

Furman University

Furman University Scholar Exchange

Chemistry Publications

Chemistry

2020

Water-Mediated Peptide Bond Formation in the Gas Phase: A Model Prebiotic Reaction

George C. Shields
Furman University

Ariel Gale

Tuguldur T. Odbadrakh
Furman University

Tyler Ball

Follow this and additional works at: <https://scholarexchange.furman.edu/chm-publications>

 Part of the [Physical Chemistry Commons](#)

Recommended Citation

Shields, George C.; Gale, Ariel; Odbadrakh, Tuguldur T.; and Ball, Tyler, "Water-Mediated Peptide Bond Formation in the Gas Phase: A Model Prebiotic Reaction" (2020). *Chemistry Publications*. 5.
<https://scholarexchange.furman.edu/chm-publications/5>

This Article (Journal or Newsletter) is made available online by Chemistry, part of the Furman University Scholar Exchange (FUSE). It has been accepted for inclusion in Chemistry Publications by an authorized FUSE administrator. For terms of use, please refer to the [FUSE Institutional Repository Guidelines](#). For more information, please contact scholarexchange@furman.edu.

Water-Mediated Peptide Bond Formation in the Gas Phase: A Model Prebiotic Reaction

Ariel G. Gale, Tuguldur T. Odbadrakh, Benjamin T. Ball, and George C. Shields*

Cite This: *J. Phys. Chem. A* 2020, 124, 4150–4159

Read Online

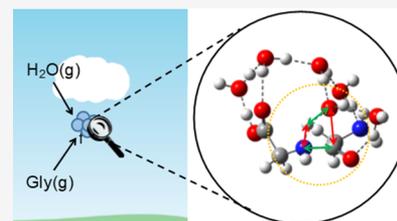
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The emergence of life on the prebiotic Earth must have involved the formation of polypeptides, yet the polymerization of amino acids is thermodynamically unfavorable under biologically relevant aqueous conditions because amino acids are zwitterions in solution and because of the production of a water molecule through a condensation reaction. Many mechanisms for overcoming this thermodynamic unfavorability have been proposed, but the role of gas phase water clusters has not been investigated. We present the thermodynamics of the water-mediated gas phase dimerization reaction of glycine as a model for the atmospheric polymerization of amino acids prior to the emergence of biological machinery. We hypothesize that atmospheric aerosols may have played a major role in the prebiotic formation of peptide bonds by providing the thermodynamic driving force to facilitate increasingly stable linear oligopeptides. In addition, we hypothesize that small aerosols orient amino acids on their surfaces, thus providing the correct molecular orientations to funnel the reaction pathways of peptides through transition states that lead eventually to polypeptide products. Using density functional theory and a thorough configurational sampling technique, we show that the thermodynamic spontaneity of the linear dimerization of glycine in the gas phase can be driven by the addition of individual water molecules.



INTRODUCTION

How living matter emerged from nonliving matter in the prebiotic chemistry of the early Earth is a profound and interesting question and has kept prebiotic chemists busy for decades.^{1–4} The grand scheme, as we understand it now, is as follows: the building blocks of life (amino acids, nucleotides, and lipids) formed a complex network of autocatalytic reactions and reaction cycles under prebiotic conditions, which grew in complexity and interdependence, leading eventually to the machinery of life. Amino acids were the most abundant of the three classes on prebiotic Earth, originating from meteorites,⁵ hydrothermal vents,⁶ and gaseous mixtures.^{7–9} The polymerization of amino acids is thermodynamically unfavorable in aqueous solution under biologically relevant conditions because amino acids are zwitterions in solution and because of the condensation reaction that produces a water molecule. Indeed, the experimental Gibbs free energy change for the uncatalyzed dimerization of aqueous glycine is 3.6 kcal mol⁻¹, leading to a small equilibrium constant and negligible equilibrium concentrations of the peptide versus free amino acids.¹⁰ This is because the zwitterionic form of amino acids in water deactivates the amino group and inhibits the nucleophilic attack on the carboxyl carbon of the neighboring amino acid molecule necessary for polymerization.¹¹ Biological systems overcome these thermodynamic and kinetic barriers via adenosine triphosphate (ATP) activation combined with low-energy transition states facilitated by enzymes.¹² Yet, the synthesis of polypeptides via condensation must have occurred in the

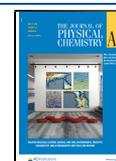
absence of such catalysts on or in the prebiotic Earth, and the search for the variety of ways this could occur has been fruitful and interesting.

One plausible mechanism for overcoming the thermodynamic barrier is to remove the solvent so that the equilibrium tends toward the polypeptide. Indeed, the theoretical gas phase uncatalyzed dimerization of glycine has a Gibbs free energy change of -2.2 kcal mol⁻¹ at the B3LYP/6-311+G(d,p) level of theory.¹² Atmospheric aerosols experience wet-dry cycles as they move through different humidity and temperature zones, which concentrate the solutes at the surface and orient them so that reactions can occur.^{13–16} Complete dehydration is not necessary as synthetic organic chemistry methodologies have confirmed the catalytic activity of stoichiometric quantities of water.¹⁷ While biotic systems are far from equilibrium, their abiotic precursors in the early atmosphere should have reached thermodynamic equilibrium. Under these conditions, thermodynamics can provide insights into the chemical composition of the atmosphere given the pressure, temperature, and volume.^{18–20} In fact, Dayhoff *et al.* state in a study of the content of the atmosphere:¹⁸

Received: April 1, 2020

Revised: April 28, 2020

Published: April 29, 2020



“In regions where special mechanisms operate to favor the production of certain compounds, there will be also a tendency to form the products required by thermodynamic equilibrium. Compounds which can be formed at equilibrium need no such special mechanism to explain their presence.”

If the standard state Gibbs free energy for polymerization of amino acids bound to small atmospheric clusters of water to form peptides has a negative value, then, given enough time, these peptides will form. Furthermore, if the gas phase products are subsequently solvated into the aqueous phase or scavenged by larger aerosols, then Le Chatelier's principle dictates that the gas phase polymerization reactions on small prenucleation nanoclusters will continue to consume the reactants. Based on these assumptions, we hypothesize that weakly interacting atmospheric clusters containing water and glycine could have assembled in such a way that the formation of diglycine through condensation is thermodynamically favorable. Such a reaction in the aqueous phase is thermodynamically prohibited due to the large excess of water and the zwitterionic nature of aqueous amino acids.

Prenucleation clusters, which precede the formation of atmospheric aerosols, are small gas phase molecular clusters up to 1 nm in diameter and are ubiquitous in experimental and theoretical research into atmospheric chemistry. Studies have shown that water molecules thermodynamically favor the formation of prenucleation clusters by stabilizing the products of hydrated sulfuric acid clusters, ammonia clusters, mixed sulfuric acid and ammonia clusters as well as those including alkylamines.^{21–40} Although various families of small molecules have been studied in great detail, research into the growth of prenucleation clusters involving amino acids has only begun.^{41,42} One intriguing aspect of these molecular clusters containing amino acids is the possibility of gas phase polymerization of the amino acids to form peptides. In this paper, we present compelling arguments for gas phase glycine dimerization as a model for the first step in atmospheric amino acid polymerization. First, we describe our methodology for computing the Gibbs free energies of the relevant reactions. Then, our results are split into three sections: (i) the abundance of hydrated glycine monomers in the atmosphere, (ii) the thermodynamics of glycine dimerization, and (iii) the kinetics of the process.

