PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

Cite this: *Phys. Chem. Chem. Phys.*, 2023, **25**, 28517

Prebiotic dimer and trimer peptide formation in gas-phase atmospheric nanoclusters of water†

Shannon E. Harold, Skyler L. Warf and George C. Shields *

Received 22nd June 2023,
Accepted 26th September 2023

DOI: 10.1039/d3cp02915h

rsc.li/pccp

Insight into the origin of prebiotic molecules is key to our understanding of how living systems evolved into the complex network of biological processes on Earth. By modelling diglycine and triglycine peptide formation in the prebiotic atmosphere, we provide a plausible pathway for peptide growth. By examining different transition states (TSs), we conclude that the formation of diglycine and triglycine in atmospheric nanoclusters of water in the prebiotic atmosphere kinetically favors peptide growth by an N-to-C synthesis of glycines through a *trans* conformation. Addition of water stabilizes the TS structures and lowers the Gibbs free activation energies. At temperatures that model the prebiotic atmosphere, the free energies of activation with a six water nanocluster as part of the TS are predicted to be 16 kcal mol^{-1} relative to the prereactive complex. Examination of the *trans vs. cis* six water transition states reveals that a homodromic water network that maximizes the acceptor/donor nature of the six waters is responsible for enhanced kinetic favorability of the *trans* N-to-C pathway. Compared to the non-hydrated *trans* TS, the *trans* six-water TS accelerates the reaction of diglycine and glycine to form triglycine by 13 orders of magnitude at 217 K. Nature uses the *trans* N-to-C pathway to synthesize proteins in the ribosome, and we note the similarities in hydrogen bond stabilization between the transition state for peptide synthesis in the ribosome and the transition states formed in nanoclusters of water in the same pathway. These results support the hypothesis that small oligomers formed in the prebiotic atmosphere and rained onto earth's surface.

Introduction

Life most likely arose on Earth 4.2 to 4.5 billion years ago,[1] sometime after the last ocean vaporizing asteroid impact.[2] One of the most puzzling questions for scientists in the last several decades revolves around the formation of the prebiotic molecules essential to the central dogma of molecular biology; DNA is transcribed into RNA and then translated into proteins for genetic expression. These three processes are complex and involve interplay between nucleic acids, peptides, and many other molecules.[3] While the building blocks of proteins, DNA and RNA, were likely synthesized in outer space on metal rich meteroites[4] and deposited on Earth from meteoric impacts,[5] how the larger oligomers formed from their precursors is an area of active research. A prominent theory regarding prebiotic macromolecule formation is the RNA world hypothesis which asserts that RNA proceeded DNA as the first genetic material.[6–8] This hypothesis is supported by the fact that RNA is self-replicating and can behave as an enzyme. Accordingly, there must have been an RNA strand that developed with the capacity to replicate itself. Over millions of replications, variations in this original strand could have led to the formation of other macromolecules like DNA and proteins. Watson–Crick base pairing is essential to this theory as well as the notion that genetically encoded proteins were unnecessary for primordial replication.[9] More recent work has begun to question this hypothesis because of the high prevalence of cross-over between classes of molecules like nucleic acids and proteins in extant biology; it seems unlikely that one of these classes could have evolved without the other. Or perhaps the primitive ribosome started out as RNA, and evolution drove the process of incorporating proteins. In processes such as ribosomal assembly and function, transcription, and polymerase-catalyzed DNA replication, there are direct interactions between proteins, RNA, and DNA.[3] If peptides existed before the replication machinery, what mechanism could have formed them? Many other possibilities have been investigated.[10,11] Thioesters may provide a plausible path for proto-peptides on the early Earth.[12] Liquid sulfur dioxide activates amino acids resulting in peptide growth.[13] Amino acids have been shown to spontaneously oligomerize without the use of enzymes or activating agents, under mild, hydroxy acid-catalyzed, dry-down

Department of Chemistry, Furman University, Greenville, South Carolina 29613, USA. E-mail: george.shields@furman.edu

† Electronic supplementary information (ESI) available: Coordinates of all optimized structures, energetics, and derivation of CBS extrapolation. See DOI: https://doi.org/10.1039/d3cp02915h

conditions.[14] In contrast to all of these solution-based methods, we aim to investigate whether prebiotic peptide formation could have occurred in the gas phase, in nanoclusters of water.

