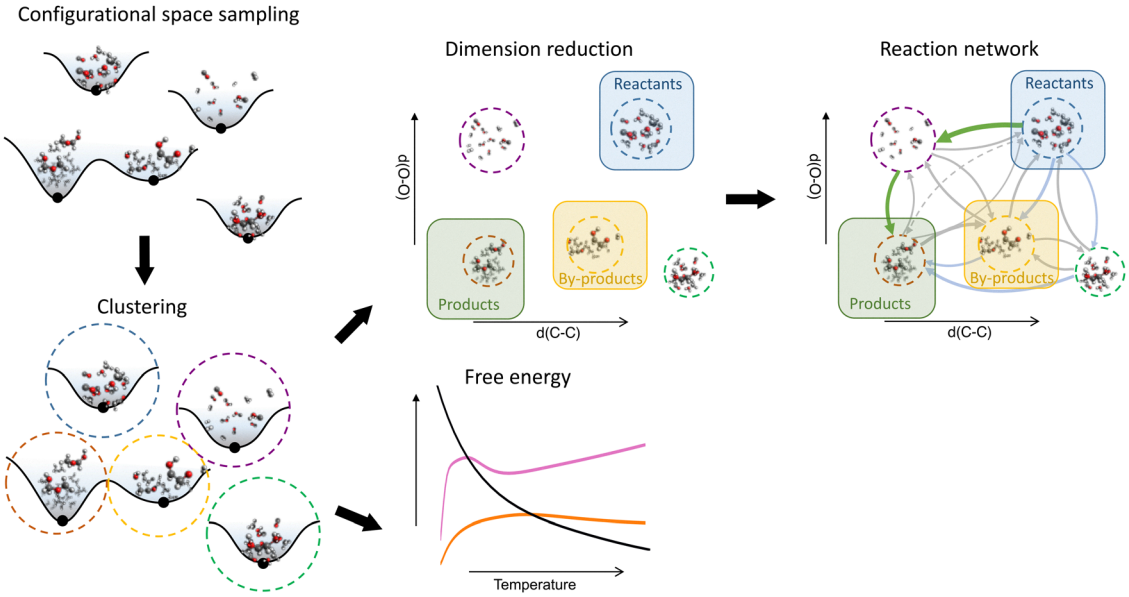


Fig. 1 Scheme of the methodology employed in this work. AIMD samples the configurational space combined with REMD. Then the sampled configurations are clustered. With the MBAR approach and MSM, the free energy curves are then estimated, and the detailed reaction network about the transitions between different clusters of configurations is constructed.

3. Loop through the structure list and calculate the high-dimensional distance between each structure and the current landmarks. Remove the structure from the structure list if the high-dimensional distance is shorter than the cutoff.

4. Choose the next structure in the structure list as the current landmark and repeat step 3.

5. Relax the structures in the structure list and repeat steps 3 and 4. The optimized structure list will become the updated landmark list.

6. Cluster the original structures to their nearest landmark by calculating their high-dimensional metric.

When using only the first four steps, the process is equivalent to the regular space clustering algorithm,[42] which uses high-dimensional distance as the metric.

2.4 Dimension reduction

The configurational space is reduced to a two-dimensional space defined by the distance between two oxygen atoms (R_{O-O}) and the distance between two carbon atoms (R_{C-C}) are effective indicators of the reaction process. Initially, the formaldehyde molecule (H_2CO) and the protonated formaldehyde molecule (H_2COH^+) are separated, resulting in large O–O and C–C distances. For the intermediates, a proton-bond dimer has a short O–O distance while the C–C distance is relatively long. In contrast, protonated glycolaldehyde ($HOCH_2CHOH^+$) has a relatively long O–O distance and a short C–C distance, while protonated methyl formate ($HC(OH)OCH_3^+$) exhibits intermediate lengths for both O–O and C–C distances.

2.5 Free energy and reaction network

The free energies of each cluster of configurations are calculated using the multistate Bennett acceptance ratio (MBAR)[43] approach, to accurately estimate free energy differences from the replica exchange molecular dynamics (REMD) simulations. In REMD, temperature serves as an implicit biasing potential that enhances conformational sampling by allowing replicas to explore higher-energy states more efficiently. However, because the ensemble of configurations collected at elevated temperatures does not directly reflect the Boltzmann distribution of the target temperature, a reweighting procedure is essential to recover unbiased thermodynamic properties. MBAR provides a statistically rigorous framework for this reweighting by combining information from all temperature replicas. It accounts for the overlap in configuration space across temperatures and yields the most probable set of free energies consistent with the sampled data. Unlike traditional histogram-based methods, MBAR does not require binning, making it well-suited for high-dimensional systems and nonequilibrium sampling scenarios where configurations may be unequally distributed. By leveraging the full potential of the REMD data, MBAR enables accurate reconstruction of free energy landscapes under the desired thermodynamic conditions.

In a convergent sampling, the ergodic hypothesis can be applied to the evolution of the system. Therefore, the reaction network is considered an undirected graph and can be constructed using the Markov state model (MSM),[44] which effectively serve as a form of sub-sampling of the simulation trajectories. In this work, we employ the PyEMMA package[45] to estimate the MSM. At the core of the MSM approach are the transition matrices. In this study, system configurations are sampled every 10 steps (*i.e.*, τ = 10 in PyEMMA) to compute transition matrices between clusters of similar configurations. From these matrices, the stationary distribution is obtained by calculating the first right eigenvector. Combined with the transition probabilities, this information enables the construction of reaction networks.

 This journal is © the Owner Societies 2025

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

The reaction rates and pathways of specified reactants and products are studied by transition path theory (TPT),[46,47] based on the transition probabilities derived from the transition matrices. First, the reactant and product states are defined. Then, Monte Carlo integration is performed using the TPT equations and the transition matrices to obtain the reaction rates.

3 Results and discussion

We aim to understand the production and thermal dissociation of protonated glycolaldehyde at various temperatures from three perspectives: static stability, dynamic stability, and connectivity within the reaction pathways. First, we investigate the species observed at each temperature. Then, we analyse the temperature dependence of the static stability of these species by evaluating their free energies. Next, the transitions between these species will be used to construct reaction networks, providing insights into their dynamic stabilities. Following this framework, we explore the reaction pathways and calculate reaction rate constants from the reactants to the products across different temperatures, highlighting the critical intermediate species involved.