METHODOLOGY

We have paid particular attention to the sampling of the vast array of possible geometries of the Gly(H₂O)_n and Digly-(H₂O)_n clusters to isolate the most promising candidates for the global minimum energy structures. In past studies, we have generated initial starting geometries of molecular clusters by manual construction⁴³ utilizing chemical intuition or by extracting geometries from molecular dynamics (MD)^{35,44–47} and Monte Carlo (MC)^{48,49} simulations. Recently, we have employed a new sampling approach^{50–52} based on an evolutionary process using a genetic algorithm (GA).⁵³ This method has been implemented in the OGOLEM⁵⁴ program and can be interfaced with any model chemistry. Our usage of this method starts with an initial pool of 250–500 randomly generated guess structures depending on the size of the system. These structures are then optimized to the nearest local minimum energy geometry on the semiempirical PM7⁵⁵ potential energy surface implemented in the MOPAC program.⁵⁶ The algorithm then iterates through a process in which each iteration splices together a randomly chosen pair of

“parent” structures from the pool and then optimizes the “child” structures at the PM7 level of theory, discarding the child with the higher energy. After 10,000 iterations, the resulting set of structures was analyzed for uniqueness using a combination of each cluster's electronic energy and rotational constants to avoid identical structures. Two structures were deemed identical if their energies are within 0.10 kcal mol⁻¹, and all three of their rotational constants are within 1% of each other.⁵⁷ The unique local minimum structures obtained from the GA are then optimized using density functional theory (DFT) with the PW91 exchange-correlation functional⁵⁸ and the 6-31+G* Pople basis set⁵⁹ implemented in the Gaussian 09 program.⁶⁰ The PW91 density functional was used due to its favorable performance toward gas phase molecular clusters.^{61,62} The density functionals PW91, WB97, and M06-2X predict the same correlation and similar ΔG° values for the formation of mixed base prenucleation clusters.⁵¹ We found that PW91 ΔG^\ddagger values for the dimerization of glycine are within 0.1–2.5 kcal mol⁻¹ of the average of the same three density functionals. The unique structures from the PW91/6-31+G* optimizations were then further optimized at the PW91/6-311++G** level of theory with the strict convergence criteria of $\Delta E < 1 \times 10^{-6}$ au, root mean squared gradient $< 3 \times 10^{-5}$ au \AA^{-1} , and root mean squared displacement $< 6 \times 10^{-4}$ \AA . Finally, the harmonic vibrational frequencies of the resulting unique structures were computed at the same level of theory to ensure that no imaginary frequencies were present. The possibility of different protonation states of the peptides was accounted for by pursuing the above method for both the canonical and the zwitterionic peptides independently.

We also examined the effect of water molecules on the transition states involved in glycine dimerization. Here, we chose the cis and trans gas phase transition state structures identified by Dornshuld *et al.* as a starting point for the initial geometries.⁶³ Using the GA process, we generated 500 clusters with *n* water molecules placed around the initial transition state structure. The solvating water molecules were allowed to relax on the PM7 potential energy surface while keeping the central transition state structure frozen. The resulting structures were screened for uniqueness, and the highest energy structures were optimized to the nearest PW91/6-31+G* transition state using Gaussian 09.⁶⁰ We found that, if we used the lower energy structures as in the geometry optimization procedure, then the energy difference between the starting and the saddle point was often too great for the optimization algorithm to converge to the desired transition state structure. Beginning with the higher energy PM7 structures addressed this issue and resulted in successful convergence to the PW91 transition states. The resulting structures were analyzed visually to determine if the transition state structure was, in fact, at the saddle point of the dimerization reaction. In addition, the harmonic vibrational frequencies and normal mode displacements were used to confirm the validity of the transition states. Structures that showed lengthening and shortening of the correct bonds in the displacements corresponding to the one imaginary frequency were then used as the input geometry for intrinsic reaction coordinate calculations^{64,65} at the same level of theory to confirm that the transition state lies on the correct reaction path. The structures corresponding to the last point on both the forward and reverse reaction paths were further optimized to complete the reaction coordinate. Finally, the optimized reactants and products were inspected visually to

confirm that the reactant, transition state, and product clusters correspond to the same reaction path.

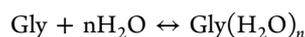
The final electronic and vibrational energies of all converged structures were used to compute the zero-point corrected energy $[E(0)]$, finite temperature corrected energy $[E(T)^\circ]$, enthalpy $[H(T)^\circ]$, and Gibbs free energy $[G(T)^\circ]$ for each cluster at a standard pressure of 1 atm. This was accomplished by using the rigid rotor/harmonic oscillator (RRHO) approximation⁶⁶ in the THERMO.PL script from the National Institute of Standards and Technology.⁶⁷ We computed concentrations of the various clusters by solving sets of equilibrium equations at three different temperatures corresponding to various altitudes within the troposphere.⁶⁸ A pictorial representation of our full methodology for configurational sampling can be found in Figure S1 of the Supporting Information. The above procedure was repeated using the density-based SMD implicit solvation model⁶⁹ implemented in Gaussian 09 to evaluate the effects of local hydration in the aqueous phase.

RESULTS AND DISCUSSION

Hydrated Glycine Content of the Atmosphere. For the dimerization of glycine to occur in the atmosphere, there must be a significant amount of glycine to start with. To establish the abundance of atmospheric glycine, we present the concentrations of $\text{Gly}(\text{H}_2\text{O})_{n=0-5}$ computed under 1 atm pressure and at temperatures of 217, 273, and 298 where 217 K represents the temperature at an altitude that is at the top of the troposphere and 298 K is at sea level. The Cartesian coordinates of the monomers employed in this step are listed in Data S1 of the Supporting Information, while coordinates of the bound clusters can be found in Data S2. We accounted for the possibility of the zwitterionic form of glycine by pursuing two separate configurational sampling runs: one starting with the canonical glycine monomer as an input for the GA and the other starting with the zwitterionic form of glycine. The resulting clusters' energies were compared, and the global minimum energy structure was chosen from the combined set of clusters containing both neutral and ionic forms. We found that the $n = 5$ structure, labeled G_5 in Figure 1 below, is a zwitterion, indicating that the hydration shell has deprotonated the carboxyl group and protonated the amino group in each of these structures. This agrees with previous studies that have

suggested that the gas phase glycine molecule first accesses the zwitterionic form when five water molecules are bound to it.⁷⁰ Figure 1 shows the geometries of $\text{Gly}(\text{H}_2\text{O})_{n=0-5}$ and reveals that water molecules prefer to bind to the carboxylic acid group rather than the amine group, except for the zwitterion G_5 . This is most likely due to the stronger electronegativity of oxygen compared to nitrogen, resulting in stronger hydrogen bonds. For $n = 1, 2, 3,$ and 4 , labeled $\text{G}_1, \text{G}_2, \text{G}_3,$ and G_4 , respectively, the water molecules are observed to form stable hydrogen bond networks on the carboxylic acid group with a water molecule donating a hydrogen bond to the deprotonated side of the carboxylic acid group ($\text{HO}-\text{H} \rightarrow \text{O}=\text{C}-\text{OH}$), while another water molecule accepts a hydrogen bond from the opposite side of the carboxylic acid ($\text{H}_2\text{O} \leftarrow \text{H}-\text{O}-\text{C}=\text{O}$). Interestingly, the hydrogen-bonded end of G_2 resembles the cyclic water tetramer with two of the water molecules of the tetramer replaced by the carboxylic acid group of glycine. G_3 shows a similar correspondence to the cyclic water pentamer, and both structures have sequential donor–acceptor hydrogen bonds. This kind of structure is called a homodromic hydrogen bond network and has been shown to stabilize small water clusters through nonadditive many-body interactions.^{44,71,72}