One of the primary challenges faced by prebiotic chemists is understanding how peptides formed despite the substantial thermodynamic and kinetic barriers of the oligomerization of amino acids in aqueous solution. In this context, the dimerization of amino acids is thermodynamically unfavorable because a water molecule is released. Additionally, when free amino acids are in aqueous environments, they exist in their zwitterionic form which prevents amide bond formation.[15] Formation of the zwitterion of amino acids consists of protonation of the amino group and deprotonation of the carboxyl group. When glycine is in its zwitterionic form, there is no longer a lone pair of electrons available for nucleophilic attack on the carbonyl carbon of the neighboring electron. This problem is circumvented in biological organisms due to the role of enzymes.[5] Glycine and other amino acids exist in their canonical (neutral) form in the gas phase when one to four waters are hydrogen bonded to them.[16] Thus, it is plausible that the gas-phase environment was hospitable for peptide bond formation in the prebiotic atmosphere.

Amino acids such as glycine, alanine, proline, valine, glutamine, asparagine, and leucine have been found in carbonaceous chondrite meteorites.[5] These meteorite fragments come from asteroids that are thought to have remained relatively unchanged since the formation of the solar system. If amino acids or dipeptides of amino acids were formed and then incorporated into meteorites, they could have been delivered to earth *via* comets or meteorites to act as biological catalysts. In interstellar environments, it has been proposed that cosmic radiation or gas phase collisions could lead to the excitation of protonated glycine dimers. These protonated dipeptides could further elongate into tripeptide chains, as verified by mass spectrometry.[17] Interstellar ices have been studied for their potential to contribute organic material to meteorites and comets. Dipeptide formation has also been shown to occur in interstellar ice models subjected to electron irradiation mimicking galactic cosmic rays,[18] and small peptides are synthesized abiotically in aqueous microdroplets.[19]

Another potential environment is the air–water interface of microdroplets of water.[20] An experimental study on nucleotide formation examined a potential method to overcome the thermodynamic barrier of the phosphorylation of ribose.[21,22] It was proposed that microdroplets would be able to accelerate this and similar reactions because they have properties that allow them to orient molecules in a more favorable way at the air–water interface. They found that when they put ribose and phosphoric acid in a microdroplet environment, the experimental value of ΔG decreased from 5.4 kcal mol^{-1} in bulk solution to -1.1 kcal mol^{-1}. Additionally, it has been shown that uridine, adenosine, inosine, and cytidine could be generated in these microdroplets. Similar experiments using mass spectrometry have shown that dipeptide bond formation can occur under ambient conditions with Mg^{2+} as a catalyst in aqueous microdroplets.[23] Free amino acids also react to form small peptides composed of alanine and glycine, with no need for metal catalysts.[19] Thermally accessible prebiotic pathways for the formation of ribonucleic acid and protein precursors from aqueous hydrogen cyanide has recently been explored using a computational network methodology, revealing a diverse reactivity landscape.[24] The reaction pathways provide lower activation energies than previous proposed pathways, and water can catalyze these reactions. QM/MM MD simulations revealed that fluctuations of the electrostatic potential results from the dynamics of the water solvation shells, thus leading to enhanced rate acceleration at the air–water interface.[25]

Evidence that prebiotic molecule formation spontaneously occurs in microdroplets has given rise to research in larger aerosols known to contain up to 50% organic material.[26] Atmospheric aerosols are suspensions of liquid and solid particles in the air that come from natural and anthropogenic sources such as fires, vehicle exhaust, desert dust, and sea spray. They scatter and absorb sunlight, thus playing an important role in the temperature control of earth. Furthermore, atmospheric aerosols provide a unique environment for chemical reactions because repeated wet-dry cycles can function to remove water and allow organic molecules to concentrate at the air–water interface. Models of aerosols from mass spectroscopy reveal that aerosols may exist as inverted micelles, with most of the organic content being at the surface.[27] This kind of structure can allow atmospheric aerosols to align molecules in such a way that reactions become more favorable with lower activation energy barriers. Atmospheric research on Titan, Saturn's largest moon, has provided evidence for the plausibility of aerosols as an environment for the formation of prebiotic molecules. Experimentalists reproduced tholins, or aerosol analogs, and created an environment aligning with that of Titan's atmosphere.[28] The results showed that all five nucleobases as well as alanine and glycine were formed, revealing that prebiotic molecules could form in Titan's atmosphere even in the absence of liquid water. We propose a similar process may have occurred in the reducing upper atmosphere of early Earth, with prebiotic molecules formed through gas phase interactions of amino acids and small nanoclusters of water.

The prebiotic atmosphere was reducing, since no oxygen was present, and cold, as a result of weathering of ejecta from intense meteorite bombardment.[29] Ultraviolet radiation could have generated simple sugars necessary for prebiotic synthesis,[30] and epochs of moderately high volcanism would have shielded the atmosphere and been conducive to cyanosulfidic prebiotic chemistry.[31,32] The surface temperature has been modeled, and simulations predict a 70% probability of a cold surface environment below the freezing point of water based on the expected concentration of CO_2 in the atmosphere 4.3 billion years ago, with an average surface temperature of 259.2 K.[29] In addition, lakes and oceans are predicted to be mildly alkaline,[33] making them more hospitable to life. For this reason, cold lakes have been proposed as sites for the origin of life, as they would overcome the concentration problem, which is that reactants must be sufficiently concentrated to drive reactions.[34] We note that the favorable formation of

polypeptides in the Earth's atmosphere could have been a source for the accumulation of peptides in bodies of water from rainfall.