3.1 Exploring relevant species

In principle, every species has a probability of being observed at all temperatures, regardless of how small that probability may be. However, with a finite number of observations, species with extremely low probabilities are unlikely to be observed in practice. As the stability of each species changes with temperature, the set of species observed varies accordingly. Table 1 lists the 20 species observed during the simulation, obtained by coarse-graining the sampled configurations into groups based on their chemical bond connections. These species include $CH_4 + CO_2H^+$, $CO + CH_3OH_2^+$, $H_2CO + H_2COH^+$, $H_2O + CH_3CO^+$, $HCOOHCH_3^+$, $HOCH_2CHOH^+$, $HOCHOCH_3^+$, "dissociated", "proton bond dimer", $CO + H_2 + H_2COH^+$, $H_2O + CO + CH_3^+$, $CO + H_2 + H_2 + COH^+$, $H_2 + CO_2 + CH_3^+$, $H_2 + HCOCHOH^+$, $H_2 + HCOOCH_2^+$, $HCOH + H_2 + COH^+$, $HCOOH + CH_3^+$, and $HCOH + H_2COH^+$. $H_2O + H_2 + OCCH^+$, and $H_2 + H_2 + HCOCO^+$. As noted above, these species are not always observed at every sampled temperature due to the finite number of observations at each temperature. Nevertheless, the observation of a species indicates its relatively high stability under the corresponding conditions.

3.2 Temperature effects on free energy

The free energy for each of the identified structures (see Table 1) as a function of the temperature is shown in Fig. 2. Among all the species, nine are observed (indices 0–7 and 9) at the lowest simulated temperature of 400 K, represented by solid curves. In contrast, species relevant at higher temperatures are depicted as dashed curves. The temperature dependence of the free energies for the nine species is significantly more complex than that of the species relevant at high temperatures. Overall, the reactant $H_2CO + H_2COH^+$ remains one of the most stable species.

Table 1 Index of clusters with corresponding configurations

Index	Name/chemical formula	Label
0	$CH_4 + CO_2H^+$	
1	$CO + CH_3OH_2^+$	
2	$H_2CO + H_2COH^+$	
3	$H_2O + CH_3CO^+$	
4	$HCOOHCH_3^+$	
5	$HOCH_2CHOH^+$	
6	$HOCHOCH_3^+$	
7	dissociated	
8	proton bond dimer	
9	$CO + H_2 + H_2COH^+$	
10	$H_2O + CO + CH_3^+$	
11	$CO + H_2 + H_2 + COH^+$	
12	$H_2 + CO_2 + CH_3^+$	
13	$H_2 + HCOCHOH^+$	
14	$H_2 + HCOOCH_2^+$	
15	$HCOH + H_2 + COH^+$	
16	$HCOOH + CH_3^+$	
17	$HCOH + H_2COH^+$	
18	$H_2O + H_2 + OCCH^+$	
19	$H_2 + H_2 + HCOCO^+$	

However, two species are slightly more stable under specific conditions: $CO + CH_3OH_2^+$ at 400 K, and the "proton-bond dimer" from 500 K to 800 K. It is worth noting that these species are formed by the combination of two small molecules, indicating that entropy plays a significant role at temperatures above 400 K. The free energy of $CO + CH_3OH_2^+$ increases with rising temperature up to 1600 K, after which it gradually decreases. In contrast,

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

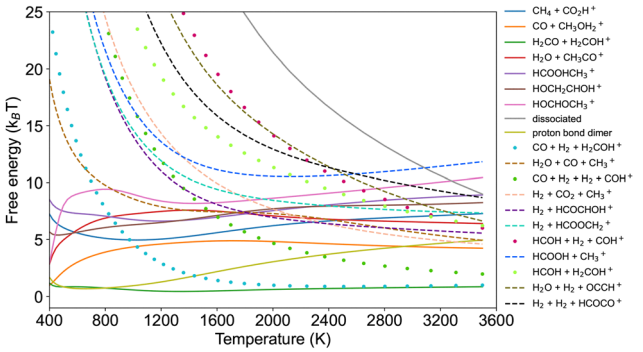


Fig. 2 Free energies of configurations changing with the temperature. The solid lines indicate the free energy of the 9 species observed at 400 K.

the "proton-bond dimer" exhibits an opposite trend: its free energy decreases until 600 K, and then rises with increasing temperature. The free energy of the product, protonated glycolaldehyde ($HOCH_2CHOH^+$), consistently increases with temperature.

It is worth pointing out that the two variants of protonated methyl formate, $HCOOHCH_3^+$ and $HOCHOCH_3^+$, display different temperature-dependent behaviors. The free energy of $HCOOHCH_3^+$ decreases until 1300 K and then begins to increase, while $HOCHOCH_3^+$ increases rapidly until 800 K, decreases until 1400 K, and then rises again at higher temperatures. The free energies of the two dissociated configurations of formaldehyde, $CO + H_2 + H_2COH^+$, and $CO + H_2 + H_2 + COH^+$, decrease sharply with rising temperature. $CO + H_2 + H_2COH^+$, in particular, becomes the second most stable species at temperatures above 1500 K. This suggests that $CO + H_2$ may actively participate in the reactions at higher temperatures.

This temperature-dependent behavior of free energy determines the relative stability of the species. As the probability of observing each species varies with temperature, the reaction mechanisms are expected to differ between low and high temperatures. This results in distinct reaction pathways in the production and dissociation of protonated glycolaldehyde, as discussed below.