Table 1 lists the formation energies of $\text{Gly}(\text{H}_2\text{O})_{n=1-5}$ for the reactions



where $n = 1-5$ at three different temperatures. Energies are presented to the hundredth decimal place to avoid rounding errors; however, we expect the accuracy of these calculations to be in the 10th decimal place or higher. The harmonic vibrational frequencies used to compute the thermodynamics of the equation above are listed in Tables S1, S3, and S6 to S11 in the Supporting Information. The temperatures considered below correspond to a latitude of 50° from the equator where $T = 298.15$ K at sea level, $T = 273.15$ K at 2 km above sea level, and $T = 216.65$ K at the top of the troposphere 10 km above sea level.⁷⁴ We note that, while we accounted for the altitude–temperature dependence of the concentration of water in our calculations, we did not vary the concentration of isolated glycine as we have not found experimental data that shows variation with altitude. The Gibbs free energies of formation show a dramatic decrease from -1.07 kcal mol⁻¹ for G_1 to -4.06 kcal mol⁻¹ for G_2 at room temperature. Under colder conditions where entropy effects are less prevalent, G_1 has an even more negative Gibbs free energy. The geometries shown in Figure 1 reveal that the water molecule in G_1 is hydrogen-bonded to the carboxyl group of glycine in a rather strained geometry, while G_2 has three hydrogen bonds and relaxed water molecules bound to the carboxyl group of glycine. These three hydrogen bonds are of the donor–acceptor type, forming a stable ring using the carboxylic acid group of glycine to a negative enthalpy of formation with a magnitude of 23 kcal mol⁻¹. A similar hydrogen bond ring forms in G_3 with a free energy change of -5.38 kcal mol⁻¹, creating a five-member ring of donor–acceptor hydrogen bonds with a negative enthalpy of formation of -32.9 kcal mol⁻¹. Perhaps the most interesting structure is G_4 , which, at a room temperature Gibbs formation energy of -5.78 kcal mol⁻¹, is the most thermodynamically favored. Here, the carboxylic acid group of glycine is hydrogen-bonded to the very stable water tetramer^{71,75} in the donor–acceptor connectivity. The planar structure observed in $\text{G}_1, \text{G}_2,$ and

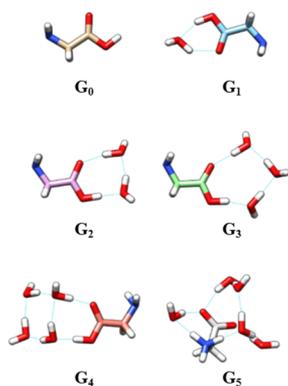


Figure 1. PW91/6-311++G** minimum energy structures of $\text{Gly}(\text{H}_2\text{O})_{n=0-5}$ generated using UCSF Chimera.⁷³ Hydrogen bonds are drawn in cyan, O in red, N in blue, and H in white. Cartesian coordinates are included in Data S1 and S2 of the supporting Information.

Table 1. PW91/6-311++G** Energies, Enthalpies, and Free Energies of Formation for the Reaction $\text{Gly} + n\text{H}_2\text{O} \leftrightarrow \text{Gly}(\text{H}_2\text{O})_{n=1-5}$ at 0, 217, 273, and 298 K, for a Standard State of 1 atm Pressure^a

<i>n</i>	0 K		216.65 K		273.15 K		298.15 K	
	ΔE°	ΔH°	ΔG°	ΔH°	ΔG°	ΔH°	ΔG°	
1	-9.85	-10.61	-3.68	-10.61	-1.87	-10.59	-1.07	
2	-21.53	-23.10	-9.27	-23.11	-5.66	-23.09	-4.06	
3	-30.72	-32.88	-12.90	-32.87	-7.69	-32.82	-5.38	
4	-40.77	-43.92	-16.22	-43.96	-8.98	-43.92	-5.78	
5	-49.90	-54.28	-18.08	-54.44	-8.62	-54.44	-4.42	

^aAll energies are zero-point-corrected using harmonic vibrational frequencies and are in units of kcal mol⁻¹. At 0 K, $\Delta E^\circ = \Delta H^\circ = \Delta G^\circ$. The lowest Gibbs free energies are emphasized in bold font.

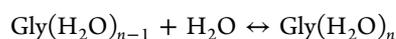
Table 2. PW91/6-311++G** Energies, Enthalpies, and Free Energies for $\text{Gly}(\text{H}_2\text{O})_{n-1} + \text{H}_2\text{O} \leftrightarrow \text{Gly}(\text{H}_2\text{O})_n$ at 0, 217, 273, and 298 K, for a Standard State of 1 atm Pressure^a

<i>n</i>	0 K		216.65 K		273.15 K		298.15 K	
	ΔE°	ΔH°	ΔG°	ΔH°	ΔG°	ΔH°	ΔG°	
1	-9.85	-10.61	-3.68	-10.61	-1.87	-10.59	-1.07	
2	-11.68	-12.49	-5.59	-12.51	-3.78	-12.50	-2.99	
3	-9.20	-9.78	-3.63	-9.76	-2.03	-9.73	-1.33	
4	-10.04	-11.04	-3.32	-11.09	-1.30	-11.10	-0.40	
5	-9.14	-10.36	-1.86	-10.48	0.37	-10.51	1.36	

^aAll energies are zero-point-corrected using harmonic vibrational frequencies and are in units of kcal mol⁻¹. At 0 K, $\Delta E^\circ = \Delta H^\circ = \Delta G^\circ$. The lowest Gibbs free energies are emphasized in bold font.

G₃ gives way to a more three-dimensional shape with the water tetramer approximately 113° bent out of the plane of glycine. In fact, this closely resembles the “book” isomer of the gas phase water hexamer but with one “page” of the book replaced by the glycine molecule’s carboxylic acid group. The water tetramer has been proposed as an important workhorse in the atmosphere, absorbing IR radiation in its ring system, which is then available for carrying out chemical reactions.⁷⁶ Samala and Agmon argue that logarithmically small tails in the radial probability densities of (H₂O)_{n=2–6} clusters provide direct evidence that the cyclic water tetramer is stable in atmospheric layers.⁷⁶ They deduce a temperature for vaporization of 240 K from tails in their molecular dynamics simulations. Previous high level quantum chemistry results predict that the equilibrium constant for the reaction of four waters to form the cyclic water tetramer is one at 268–269 K, which is another way to think about when the concentration of monomers and tetramers are at equilibrium ($\Delta G^\circ = 0$, so $K = 1$).⁷⁵ Indeed, its unique stability^{44,75} is apparent in **G**₄. The final cluster, **G**₅, displays a completely three-dimensional cage-like structure with the water molecules and the carbon backbone of glycine facing in opposite directions. This can be interpreted as a microscopic form of hydrophobic interactions where the alkyl backbone avoids the water molecules. Additionally, the zwitterionic form of glycine in **G**₅ breaks the homodromic donor–acceptor hydrogen bond network by forcing the amine group into a donor–donor connectivity and the carboxylic acid group into an acceptor–acceptor connectivity. This inability to maximize the number of resonant donor–acceptor hydrogen bonds results in a smaller formation free energy than **G**₃ and **G**₄ at higher temperatures.

Table 2 lists the sequential hydration energies of glycine representing the energetics of the process of adding water molecules one at a time as



for $n = 1–5$. Based on these values, we have calculated the atmospheric concentrations of $\text{Gly}(\text{H}_2\text{O})_{n=0–5}$ under three different relative humidity values corresponding to local variations of water saturation by solving sets of equilibrium equations⁶⁸ at the previously mentioned altitudes and temperatures. The derivations for these equations are given in ref S2. Figure 2 shows the results grouped by temperature.

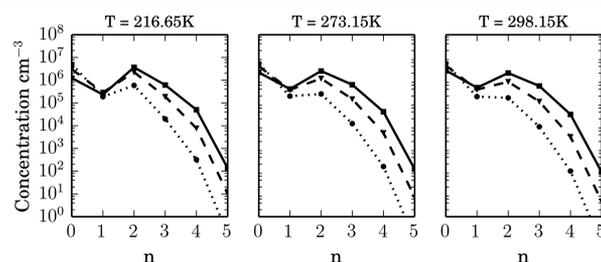


Figure 2. Equilibrium distribution of $\text{Gly}(\text{H}_2\text{O})_{n=0-5}$ in units of clusters cm^{-3} at relative humidity values of 100 (solid line), 50 (dashed line), and 20% (dotted line) and temperatures of 216.65, 273.15, and 298.15 K computed using the Gibbs free energies of sequential hydration listed in Table 2. $[\text{Gly}]_0 = 2.9 \times 10^6 \text{ cm}^{-3}$,^{80,81} $[\text{H}_2\text{O}]_{\text{RH}=100\%} = 9.9 \times 10^{14}$,¹⁴ 1.6×10^{17} ,¹⁷ and $7.7 \times 10^{17} \text{ cm}^{-3}$ at $T = 216.65$, 273.15, and 298.15 K, respectively, at 1 atm pressure.^{35,70} Note that the y-axis is logarithmic.

G₀ and **G**₂ are found to be the most abundant clusters at all temperatures at 100% relative humidity, as shown by the solid black line, while the concentrations of **G**₃ and **G**₄ are each smaller by around 1 order of magnitude, and the **G**₅ concentration is smaller by about 2 orders of magnitude. The equilibrium abundances of these clusters seem to be insensitive to altitude based on the lack of appreciable change between the three temperatures considered. This is because of the interplay between the magnitude of the free energy changes and the variation in water concentration as a function of temperature (and thus altitude). As altitude increases,

Table 3. PW91/6-311++G** Energies, Enthalpies, and Gibbs Free Energy Changes for the Reaction $2\text{Gly} + n\text{H}_2\text{O} \leftrightarrow \text{Digly}(\text{H}_2\text{O})_{n+1}$ where $n = 0-8$, at 0, 217, 273, and 298 K at a Standard State Pressure of 1 atm^a

n	0 K		216.65 K		273.15 K		298.15 K	
	ΔE°	ΔH°	ΔG°	ΔH°	ΔG°	ΔH°	ΔG°	
aq	-5.53	-5.60	2.75	-5.35	4.90	-5.23	5.84, 3.6 ^b	
0	-14.30	-14.41	-6.00	-14.24	-3.83	-14.16	-2.88	
1	-26.69	-27.44	-12.67	-27.25	-8.85	-27.15	-7.16	
2	-37.56	-39.47	-16.25	-39.37	-10.20	-39.29	-7.53	
3	-47.11	-49.58	-20.26	-49.47	-12.62	-49.36	-9.25	
4	-59.98	-63.64	-25.30	-63.58	-15.31	-63.49	-10.89	
5	-70.92	-75.34	-29.97	-75.30	-18.13	-75.21	-12.91	
6	-81.14	-85.96	-34.76	-85.85	-21.41	-85.71	-15.52	
7	-92.62	-98.29	-40.35	-98.25	-25.23	-98.13	-18.55	
8	-102.16	-108.56	-43.11	-108.48	-26.04	-108.33	-18.50	

^aAll energies are zero-point-corrected using harmonic vibrational frequencies and are in units of kcal mol⁻¹. The row labeled as "aq" represents the uncatalyzed dimerization energy within the SMD implicit solvation model, while the rest of the table represents the gas phase dimerization energies at $n = 0-8$. The lowest Gibbs free energies are emphasized in bold font. ^bThe experimental numbers are from refs 10 and 82.

temperature drops, and the concentration of water decreases by 3 orders of magnitude from sea level to the top of the troposphere, while the free energy becomes more negative. So, at the top of the troposphere (the tropopause), the greater magnitude of the free energy is opposed by a smaller concentration of water. However, at each temperature, the dependence of concentration on relative humidity is pronounced. For the $n > 1$ clusters, a decrease in relative humidity corresponds to a decrease in concentration, while the unhydrated glycine concentration increases. This is simply due to the decrease in the amount of hydrated glycine, leaving more unhydrated glycine in the atmosphere. This equilibrium distribution of $\text{Gly}(\text{H}_2\text{O})_{n=0-5}$ shows that there are upwards of 1 million hydrated glycine clusters with one through four water molecules in 1 mL of air at thermodynamic equilibrium. Combined with sea-level humidity data measured from satellites,^{77,78} we can postulate about the local abundances of hydrated glycine based on relative humidity. For example, data collected by Willett *et al.* shows that tropical and coastal areas stay close to 100% relative humidity, while deserts and high plateaus such as the Sahara and the American West maintain 20% relative humidity.⁷⁹ This suggests that there are more $n = 2-4$ clusters in the air above the tropics and coasts and fewer of these clusters further inland. Based on this equilibrium distribution of concentrations, we assume that there are enough glycine and hydrated glycine clusters in the atmosphere for dimerization reactions to occur.

Water-Mediated Formation of Diglycine. Next, we examine the gas phase dimerization of glycine as a function of hydration. Table 3 lists the PW91/6-311++G** changes in energy for the reactions



for $n = 0-8$ in the gas phase as well as for the uncatalyzed reaction ($n = 0$) computed within the SMD implicit solvation model. The starting monomers for this step are listed in Data S1, and the resulting clusters are listed in Data S3. This process differs from the hydration of a single glycine molecule in the previous section in that a peptide bond is formed between the amine group of one glycine molecule and the carboxyl group of another, resulting in the formation of one additional water molecule. Using the vibrational frequencies listed in Tables S1 to S5 and S12 to S20 in the Supporting Information, we observe that, while the dimerization energy is enthalpically

favored in the aqueous phase, the inclusion of entropic effects makes it thermodynamically unfavorable and nonspontaneous. In contrast, even the uncatalyzed gas phase dimerization is thermodynamically favorable. Figure 3 displays the geometries

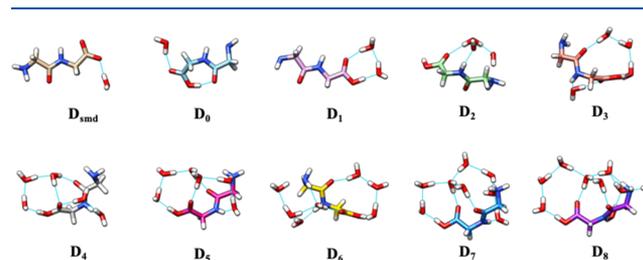


Figure 3. PW91/6-311++G** minimum energy structures of $\text{Digly}(\text{H}_2\text{O})_{n=1-9}$. Hydrogen bonds are drawn in cyan, O atoms in red, N in blue, and H in white. The aqueous phase (SMD solvated) structure is labeled D_{smd} , and the gas phase structures are labeled D_n where n is the number of catalytic water molecules in the dimerization reaction. Cartesian coordinates are presented in Data S3 of the Supporting Information.

of the minimum energy $\text{Digly}(\text{H}_2\text{O})_{n=1-9}$ produced from the dimerization of two glycine molecules and $n = 0-8$ water molecules. The product $\text{Digly}(\text{H}_2\text{O})$ was optimized in the aqueous phase using the SMD implicit solvation method (D_{smd}) and in the gas phase (D_0) at the same level of theory. D_{smd} is a zwitterion and has an elongated linear structure because of the solvation environment's influence, while the gas phase diglycine molecule is free to form one intramolecular hydrogen bond, as shown in D_0 . Most importantly, the aqueous phase dimerization energy is positive, while the gas phase dimerization energies are negative at all temperatures in the troposphere. This is direct evidence of thermodynamically favored peptide bond formation in the gas phase.

At $n = 1$, the room temperature dimerization free energy drops to -7.16 kcal mol⁻¹, a change of about 4.28 kcal mol⁻¹, increasing the formation of products by a factor of 3 orders of magnitude. The energetic change is a direct result of the formation of a hydrogen bond ring at the carboxyl end of the diglycine molecule similar to the doubly hydrated glycine monomer G_2 . The inclusion of one water molecule in the reactants seems to have the largest effect on dimerization when compared to the $n > 1$ reactions. With one additional water molecule, the dimerization energy does not change appreciably

at higher temperatures, but the geometry of the cluster, D_2 , undergoes a drastic change. This is the first occurrence of a three-dimensional hydrogen bond network. Further hydration results in a decrease of about $1.5 \text{ kcal mol}^{-1}$ per additional water molecule and continues until D_7 . We attribute this to the stepwise stabilization produced by the rearrangement of the hydrogen bond network bound to the diglycine molecule as water molecules are added, which releases the strain in the network. Each $1.5 \text{ kcal mol}^{-1}$ decrease shifts the equilibrium in the direction of the product dipeptide by about another order of magnitude at 298 K. Note that, at the top of the troposphere, at 217 K, the increase in the magnitude of the free energy is greater, about $4\text{--}6 \text{ kcal mol}^{-1}$ per water up to seven waters, with each additional water shifting the equilibrium further to the right. The addition of one more water molecule in D_8 has no appreciable effect on the dimerization energy at 298 K, which shows that the hydrogen bond network is flexible enough at this point to accommodate one more water molecule without affecting the diglycine molecule in an energetically favorable or unfavorable way.