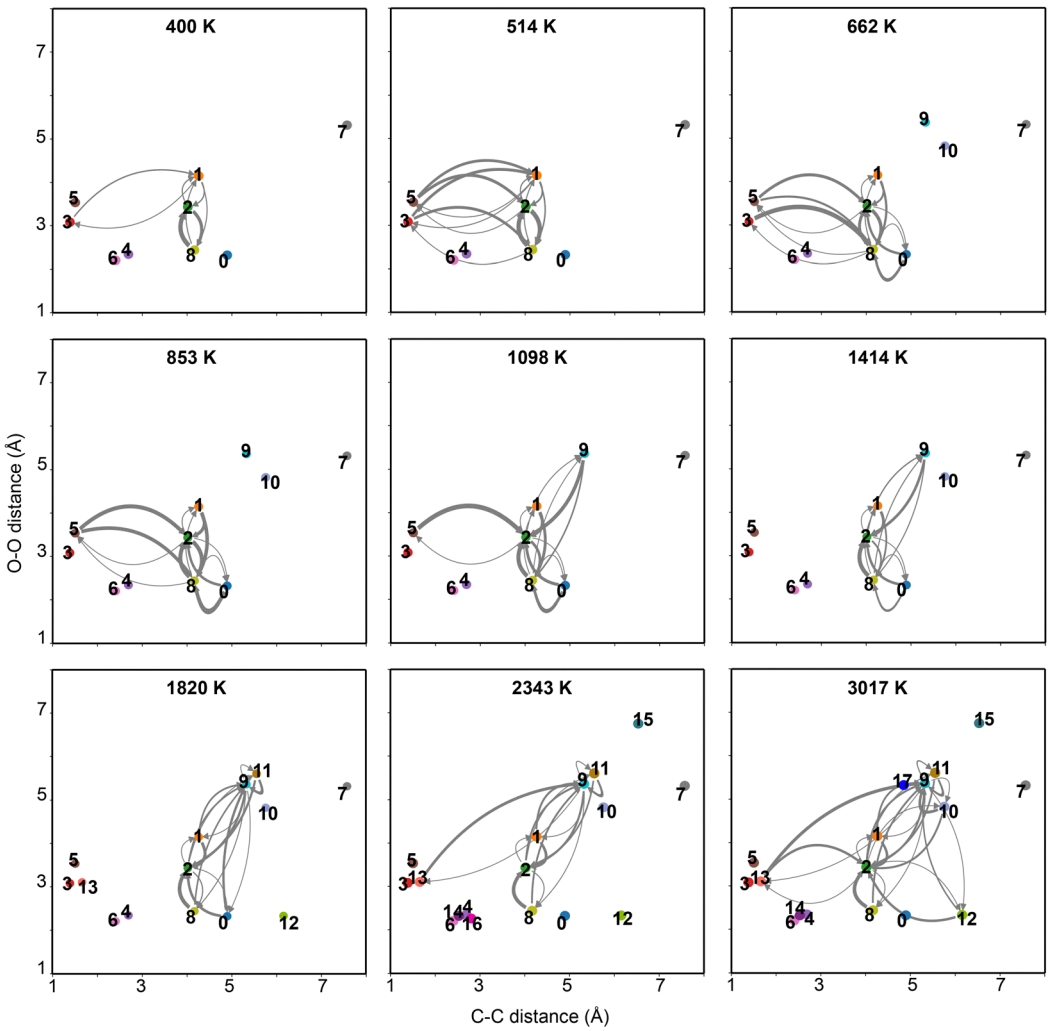


Fig. 3 Reaction networks at various temperatures. The line widths represent the transition probabilities of the respective transitions. Only transitions with probabilities exceeding 0.05% are included in the figure. The names or chemical formulas of the species are listed in Table 1.

 This journal is © the Owner Societies 2025

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

3.3 Temperature-dependent reaction networks

The free energy of all species provides insight into their positions on the free energy surface, but offers no information about the the structural or morphological features of the surrounding configurations. In other words, the free energy of a certain species alone does not reveal whether the free energy surface around this species is smooth or rough, or whether the slope is steep or not. This surrounding morphological information is crucial because it governs the dynamic transformations of the species. To obtain such information, we can calculate the transition probabilities between species. These probabilities capture the likelihood of transition from one species to another and reflect the local environment of the free energy surface. Collectively, these transition probabilities form a reaction network, which provides a comprehensive view of the dynamics governing the system. The transition probabilities in reaction networks provide insights into the activity of species and highlight the emergence of important intermediate species at different temperatures.

The constructed reaction networks at various temperatures are shown in Fig. 3. Across all temperatures, species with C–C distances ranging from 4 to 4.5 Å significantly contribute to the reactions. These species include $CO + CH_3OH_2^+$, $H_2CO + H_2COH^+$, and the "proton-bond dimer". A common feature of these three species is that they do not form C–C or O–O bonds but consistently maintain two C–O bonds. This indicates that transitions between them primarily involve the transfer of H atoms or a proton. The high transition probability between $H_2CO + H_2COH^+$ and the proton-bond dimer involves only forming and breaking a hydrogen bond, without changing any chemical bonds. In contrast, the transition mechanism between $H_2CO + H_2COH^+$ and $CO + CH_3OH_2^+$ is more complex. At low temperatures, this transition may result from the recombination of $CO + H_2 + H_2COH^+$, since recombination reactions are more efficient at low collision energies.[48] $CO + H_2 + H_2COH^+$ is observed exclusively at temperatures above 662 K, where it actively participates in transitions to other species. Interestingly, transitions involving protonated glycolaldehyde and other species vary significantly with temperature. For instance, protonated glycolaldehyde is one of the reaction products at temperatures between 514 and 1098 K, as a consequence of the endothermic nature of the reaction.

3.4 The reactions of protonated glycolaldehyde

To further investigate the reactions involving protonated glycolaldehyde at finite temperatures, we calculate the reaction rate

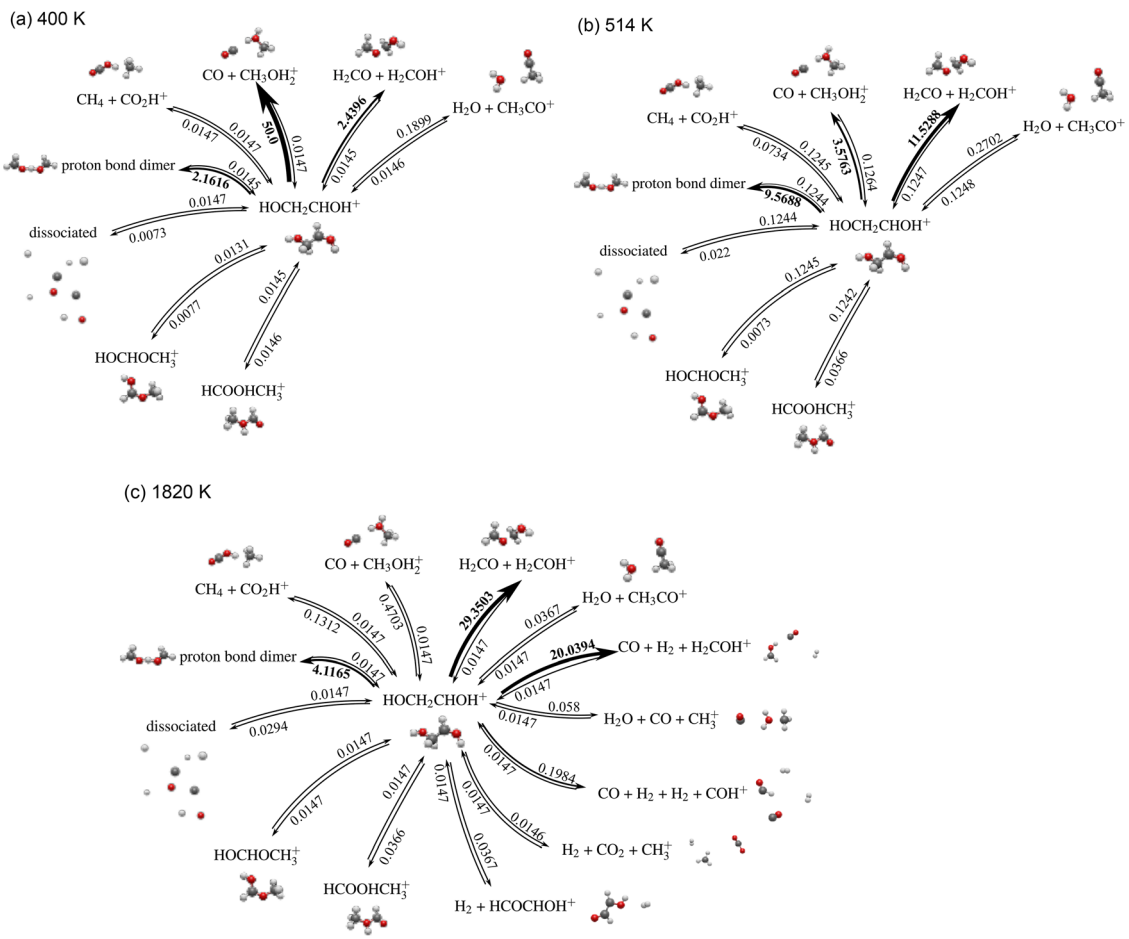


Fig. 4 Reaction rate constants of the reactions involving the protonated glycolaldehyde with other species at (a) 400 K, (b) 514 K, and (c) 1820 K. The unit of the reaction rate constants is ps^{-1}.

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

constants between protonated glycolaldehyde and other species, as shown in Fig. 4. Three representative temperatures are selected for this analysis: 400 K, which approximates the conditions under which glycolaldehyde has been observed in interstellar space; 514 K, corresponding to the biomass ignition temperature; and 1820 K, which approximates the temperature of biomass flames.

At 400 K, the rate constants for the synthesis reactions producing protonated glycolaldehyde are nearly identical, ranging narrowly from 0.0145 to 0.0147 ps^{-1}, with the exception of the reaction involving an isomer of protonated methyl formate ($HOCHOCH_3^+$), which exhibits a slightly lower rate constant of 0.0131 ps^{-1}. This uniformity suggests that these synthesis reactions likely share a common rate-determining step, as will be discussed in the next section. Among the reverse reactions, those involving the conversion of protonated glycolaldehyde – the reaction leading to $CO + CH_3OH_2^+$ displays a significantly higher rate constant compared to other decomposition pathways. This indicates that the interaction between $CO + CH_3OH_2^+$ and protonated glycolaldehyde may represent an elementary reaction step at this temperature, highlighting the critical role of this product pair.

At 514 K, the behavior of the synthesis reactions remains similar to that observed at 400 K; however, the rate constants increase by approximately a factor of nine, again with the exception of the reaction involving $CO + CH_3OH_2^+$. This further supports the hypothesis that the synthesis reactions share a common rate-determining step at elevated temperatures. For the decomposition of protonated glycolaldehyde at 514 K, the fastest pathway leads to the formation of $H_2CO + H_2COH^+$, with a rate constant of 11.528 ps^{-1}. The second fastest pathway forms a proton-bound dimer (a dimer complex of H_2CO and H_2COH^+), with a rate constant of 9.5688 ps^{-1}. The modest difference between these two rates indicates frequent interconversion between the monomer pair and the dimer, which is consistent with the behavior observed in Fig. 3. Among other products, two species – $CO + CH_3OH_2^+$ and $H_2O + OCCH_3^+$ – exhibit rate constants for their formation from protonated glycolaldehyde that are higher than those of the synthesis reactions. Since the equilibrium constant is defined as the ratio of forward to reverse rate constants, this implies that both $CO + CH_3OH_2^+$ and $H_2O + OCCH_3^+$ are more thermodynamically stable than protonated glycolaldehyde under these conditions.

At 1820 K, all reactions leading to the formation of protonated glycolaldehyde proceed with identical rate constants, including those involving species prevalent at high temperatures. In terms of decomposition, three pathways dominate, producing the proton-bound dimer, $H_2CO + H_2COH^+$, and $CO + H_2 + H_2COH^+$. The latter product is newly observed at high temperatures and is known to exhibit high stability, as shown in Fig. 2. It results from the dissociation of H_2CO in the $H_2CO + H_2COH^+$ pair into $CO + H_2$, a process favored at elevated temperatures. Given the comparable stability and the frequent transitions between $H_2CO + H_2COH^+$ and $CO + H_2 + H_2COH^+$ at high temperatures, we conclude that the decomposition of protonated glycolaldehyde primarily results in formaldehyde and its derivatives.