The products after D_1 show a striking structural similarity where homodromic rings of hydrogen bonds result in a bent diglycine structure. We observe the formation of three such rings as more water molecules are added; networks H1, H2, and H3 are shown schematically in Figure 4. In D_2 , one ring

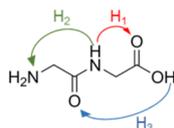


Figure 4. Schematic representation of the homodromic rings H1 (red), H2 (green), and H3 (blue) observed in the $\text{Digly}(\text{H}_2\text{O})_n$ clusters.

starts at the amide proton, passes through two water molecules, and ends at the terminal carboxyl oxygen (Figures 3 and 4, H1). The other ring starts on the same amide proton as H1, connects across two water molecules, and ends at the terminal amine nitrogen (Figures 3 and 4, H2). These two rings share one donor–donor–acceptor water molecule that acts as the point where the rings fuse. The third type of homodromic hydrogen bond ring appears in D_3 , originating from the terminal carboxyl proton and ending at the amide oxygen (Figures 3 and 4, H3). This ring passes through three water molecules in D_3 and works in combination with the H1 ring to hold the dipeptide in a bent geometry. Notably, the twisting caused in the peptide backbone by the H3 ring seems to have allowed the H1 ring to form with just one water molecule in D_3 , as opposed to two water molecules in D_2 .

With the addition of more water molecules, the H3 ring starts to dominate the overall structure of the hydrated dipeptide as it grows from three water molecules in D_{4-6} to four and six water molecules in D_7 and D_8 . The H1 ring persists in D_{4-6} but maintains the same number of water molecules; however, in D_7 , it fuses with the H3 ring through its second water molecule, and in D_8 , it fuses with H3 through its fourth water molecule. The H2 ring, on the other hand, goes through a variety of sizes as more water molecules are added. In D_4 , H2 is unchanged compared to D_2 but then fuses with the H1 ring in D_5 and D_6 . Here, the second water molecule in H1 acts as the bifurcation point, creating the H2 hydrogen bond network leading to the terminal amine nitrogen. These

structural features are vital in stabilizing the overall products $\text{Digly}(\text{H}_2\text{O})_{n=1-9}$ in the gas phase and may grow as even more water molecules are added. In addition, as the cluster grows in the number of water molecules, a noticeable grouping of the water molecules forms. This aggregation of water molecules occurs such that the alkyl backbone of the diglycine molecule points away from the center of the cluster, much in the same way as hydrophobic interactions occur. Under thermodynamic equilibrium, this interaction can hold the dipeptide in such an arrangement as to allow the next monomer to attach to one of the termini. In this way, gas phase polymerization through peptide bond formation may be possible.

Kinetic Considerations. While the dimerization reactions described above are thermodynamically favorable, we also pursued a line of inquiry into the kinetics of glycine dimerization. To this end, we computed the activation energies for dimerization as the Gibbs free energy difference between the transition state and the infinitely separated reactants while varying the number of water molecules. Our results showed that, for the $n = 0\text{--}5$ dimerization reactions, the activation energy is approximately $36\text{--}46 \text{ kcal mol}^{-1}$. A photon in this energy range would have a wavelength of around $790\text{--}620 \text{ nm}$, suggesting that the abundant visible and near-infrared radiation in the atmosphere would be sufficiently energetic to push the reaction over the dimerization barrier. Due to the completely asymmetric nature of the clusters, dipole-allowed transitions dominate the vibrational motion. The cyclic water tetramer, which stabilizes G_4 (Figure 1), has been proposed based on both molecular dynamics simulations and quantum chemistry to exist at temperatures in the troposphere.^{75,76} It is able to absorb energy from visible and IR radiation in its ring system, which can be used precisely for this reaction. Two individual gas-phase glycine molecules would not form diglycine in the absence of a third body that can absorb the heat of reaction from the formation of the peptide bond. However, gas phase molecular clusters are positioned to redistribute the energy of bond formation throughout the cluster. At a standard temperature and pressure, the collision rate of hydrated glycine clusters is on the order of 10^9 collisions per second so that the mean time between collisions is around $2 \times 10^{-10} \text{ s}$. We expect these clusters to be constantly bombarded by gases in the atmosphere, and at least sometimes, photon energy can be transferred wholly and in part to the bombarding molecules. This may result in the reaction completing as long as there is enough energy pumped into the system due to IR radiation and collisions.

Although the linear oligomerization of glycine has been shown to be spontaneous in the gas phase, there is an alternative dimer structure, which presents a potential problem. The diketopiperazine (DKP) form of glycine is a cyclic dimer with two peptide bonds formed from two separate condensation reactions. Therefore, the entropic contribution to the Gibbs free energy change is much larger than that for the linear dimer due to the production of one additional water molecule. This has long been known as the “DKP trap” and presents the largest hurdle toward the abiotic polymerization of amino acids as the highly stable six-member ring is the more favored product and faces a large uphill energy cost to reopen if polymerization is to continue.^{1,83} Potential mechanisms for overcoming this thermodynamic trap in peptide synthesis include manipulating the functionalization of the amino acids to alter the steric and torsional strains of ring formation and

the addition of protecting groups during synthesis.⁸⁴ However, in the gas phase, the DKP trap may not be a problem.

We have found that, in small atmospheric nanoclusters, two glycine molecules are oriented on the surfaces of the water clusters in favorable orientations for reactions to occur. In these clusters, the air–water interface has oriented the two glycine molecules for easy formation of the linear dipeptide, temporarily blocking the formation of the second peptide bond that would result in formation of DKP. We expect some formations of these clusters to react to form the linear dipeptide because the Gibbs activation free energy is decreased from 46 kcal mol⁻¹ for the gas phase reaction of two glycine molecules in the cis orientation to 43 kcal mol⁻¹ when one water molecule is present. In the presence of two to five waters, the Gibbs activation energy decreases by another 3–6 kcal mol⁻¹. Figure 5 shows the PW91/6-31+G* transition state

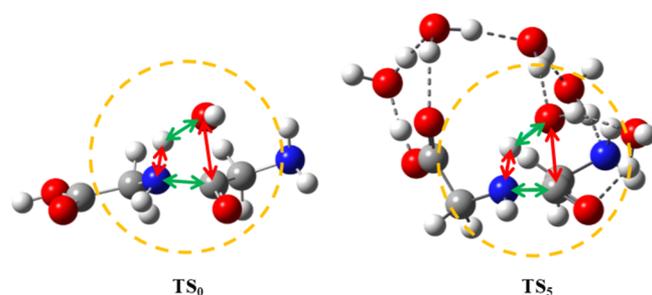


Figure 5. PW91/6-31+G* transition states for the reaction $2\text{Gly} + n\text{H}_2\text{O} \leftrightarrow \text{Digly}(\text{H}_2\text{O})_{n+1}$ for $n = 0$ (TS_0) and $n = 5$ (TS_5). Yellow dashed circles mark where covalent bonds are forming (green arrows) and breaking (red arrows). Hydrogen bonds are represented by the gray dashed lines.