3.5 Reaction pathways

The apparent reaction rate constants for the overall transformation between $H_2CO + H_2COH^+$ and $HOCH_2CHOH^+$ (protonated glycolaldehyde) have been calculated across a range of temperatures, as summarized in Table 2. The table also includes the most relevant reaction pathways connecting the reactants and the product. From 400 K (0.015 ps^{-1}) to 662 K (0.206 ps^{-1}), the forward rate constant increases rapidly with temperature, reaching a maximum at 662 K. Beyond this point, the rate constant decreases sharply, reaching 0.015 ps^{-1} at 1820 K, before rising slightly again from 2343 K (0.015 ps^{-1}) to 3017 K (0.029 ps^{-1}). This trend suggests that the optimal temperature for the formation of protonated glycolaldehyde is around 662 K.

In contrast, the backward reaction rate constant increases rapidly from 400 K (2.440 ps^{-1}) to 1098 K (38.636 ps^{-1}), after which it decreases slightly and begins to converge at higher temperatures. Regarding the contributions of different pathways at varying temperatures, since the system is at thermodynamic equilibrium, the pathways contribute the same to both the forward and backward reactions. This holds true even though the net reaction rates differ significantly between the two directions.

To understand the temperature dependence of the reaction rate, it is necessary to analyze the reaction pathways at each temperature. When considering the formation of the protonated glycolaldehyde, at 400 K, the formation of the product requires the reactant $H_2CO + H_2COH^+$ to first transition to $CO + CH_3OH_2^+$, either directly or through intermediate transitions involving the proton-bond dimer and $H_2O + OCCH_3^+$. For temperatures ranging from 500 K to 2343 K, direct transitions from reactants to products become highly probable and play the most significant role in driving the reaction. This is likely due to the formation of C–C bonds during molecular collisions, which become more probable at higher temperatures due to the increased kinetic energy of the molecules. As a result, the likelihood of direct C–C bond formation is significantly enhanced. For reactions occurring below 900 K, $CO + CH_3OH_2^+$ and the proton-bond dimer typically serve as key intermediates. For temperatures exceeding 900 K, additional reaction pathways to other possible products open up, leading to a reduction of the reaction rate toward protonated glycolaldehyde formation. Notably, the dissociated state, $CO + H_2 + H_2COH^+$, becomes increasingly significant as the temperature rises, underscoring its role in high-temperature reaction kinetics.

Although many intermediates have been observed in the formation of the protonated glycolaldehyde, their roles in the reaction remain to be fully understood. At low temperatures in interstellar space, ref. 10 suggested that the proton-bond dimer serves as the transition state from reactants ($H_2CO + H_2COH^+$) to products ($HOCH_2CHOH^+$). This elementary step is proposed to be the rate-determining step of the overall reaction. On the other hand, in ref. 49, when the hydrated proton is included in the reaction mechanism, it is proposed that the proton-bond

 This journal is © the Owner Societies 2025

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Reaction pathways and corresponding contribution to the reaction of $H_2CO + H_2COH^+ \rightleftharpoons HOCH_2CHOH^+$ (protonated glycolaldehyde). Only the reaction pathways with contributions larger than 1% are listed

Temperature	Reaction rate constants: forward/backward (ps^{-1} (10^{-11} cm^3 s^{-1}))	Reaction pathway	Contribution (%)
400 K	0.0145 (0.418)	2 ↔ 1 ↔ 5	63.85
	2.440 (70.259)	2 ↔ 8 ↔ 1 ↔ 5	31.89
		2 ↔ 8 ↔ 3 ↔ 1 ↔ 5	2.78
		2 ↔ 3 ↔ 1 ↔ 5	1.24
514 K	0.125 (3.59)	2 ↔ 5	47.62
	11.529 (332.03)	2 ↔ 1 ↔ 5	22.18
		2 ↔ 8 ↔ 1 ↔ 5	16.77
		2 ↔ 8 ↔ 5	11.86
		2 ↔ 8 ↔ 3 ↔ 1 ↔ 5	1.21
662 K	0.206 (5.93)	2 ↔ 5	50.34
	23.380 (673.33)	2 ↔ 8 ↔ 5	30.37
		2 ↔ 8 ↔ 1 ↔ 5	9.65
		2 ↔ 1 ↔ 5	9.4
853 K	0.177 (5.09)	2 ↔ 5	52.22
	29.280 (843.28)	2 ↔ 8 ↔ 5	45.58
		2 ↔ 1 ↔ 5	1.85
1098 K	0.140 (4.02)	2 ↔ 5	60.63
	38.636 (1112.7)	2 ↔ 8 ↔ 5	34.15
		2 ↔ 9 ↔ 5	4.47
1414 K	0.0735 (2.12)	2 ↔ 5	60.06
	28.434 (818.90)	2 ↔ 8 ↔ 5	19.99
		2 ↔ 9 ↔ 5	18.48
1820 K	0.0147 (0.423)	2 ↔ 5	50.01
	29.350 (845.29)	2 ↔ 9 ↔ 5	45.28
		2 ↔ 8 ↔ 9 ↔ 5	3.61
2343 K	0.0147 (0.423)	2 ↔ 5	50.01
	22.519 (648.56)	2 ↔ 9 ↔ 5	46.06
		2 ↔ 8 ↔ 9 ↔ 5	1.64
		2 ↔ 11 ↔ 9 ↔ 5	1.28
3017 K	0.0294 (0.845)	2 ↔ 9 ↔ 5	46.65
	29.381 (846.17)	2 ↔ 5	37.54
		2 ↔ 8 ↔ 5	9.31
		2 ↔ 11 ↔ 9 ↔ 5	3.35

dimer functions not as a transition state but as a precursor to HCOH + H_2COH^+, which constitutes the transition state of the rate-determining step leading to the formation of $HOCH_2CHOH^+$. Additionally, it has been suggested that hydroxymethylene is a critical active species in the reaction pathway from neutral H_2CO to $HOCH_2CHO$.[50] Once H_2CO is converted to the isomer HCOH, the reaction is believed to occur rapidly.