structure for the gas phase reaction without waters, TS_0 , and for the reaction with five waters, TS_5 . The figure illustrates bonds broken (red arrows) and formed (green arrows) in this reaction where the peptide bond is in the process of formation during the condensation reaction that produces a water molecule. One can see that, for the $\text{Digly}(\text{H}_2\text{O})_5$ structure, the five waters form a hydrogen-bonded network, which stabilizes the transition state. The hydroxyl group dissociating from the second glycine molecule is stabilized by the water network, and the C...N peptide bond that is being formed is reduced from a length of 1.614 Å in the absence of the water network to 1.554 Å in the presence of five waters. Similarly, the breaking bonds, N...H and C...O are both elongated slightly, while the formation of the H...O bond to produce the sixth water is decreased slightly. This activated complex has a PW91/6-31+G* Gibbs free energy of 36.6 kcal mol⁻¹, almost 10 kcal mol⁻¹ lower than the gas phase reaction in the absence of water. Further details on the transition state calculations will be published in a separate paper.⁸⁵

CONCLUSIONS

We have shown that the dimerization of glycine through condensation is thermodynamically feasible in the gas phase in the presence of water and predict that monomeric glycine does persist in the atmosphere in abundant hydrated forms. Furthermore, the thermodynamic favorability of this reaction increases with the number of water molecules for $\text{Digly}(\text{H}_2\text{O})_{n=1-9}$ products. This increase is shown to be driven by the stability of the product clusters in which diglycine can bend

to maximize the intra- and intermolecular interactions, specifically hydrogen bonding through the formation of homodromic networks of self-reinforcing hydrogen bonds. This does not occur in the aqueous phase due to strong intermolecular interactions between the zwitterionic product and its solvation environment, which creates a high energy barrier to any bending motion to form intramolecular hydrogen bonds. Therefore, the intramolecular interactions and hydrogen bond rings seem to make a large contribution to the thermodynamic favorability of glycine dimerization. Additionally, we found that the $n = 3-9$ clusters have very similar homodromic hydrogen bond networks dominated by three general topologies: (i) connecting the amide proton to the terminal carboxyl oxygen, (ii) connecting the amide proton to the terminal amine nitrogen, and (iii) connecting the terminal carboxyl proton to the amide oxygen atom. Based on the energetic and structural similarities between D_7 and D_8 , we expect the linear diglycine to remain at the surface of the growing water cluster. We have also shown that the reaction barrier for linear dimerization of glycine in the gas phase is catalyzed by the addition of up to five water molecules. This suggests that the driving thermodynamic force of this process could dominate in the gas phase as hydration increases due to the increasingly negative Gibbs free energy change at all three temperatures considered. We hope that the gas phase model reaction described in this work will promote further research into the possible role of atmospheric molecular clusters in the prebiotic polymerization of biologically important molecules.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.0c02906>.

Our computational methodology for configurational sampling and computing thermodynamic corrections (Figure S1); harmonic vibrational frequencies and intensities of the gas phase water monomer (Table S1); harmonic vibrational frequencies and intensities of the aqueous water monomer (Table S2); harmonic vibrational frequencies and intensities of the gas phase glycine monomer (Table S3); harmonic vibrational frequencies and intensities of the aqueous glycine and gas phase diglycine monomers (Table S4); harmonic vibrational frequencies and intensities of the aqueous diglycine monomer (Table S5); harmonic vibrational frequencies and intensities of gas phase $\text{Gly}(\text{H}_2\text{O})_{n=1-5}$ (Tables S6 to S11); harmonic vibrational frequencies and intensities of aqueous $\text{Digly}(\text{H}_2\text{O})$ (Table S12a); harmonic vibrational frequencies and intensities of gas phase $\text{Digly}(\text{H}_2\text{O})$ (Table S12g); harmonic vibrational frequencies and intensities of gas phase $\text{Digly}(\text{H}_2\text{O})_{n=1-9}$ (Tables S13 to S20); Cartesian coordinates of water, glycine, and diglycine monomers in the gas and aqueous phase (Data S1); Cartesian coordinates of gas phase $\text{Gly}(\text{H}_2\text{O})_{n=1-5}$ clusters (Data S2); and Cartesian coordinates of gas phase $\text{Digly}(\text{H}_2\text{O})_{n=1-9}$, including $\text{Digly}(\text{H}_2\text{O})$ in the aqueous phase (Data S3) (PDF)

AUTHOR INFORMATION

Corresponding Author

George C. Shields – Department of Chemistry, Furman University, Greenville, South Carolina 29613, United States;

orcid.org/0000-0003-1287-8585; Email: george.shields@furman.edu

Authors

Ariel G. Gale – Department of Chemistry, Furman University, Greenville, South Carolina 29613, United States

Tuguldur T. Odbadrakh – Department of Chemistry, Furman University, Greenville, South Carolina 29613, United States

Benjamin T. Ball – Department of Chemistry, Furman University, Greenville, South Carolina 29613, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpca.0c02906>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants CHE-1229354, CHE-1662030, CHE-1721511, and CHE-1903871 from the National Science Foundation (GCS), the Arnold and Mabel Beckman Foundation Beckman Scholar Award (AGG), and the Barry M. Goldwater Scholarship (AGG). High-performance computing resources were provided by the MERCURY Consortium (<http://www.mercuryconsortium.org>), and molecular graphics were generated with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. Graphs and schematics were created with Python 2.7, Pyplot in Matplotlib 2.0.2, Gaussview 6.0.16. We thank Greg Springsteen for helpful conversations and Berhane Temelso for scientific and technical support.