Here, we aim to investigate whether these intermediate species play a significant role within the temperature range considered in this study. Within the current framework, it is possible to analyse in depth the relevance of intermediate species and identify the critical rate-determining steps toward the production of protonated glycolaldehyde. However, due to the stochastic nature of the Markov state model, the high participation of key intermediates in the reaction cannot be directly correlated with a significant influence on the reaction rate. To evaluate the role of the mentioned intermediates in determining the reaction rate, three specific reaction rate constants are considered for each intermediate:

(1) The reaction rate constant from the intermediate to the product (r_{IP}).

(2) The forward reaction rate constant between the intermediate and the reactant (r_{IR}).

(3) The backward reaction rate constant from the intermediate to the reactant (r_{RI}).

The rate-determining step should occur either before or after the intermediate. If r_{IP} is closer to the total reaction rate constant (r_{total}), the rate-determining step likely occurs after the intermediate transitions to the product. Conversely, if r_{RI} is closer to r_{total}, the rate-determining step likely precedes the transition involving the intermediate. Apart from these two reactions, r_{IR} is also considered to clarify the intermediate's role. Using these three reaction rate constants (r_{RI}, r_{IR}, and r_{IP}), the overall reaction can be reconstructed as a model featuring a central hub intermediate. This approach helps elucidate the intermediates' role and their impact on the reaction dynamics.

As a result, in Fig. 5(c), we calculate the three reaction rate constants (r_{IP}, r_{IR}, and r_{RI}) for all identified intermediates in the literature: proton-bond dimer[10] and CO + $CH_3OH_2^+$,[49] along with an additional intermediate containing an HCOH molecule. The results reveal that, in all cases, r_{IP} is equal to r_{total}, while r_{IR} and r_{RI} are significantly higher than r_{IP}. Specifically, for the proton-bond dimer, its r_{RI} and r_{IR} are several orders of magnitude higher than r_{IP}. This indicates that the transitions between the reactant and the proton-bond dimer establish rapid equilibrium within the reaction. Consequently, the proton-bond dimer does not influence the overall reaction rate constant. Similarly, dissociated formaldehyde (CO + H_2) exhibits behavior analogous to that of the proton-bond dimer. For the species CO + $CH_3OH_2^+$, while r_{IR} is an order of magnitude greater than r_{RI}, both are still significantly higher than r_{IP}. This results in a concentration of CO + $CH_3OH_2^+$ that is much lower than that of the reactant, rendering its impact on the reaction rate constant negligible.

In the case of H_2O + CH_3CO^+ and HCOH, their r_{IR} values are several orders of magnitude higher than r_{RI}. As a result, the

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

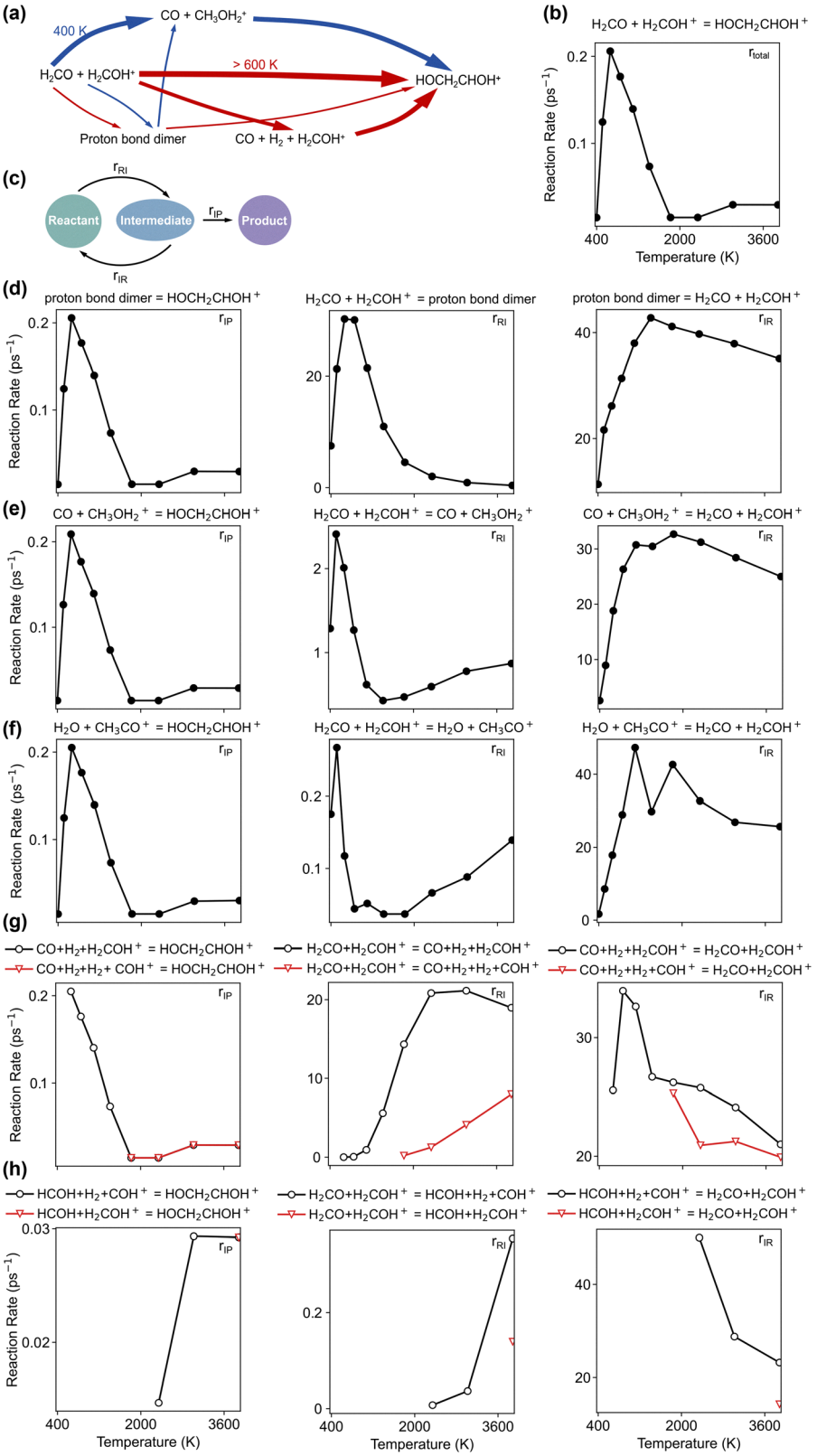


Fig. 5 Reaction rate constants of the reactions involving the important intermediates. (a) Summary of the reaction pathways at different temperatures. (b) The reaction rate constant of the total reaction as a function of temperature. (c) The reaction rate constants of the reactions involving the important intermediates. (d)–(h) The calculated reaction rate constants r_{IP}, r_{RI} and r_{IR} as a function of temperature for each intermediate.