REFERENCES

- (1) Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529.
- (2) Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev* **2014**, *114*, 285–366.
- (3) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.
- (4) Hud, N. V.; Fialho, D. M. RNA nucleosides built in one prebiotic pot. *Science* **2019**, *366*, 32–33.
- (5) Pizzarello, S.; Shock, E. The organic composition of carbonaceous meteorites: the evolutionary story ahead of biochemistry. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a002105.
- (6) Russell, M. J.; Hall, A. J.; Boyce, A. J.; Fallick, A. E. 100th anniversary special paper: On hydrothermal convection systems and the emergence of life. *Econ. Geol.* **2005**, *100*, 419–438.
- (7) Fitz, D.; Reiner, H.; Rode, B. M. Chemical evolution toward the origin of life. *Pure Appl. Chem.* **2007**, *79*, 2101–2117.
- (8) Kvenvolden, K.; Lawless, J.; Pering, K.; Peterson, E.; Flores, J.; Ponnampertuma, C.; Kaplan, I. R.; Moore, C. Evidence for extraterrestrial amino-acids and hydrocarbons in the Murchison meteorite. *Nature* **1970**, *228*, 923–926.
- (9) Bernstein, M. P.; Dworkin, J. P.; Sandford, S. A.; Cooper, G. W.; Allamandola, L. J. Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues. *Nature* **2002**, *416*, 401–403.
- (10) Martin, R. B. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers* **1998**, *45*, 351–353.
- (11) Danger, G.; Plasson, R.; Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **2012**, *41*, 5416–5429.
- (12) Remko, M.; Rode, B. M. Catalyzed peptide bond formation in the gas phase. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4667–4673.
- (13) Tuck, A. The role of atmospheric aerosols in the origin of life. *Surv. Geophys.* **2002**, *23*, 379–409.
- (14) Lee, J. K.; Banerjee, S.; Nam, H. G.; Zare, R. N. Acceleration of reaction in charged microdroplets. *Q. Rev. Biophys.* **2015**, *48*, 437–444.
- (15) Griffith, E. C.; Tuck, A. F.; Vaida, V. Ocean-atmosphere interactions in the emergence of complexity in simple chemical systems. *Acc. Chem. Res.* **2012**, *45*, 2106–2113.
- (16) Griffith, E. C.; Vaida, V. In situ observation of peptide bond formation at the water-air interface. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15697–15701.
- (17) Ribe, S.; Wipf, P. Water-accelerated organic transformations. *Chem. Commun.* **2001**, *4*, 299–307.
- (18) Dayhoff, M. O.; Lippincott, E. R.; Eck, R. V. Thermodynamic equilibria in prebiological atmospheres. *Science* **1964**, *146*, 1461–1464.
- (19) Eck, R. V.; Lippincott, E. R.; Dayhoff, M. O.; Pratt, Y. T. Thermodynamic equilibrium and the inorganic origin of organic compounds. *Science* **1966**, *153*, 628–633.
- (20) Morowitz, H. J. A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **1999**, *4*, 39–53.
- (21) Tiszenkel, L.; Stangl, C.; Krasnomowitz, J.; Ouyang, Q.; Yu, H.; Apsokardu, M. J.; Johnston, M. V.; Lee, S. H. Temperature effects on sulfuric acid aerosol nucleation and growth: initial results from the TANGENT study. *Atmos. Chem. Phys.* **2019**, *19*, 8915–8929.
- (22) Glasoe, W. A.; Volz, K.; Panta, B.; Freshour, N.; Bachman, R.; Hanson, D. R.; McMurry, P. H.; Jen, C. Sulfuric acid nucleation: An experimental study of the effect of seven bases. *J. Geophys. Res.: Atmos.* **2015**, *120*, 1933–1950.
- (23) Elm, J.; Jen, C. N.; Kurtén, T.; Vehkamäki, H. Strong Hydrogen Bonded Molecular Interactions between Atmospheric Diamines and Sulfuric Acid. *J. Phys. Chem. A* **2016**, *120*, 3693–3700.
- (24) Jen, C. N.; McMurry, P. H.; Hanson, D. R. Stabilization of sulfuric acid dimers by ammonia, methylamine, dimethylamine, and trimethylamine. *J. Geophys. Res.: Atmos.* **2014**, *119*, 7502–7514.
- (25) Chen, H.; Chee, S.; Lawler, M. J.; Barsanti, K. C.; Wong, B. M.; Smith, J. N. Size resolved chemical composition of nanoparticles from reactions of sulfuric acid with ammonia and dimethylamine. *Aerosol Sci. Technol.* **2018**, *52*, 1120–1133.
- (26) Doyle, G. J. Self-Nucleation in the Sulfuric Acid-Water System. *J. Chem. Phys.* **1961**, *35*, 795–799.
- (27) Sipilä, M.; Berndt, T.; Petäjä, T.; Brus, D.; Vanhanen, J.; Stratmann, F.; Patokoski, J.; Mauldin, R. L.; Hyvärinen, A.-P.; Lihavainen, H.; Kulmala, M. The Role of Sulfuric Acid in Atmospheric Nucleation. *Science* **2010**, *327*, 1243–1246.
- (28) Myllys, N.; Kubečka, J.; Besel, V.; Alfaouri, D.; Olenius, T.; Smith, J. N.; Passananti, M. Role of base strength, cluster structure and charge in sulfuric-acid-driven particle formation. *Atmos. Chem. Phys.* **2019**, *19*, 9753–9768.
- (29) Berndt, T.; Böge, O.; Stratmann, F.; Heintzenberg, J.; Kulmala, M. Rapid formation of sulfuric acid particles at near-atmospheric conditions. *Science* **2005**, *307*, 698–700.
- (30) Temelso, B.; Morrell, T. E.; Shields, R. M.; Allodi, M. A.; Wood, E. K.; Kirschner, K. N.; Castonguay, T. C.; Archer, K. A.; Shields, G. C. Quantum mechanical study of sulfuric acid hydration: Atmospheric implications. *J. Phys. Chem. A* **2012**, *116*, 2209–2224.
- (31) Zhang, R.; Wooldridge, P. J.; Abbatt, J. P. D.; Molina, M. J. Physical chemistry of the sulfuric acid/water binary system at low temperatures: stratospheric implications. *Journal of Physical Chemistry* **1993**, *97*, 7351–7358.
- (32) Ge, P.; Luo, G.; Luo, Y.; Huang, W.; Xie, H.; Chen, J. A molecular-scale study on the hydration of sulfuric acid-amide complexes and the atmospheric implication. *Chemosphere* **2018**, *213*, 453–462.
- (33) Myllys, N.; Chee, S.; Olenius, T.; Lawler, M.; Smith, J. N. Molecular-level understanding of synergistic effects in sulfuric acid-amide-ammonia mixed clusters. *J. Phys. Chem. A* **2019**, *123*, 2420–2425.