This journal is © the Owner Societies 2025

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

concentrations of these two species are exceedingly low, and their contribution to the reaction is insignificant when compared to other species with higher static and kinetic stabilities. As a result, none of the transitions involving the mentioned intermediates serve as the rate-determining step at high temperatures. Instead, the bottleneck of the reaction lies in the final step: the transition to the product itself.

4 Conclusions

In this study, we investigate model gas-phase reactions relevant to atmospheric and interstellar conditions, focusing on the interconversion between protonated glycolaldehyde (and its isomers) and formaldehyde/protonated formaldehyde through their formation and thermal dissociation pathways. To understand the reaction comprehensively, we analyze the reaction pathways and the stability of the species involved across a range of finite temperatures. Unlike the reaction mechanism predicted by traditional models, our findings reveal that the reaction mechanism is highly temperature-dependent: the important intermediates and reaction pathways vary significantly at different temperatures. At finite temperatures, the participation of intermediates in the reaction is dependent on their free energies, which in turn causes the overall reaction rate constant to depend strongly and non-monotonically on temperature. At intermediate temperatures (514–2343 K), the direct reaction from formaldehyde to glycolaldehyde accounts for approximately half of the total reaction flux. At higher temperatures, the opening of additional reaction channels reduces the relative contribution of this direct pathway. In contrast, at lower temperatures (*e.g.*, 400 K), the contribution from the direct pathway is less than 1%. This indicates that the activation energy for C–C bond formation and breaking is relatively high, and at lower temperatures, alternative pathways involving intermediate species are required to bypass the energy barrier. This also explains the significantly lower reaction rate observed at 400 K compared to higher temperatures.

Our approach relies on the stability of the species, which is determined not by the zero-Kelvin potential energies but by the free energies obtained from MD simulations at finite temperatures, which naturally incorporates temperature effects and anharmonicity. Rather than relying on transition state theory applied to static models based on zero-Kelvin energies, the reaction rates are directly obtained from transition probabilities through the reaction pathways explored in the MD simulations at finite temperatures. Therefore, this approach provides a more realistic representation of the reaction mechanisms with better predictions of reaction rate constants and relative abundance and helps identify key intermediates that might otherwise be overlooked in conventional models. These results underscore the importance of incorporating finite temperature effects into kinetic analyses to achieve a comprehensive understanding of the reactions involved in the formation and dissociation of protonated glycolaldehyde.

It is worth noting that the current probabilistic model, specifically the Markov state model, does not capture the detailed dynamical processes underlying the reaction. To gain a deeper understanding of the mechanisms of chemical bond formation and breaking, future research will focus on the scattering processes that lead to the formation of $HOCH_2$-$CHOH^+$. In addition, the findings of this study could offer valuable insights into the formation of complex organic molecules in interstellar environments and the origins of prebiotic chemistry. To advance this line of investigation, it will be important to extend the current AIMD approach to account for nuclear quantum effects, which can be significant under the low-temperature conditions typical of the interstellar medium.

Author contributions

W. W.: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft. X. L.: formal analysis, investigation, validation, visualization, writing – original draft. J. P.: conceptualization, funding acquisition, project administration, resources, supervision, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data of relevance to this study may be accessed at the following DOI: **https://doi.org/10.5281/zenodo.15363757**.

Acknowledgements

W. W. acknowledges funding from the Max Planck-Radboud University Center for Infrared Free Electron Laser Spectroscopy. J. P.-R. acknowledges the support of the Simons Foundation and the hospitality of the Radboud University and Fritz-Haber Institute of the Max Planck Society. The authors thank Sandra Brünken and Hunarpreet Kaur for fruitful discussions. Open Access funding provided by the Max Planck Society.

Notes and references

1 T. J. Johnson, R. L. Sams, L. T. Profeta, S. K. Akagi, I. R. Burling, R. J. Yokelson and S. D. Williams, *J. Phys. Chem. A*, 2013, **117**, 4096–4107.
2 N. I. Butkovskaya, N. Pouvesle, A. Kukui and G. Le Bras, *J. Phys. Chem. A*, 2006, **110**, 13492–13499.
3 C. E. Harman, J. F. Kasting and E. T. Wolf, *Origins Life Evol. Biospheres*, 2013, **43**, 77–98.
4 S. D. Domagal-Goldman, V. S. Meadows, M. W. Claire and J. F. Kasting, *Astrobiology*, 2011, **11**, 419–441.
5 L. Allou, L. El Maimouni and S. Le Calvé, *Atmos. Environ.*, 2011, **45**, 2991–2998.

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

6 J. K. Jørgensen, C. Favre, S. E. Bisschop, T. L. Bourke, E. F. Van Dishoeck and M. Schmalzl, *Astrophys. J., Lett.*, 2012, **757**, L4.
7 J. K. Jørgensen, M. H. D. van der Wiel, A. Coutens, J. M. Lykke, H. S. P. Müller, E. F. van Dishoeck, H. Calcutt, P. Bjerkeli, T. L. Bourke, M. N. Drozdovskaya, C. Favre, E. C. Fayolle, R. T. Garrod, S. K. Jacobsen, K. I. Öberg, M. V. Persson and S. F. Wampfler, *Astron. Astrophys.*, 2016, **595**, A117.
8 J. M. Hollis, F. J. Lovas and P. R. Jewell, *Astrophys. J.*, 2000, **540**, L107.
9 J. M. Hollis, P. R. Jewell, F. J. Lovas and A. Remijan, *Astrophys. J.*, 2004, **613**, L45.
10 D. Halfen, A. Apponi, N. Woolf, R. Polt and L. Ziurys, *Astrophys. J.*, 2006, **639**, 237.
11 M. L. Vant Hoff, D. Harsono, J. J. Tobin, A. D. Bosman, E. F. van Dishoeck, J. K. Jørgensen, A. Miotello, N. M. Murillo and C. Walsh, *Astrophys. J.*, 2020, **901**, 166.
12 K. Altwegg, H. Balsiger and S. A. Fuselier, *Annu. Rev. Astron. Astrophys.*, 2019, **57**, 113–155.
13 S. Manigand, J. Jørgensen, H. Calcutt, H. Müller, N. F. W. Ligterink, A. Coutens, M. N. Drozdovskaya, E. van Dishoeck and S. Wampfler, *Astron. Astrophys.*, 2020, **635**, A48.
14 A. Belloche, R. Garrod, H. Müller, K. Menten, I. Medvedev, J. Thomas and Z. Kisiel, *Astron. Astrophys.*, 2019, **628**, A10.
15 V. M. Rivilla, J. Martín-Pintado, I. Jiménez-Serra, S. Martín, L. F. Rodrguez-Almeida, M. A. Requena-Torres, F. Rico-Villas, S. Zeng and C. Briones, *Astrophys. J., Lett.*, 2020, **899**, L28.
16 A. Belloche, A. Maury, S. Maret, S. Anderl, A. Bacmann, P. André, S. Bontemps, S. Cabrit, C. Codella and M. Gaudel, *et al.*, *Astron. Astrophys.*, 2020, **635**, A198.
17 D. Skouteris, N. Balucani, C. Ceccarelli, F. Vazart, C. Puzzarini, V. Barone, C. Codella and B. Lefloch, *Astrophys. J.*, 2018, **854**, 135.
18 E. Herbst, *Int. Rev. Phys. Chem.*, 2017, **36**, 287–331.
19 L. Podio, A. Garufi, C. Codella, D. Fedele, E. Bianchi, F. Bacciotti, C. Ceccarelli, C. Favre, S. Mercimek and K. Rygl, *et al.*, *Astron. Astrophys.*, 2020, **642**, L7.
20 M. Van Gelder, B. Tabone, E. van Dishoeck, H. Beuther, A. Boogert, A. C. O. Garatti, P. Klaassen, H. Linnartz, H. Müller and V. Taquet, *et al.*, *Astron. Astrophys.*, 2020, **639**, A87.
21 E. Bianchi, C. Codella, C. Ceccarelli, F. Vazart, R. Bachiller, N. Balucani, M. Bouvier, M. De Simone, J. Enrique-Romero and C. Kahane, *et al.*, *Mon. Not. R. Astron. Soc.*, 2019, **483**, 1850–1861.
22 S. Zeng, I. Jiménez-Serra, V. Rivilla, S. Martín, J. Martín-Pintado, M. Requena-Torres, J. Armijos-Abendaño, D. Riquelme and R. Aladro, *Mon. Not. R. Astron. Soc.*, 2018, **478**, 2962–2975.
23 Y. Chen, W. R. M. Rocha, E. F. van Dishoeck, M. L. van Gelder, P. Nazari, K. Slavicinska, L. Francis, B. Tabone, M. Ressler and P. Klaassen, *et al.*, *Astron. Astrophys.*, 2024, **690**, A205.
24 V. Taquet, E. S. Wirström and S. B. Charnley, *Astrophys. J.*, 2016, **821**, 46.
25 R. T. Garrod, M. Jin, K. A. Matis, D. Jones, E. R. Willis and E. Herbst, *Astrophys. J., Suppl. Ser.*, 2022, **259**, 1.
26 P. Ehrenfreund and S. B. Charnley, *Annu. Rev. Astron. Astrophys.*, 2000, **38**, 427–483.
27 W. T. Huntress, Jr. and G. F. Mitchell, *Astrophys. J.*, 1979, **231**, 456–467.
28 J. M. Bowman, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16061–16062.
29 J. M. Bowman and A. G. Suits, *Phys. Today*, 2011, **64**, 33–37.
30 J. M. Bowman and P. L. Houston, *Chem. Soc. Rev.*, 2017, **46**, 7615–7624.
31 H. Kaur, W. Wang, J. Pérez-Ríos, B. Martínez-Haya, B. Redlich and S. Brünken, *ACS Earth Space Chem.*, 2025, DOI: **10.1021/acsearthspacechem.4c00395**.
32 W. Wang, X. Liu and J. Pérez-Ros, *J. Phys. Chem. A*, 2021, **125**, 5670–5680.
33 S. Y. Haoyu, X. He, S. L. Li and D. G. Truhlar, *Chem. Sci.*, 2016, **7**, 5032–5051.
34 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.
35 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
36 W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637–649.
37 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
38 E. Marinari and G. Parisi, *Europhys. Lett.*, 1992, **19**, 451.
39 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
40 W. Wang, X. Liu and J. Pérez-Ros, *Molecules*, 2023, **29**, 222.
41 M. Ceriotti, *J. Chem. Phys.*, 2019, **150**, 150901.
42 J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.*, 2011, **134**, 174105.
43 M. R. Shirts and J. D. Chodera, *J. Chem. Phys.*, 2008, **129**, 124105.
44 J.-H. Prinz, B. Keller and F. Noé, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16912–16927.

45 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
46 E. Vanden-Eijnden, *et al.*, *J. Stat. Phys.*, 2006, **123**, 503–523.
47 E. Vanden-Eijnden, *et al.*, *Annu. Rev. Phys. Chem.*, 2010, **61**, 391–420.
48 M. Mirahmadi and J. Pérez-Ros, *Int. Rev. Phys. Chem.*, 2022, **41**, 233–267.
49 A. F. Jalbout, L. Abrell, L. Adamowicz, R. Polt, A. Apponi and L. Ziurys, *Astrobiology*, 2007, **7**, 433–442.
50 A. K. Eckhardt, M. M. Linden, R. C. Wende, B. Bernhardt and P. R. Schreiner, *Nat. Chem.*, 2018, **10**, 1141–1147.

Open Access Article. Published on 12 July 2025. Downloaded on 2/12/2026 10:07:55 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.