- (34) Almeida, J.; Schobesberger, S.; Kürten, A.; Ortega, I. K.; Kupiainen-Määttä, O.; Praplan, A. P.; Adamov, A.; Amorim, A.; Bianchi, F.; Breitenlechner, M.; et al. Molecular understanding of sulphuric acid-amine particle nucleation in the atmosphere. *Nature* **2013**, *502*, 359–363.
- (35) Bustos, D. J.; Temelso, B.; Shields, G. C. Hydration of the sulfuric acid-methylamine complex and implications for aerosol formation. *J. Phys. Chem. A* **2014**, *118*, 7430–7441.
- (36) Henschel, H.; Navarro, J. C. A.; Yli-Juuti, T.; Kupiainen-Määttä, O.; Olenius, T.; Ortega, I. K.; Clegg, S. L.; Kurtén, T.; Riipinen, I.; Vehkamäki, H. Hydration of atmospherically relevant molecular clusters: computational chemistry and classical thermodynamics. *J. Phys. Chem. A* **2014**, *118*, 2599–2611.
- (37) Nie, B.; Wang, J.; Qu, B.; Sun, L.; Yan, S. The growth mechanism of sulfuric acid clusters: Implication for the formation of cloud condensation nuclei. *J. Aerosol Sci.* **2017**, *114*, 169–179.
- (38) Loukonen, V.; Kurtén, T.; Ortega, I. K.; Vehkamäki, H.; Pádua, A. A. H.; Sellegri, K.; Kulmala, M. Enhancing effect of dimethylamine in sulfuric acid nucleation in the presence of water - a computational study. *Atmos. Chem. Phys.* **2010**, *10*, 4961–4974.
- (39) Myllys, N.; Elm, J.; Halonen, R.; Kurtén, T.; Vehkamäki, H. Coupled Cluster Evaluation of the Stability of Atmospheric Acid-Base Clusters with up to 10 Molecules. *J Phys Chem A* **2016**, *120*, 621–630.
- (40) Elm, J. An atmospheric cluster database consisting of sulfuric acid, bases, organics, and water. *ACS Omega* **2019**, *4*, 10965–10974.
- (41) Wang, C.-Y.; Ma, Y.; Chen, J.; Jiang, S.; Liu, Y.-R.; Wen, H.; Feng, Y.-J.; Hong, Y.; Huang, T.; Huang, W. Bidirectional Interaction of Alanine with Sulfuric Acid in the Presence of Water and the Atmospheric Implication. *J. Phys. Chem. A* **2016**, *120*, 2357–2371.
- (42) Ge, P.; Luo, G.; Luo, Y.; Huang, W.; Xie, H.; Chen, J.; Qu, J. Molecular understanding of the interaction of amino acids with sulfuric acid in the presence of water and the atmospheric implication. *Chemosphere* **2018**, *210*, 215–223.
- (43) Herb, J.; Nadykto, A. B.; Yu, F. Large ternary hydrogen-bonded pre-nucleation clusters in the Earth's atmosphere. *Chem. Phys. Lett.* **2011**, *518*, 7–14.
- (44) Day, M. B.; Kirschner, K. N.; Shields, G. C. Global search for minimum energy $(\text{H}_2\text{O})_n$ clusters, $n = 3$ -5. *J. Phys. Chem. A* **2005**, *109*, 6773–6778.
- (45) Shields, G. C.; Temelso, B.; Archer, K. A.; Morrell, T. E.; Shields, G. C. Accurate predictions of water cluster formation, $(\text{H}_2\text{O})_{n=2-10}$. *J. Phys. Chem. A* **2010**, *114*, 11725–11737.
- (46) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.
- (47) Temelso, B.; Shields, G. C. The role of anharmonicity in hydrogen-bonded systems: The case of water clusters. *J. Chem. Theory Comput.* **2011**, *7*, 2804–2817.
- (48) Von Freyberg, B.; Braun, W. Efficient search for all low energy conformations of polypeptides by Monte Carlo methods. *J. Comput. Chem.* **1991**, *12*, 1065–1076.
- (49) Rakshit, A.; Yamaguchi, T.; Asada, T.; Bandyopadhyay, P. Understanding the structure and hydrogen bonding network of $(\text{H}_2\text{O})_{32}$ and $(\text{H}_2\text{O})_{33}$: An improved Monte Carlo temperature basin paving (MCTBP) method of quantum theory of atoms in molecules (QTAIM) analysis. *RSC Adv.* **2017**, *7*, 18401–18417.
- (50) Steber, A. L.; Perez, C.; Temelso, B.; Shields, G. C.; Rijis, A. M.; Pate, B. H.; Kisiel, Z.; Schnell, M. Capturing the Elusive Water Trimer from the Stepwise Growth of Water on the Surface of the Polycyclic Aromatic Hydrocarbon Acenaphthene. *J. Phys. Chem. Lett.* **2017**, *8*, 5744–5750.
- (51) Temelso, B.; Morrison, E. F.; Speer, D. L.; Cao, B. C.; Appiah-Padi, N.; Kim, G.; Shields, G. C. Effect of mixing ammonia and alkylamines on sulfate aerosol formation. *J. Phys. Chem. A* **2018**, *122*, 1612–1622.
- (52) Odbadrakh, T. T.; Gale, A. G.; Ball, B. T.; Temelso, B.; Shields, G. C. Computation of atmospheric concentrations of molecular clusters from *ab initio* thermochemistry. *J. Visualized Exp.* **2020**, *158*, No. e60964.
- (53) Vilhelmsen, L. B.; Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *J Chem Phys* **2014**, *141*, No. 044711.
- (54) Dieterich, J. M.; Hartke, B. OGOLEM: Global cluster structure optimisation for arbitrary mixtures of flexible molecules. A multi-scaling, object-oriented approach. *Mol. Phys.* **2010**, *108*, 279–291.
- (55) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.
- (56) Stewart, J. J. P. *MOPAC2012 Computational Chemistry*, 2012.
- (57) Temelso, B.; Mabey, J. M.; Kubota, T.; Appiah-Padi, N.; Shields, G. C. ArbAlign: A Tool for Optimal Alignment of Arbitrarily Ordered Isomers Using the Kuhn-Munkres Algorithm. *J. Chem. Inf. Model.* **2017**, *57*, 1045–1054.
- (58) Burke, K.; Perdew, J. P.; Wang, Y. Derivation of a generalized gradient approximation: The PW91 density functional. In *Electronic Density Functional Theory*; Dobson, J. F.; Vignale, G.; Das, M. P., Eds. Springer: Boston, MA, USA, 1998.
- (59) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C., et al. *GAUSSIAN 09 (Revision A.02)*, Gaussian, Inc.: Wellington, CT, 2009.
- (61) Elm, J.; Bilde, M.; Mikkelsen, K. V. Assessment of binding energies of atmospherically relevant clusters. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16442–16445.
- (62) Elm, J.; Bilde, M.; Mikkelsen, K. V. Assessment of density functional theory in predicting structures and free energies of reaction of atmospheric prenucleation clusters. *J. Chem. Theory Comput.* **2012**, *8*, 2071–2077.
- (63) Van Dorshuld, E.; Vergenz, R. A.; Tschumper, G. S. Peptide bond formation via glycine condensation in the gas phase. *J. Phys. Chem. B* **2014**, *118*, 8583–8590.
- (64) Fukui, K. The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* **2002**, *14*, 363–368.
- (65) Dykstra, C. E. *Theory and applications of computational chemistry: the first forty years*; 1 ed.; Elsevier: Boston, MA, 2005.
- (66) Irikura, K. K. Computational thermochemistry: Prediction and estimation of molecular thermodynamics. *ACS Symp. Ser.* **1998**, *677*, 285.
- (67) Irikura, K. K. *THERMO.PL*; NIST: 2002.
- (68) Morrell, T. E.; Shields, G. C. Atmospheric Implications for Formation of Clusters of Ammonium and 1-10 Water Molecules. *J. Phys. Chem. A* **2010**, *114*, 4266–4271.
- (69) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (70) Perez de Tudela, R.; Marx, D. Water-Induced Zwitterionization of Glycine: Stabilization Mechanism and Spectral Signatures. *J. Phys. Chem. Lett.* **2016**, *7*, 5137–5142.
- (71) Day, M. B.; Kirschner, K. N.; Shields, G. C. Pople's Gaussian-3 Model Chemistry Applied to an Investigation of $(\text{H}_2\text{O})_8$ Water Clusters. *Int. J. Quantum Chem.* **2005**, *102*, 565–572.
- (72) Xantheas, S. S. Cooperativity and hydrogen bonding network in water clusters. *Chem. Phys.* **2000**, *258*, 225–231.
- (73) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (74) Seinfeld, J. H.; Pandis, S. N. *Atmos. Chem. Phys.: From Air Pollution to Climate Change*; 3rd ed.; John Wiley & Sons: Hoboken, NJ, 2016.

- (75) Dunn, M. E.; Pokon, E. K.; Shields, G. C. Thermodynamics of forming water clusters at various temperatures and pressures by Gaussian-2, Gaussian-3, Complete Basis Set-QM3, and Complete Basis Set-APNO model chemistries; implications for atmospheric chemistry. *J. Am. Chem. Soc.* **2004**, *126*, 2647–2653.
- (76) Samala, N. R.; Agmon, N. Thermally induced hydrogen-bond rearrangements in small water clusters and the persistent water tetramer. *ACS Omega* **2019**, *4*, 22581–22590.
- (77) Peixoto, J. P.; Oort, A. H. The climatology of relative humidity in the atmosphere. *J. Clim.* **1996**, *9*, 3443–3463.
- (78) Ruzmaikin, A.; Aumann, H. H.; Manning, E. M. Relative humidity in the troposphere with AIRS. *J. Atmos. Sci.* **2014**, *71*, 2516–2533.
- (79) Willett, K. M.; Dunn, R. J. H.; Thorne, P. W.; Bell, S.; de Podesta, M.; Parker, D. E.; Jones, P. D.; Williams, C. N., Jr. HadISDH land surface multi-variable humidity and temperature record for climate monitoring. *Clim. Past* **2014**, *10*, 1983–2006.
- (80) Zhang, Q.; Anastasio, C. Free and combined amino compounds in atmospheric fine particles (PM_{2.5}) and fog waters from Northern California. *Atmos. Environ.* **2003**, *37*, 2247–2258.
- (81) Mandalakis, M.; Apostolaki, M.; Stephanou, E. G. Trace analysis of free and combined amino acids in atmospheric aerosols by gas chromatography-mass spectrometry. *J. Chromatogr. A* **2010**, *1217*, 143–150.
- (82) Borsook, H. Peptide bond formation. *Adv. Protein Chem.* **1953**, *8*, 127–174.
- (83) Parker, E. T.; Zhou, M.; Burton, A. S.; Glavin, D. P.; Dworking, J. P.; Krishnamurthy, R.; Fernández, F. M.; Bada, J. L. A plausible simultaneous synthesis of amino acids and simple peptides on the primordial Earth. *Angew. Chem., Int. Ed.* **2014**, *53*, 8132–8136.
- (84) Shute, R. E.; Rich, D. H. Prevention of diketopiperazine formation in peptide synthesis by a simultaneous deprotection–coupling procedure: entrapment of reactive nucleophilic species by in situ acylation. *J. Chem. Soc., Chem. Commun.* **1987**, 1155–1156.
- (85) Gale, A. G.; Odbadrak, T. T.; Shields, G. C. Transition State Structures of Hydrated Glycine Clusters. *Int. J. Quantum Chem.* **2020**, , submitted, in revision