In contrast to previous work that has explored how biopolymers are synthesized in aerosols or liquids, our study examines the reaction of peptide bond formation on the nanoscale in the gas phase as a function of the number of individual water molecules (n = 0–9) forming favorable hydrogen bonding networks. Specifically, we explore the hypothesis that amino acids present in nanoclusters of water molecules in the gas phase form small peptides through thermodynamically and kinetically favorable reactions. By studying peptide formation in nanoclusters of water molecules we gain insight into the mechanism by which peptide formation could occur in the gas phase.

Methodology

To obtain activation barriers for the trimerization of glycine in water clusters, potential transition state structures were built using the Avogadro graphical user interface.[35] The *cis* diglycine TS was used from our previous work[36] and the *trans* diglycine TS was calculated in this work. Waters were added to the diglycine TS structures to examine the impact of water nanocluster hydration, and to compare the free energies of activation for the diglycine system to the triglycine system. To create starting structures for the N-to-C triglycine TSs, we added a third glycine to the carboxyl end of the diglycine TS. To create a starting structure for the C-to-N triglycine TS, we added glycine to the amine end of diglycine to obtain an initial structure for the N-to-C triglycine TS. Waters were added systematically to all possible hydrogen bonding sites for each triglycine-$(H_2O)_n$ TS in order to create input structures for the next TS, triglycine-$(H_2O)_{n+1}$. In this way, all possible *cis* TS structures were built for n = 0–6 waters and all possible *trans* TS structures for n = 0–9. From these initial models, constrained geometry optimizations were carried out with the transition state coordinates frozen, so that the waters were allowed to move freely about the TS structure. These constrained optimizations were followed by frequency calculations to ensure that the TS mode was located, followed by unconstrained TS geometry optimizations. A total of 102 input structures were created and optimized using the ωB97X-D/6-31++G(d,p) model chemistry[37–41] in Gaussian 16[42] to locate the 29 lowest free energy TS states (5 *cis*-C-to-N, 7 *cis*-N-to-C, 8 *trans*-C-to-N, and 9 *trans*-N-to-C). All transition states were verified by intrinsic reaction coordinate (IRC)[43] calculations, to ensure that each TS was connected to the prereactive and postreactive complexes. The resulting IRC endpoint geometries were optimized to confirm that the reactant and product clusters are indeed hydrated reactants and products. For every DFT structure, the harmonic vibrational frequencies were scaled by 0.971 to partially correct for anharmonicity and used to calculate finite temperature corrections with the thermo.pl script from the National Institute of Standards and Technology (NIST).[44] The Density Functional Theory (DFT) electronic energies were corrected with the domain-based local pair natural orbital coupled-cluster method with single, double, and perturbative-triple excitations (DLPNO-CCSD(T))[45–48] using ORCA 5.0.1.[49–51] The DLPNO-CCSD(T)/cc-pVnZ model chemistry electronic energies using the Dunning double, triple, and quadruple-zeta basis sets (n = D, T, Q)[52] were extrapolated using the 4–5 polynomial approximation[53] to the complete basis set (CBS) limit and then combined with the Gibbs free energy corrections from the DFT calculations to determine the final reaction ΔG° and TS $\Delta G^{\circ\#}$ values. Extrapolation to the CBS limit eliminates the need to correct for the Basis Set Superposition Error that occurs with small basis sets.[54]

Transition state free energies of activation were computed by subtracting the Gibbs free energies of the separated reactants from the Gibbs free energies of the transition states, and by subtracting the Gibbs free energies of the prereactive complexes from the Gibbs free energies of the transition states. Energies are reported to two decimal places in the tables to avoid rounding errors, although we estimate that actual uncertainties in the calculations are many tenths of a kcal mol^{-1}. The complete summarized methodology is outlined in Fig. 1.

Results and discussion

Gas phase reactions in the atmosphere

Gas phase reactions in the atmosphere are dominated by two body collisions, where two reactants collide to form the product, and then the product is stabilized by a second collision with an inert molecule (such as N_2) to carry off enough energy so that the product doesn't immediately fall apart. Eqn (1)–(3) illustrate this general concept, where the reactants A and B form C*, and then M carries off some of the energy from C*, resulting in the overall reaction (3).

$$A + B \rightarrow C^* \quad (1)$$

$$C^* + M \rightarrow C + M^* \quad (2)$$

$$A + B \rightarrow C \quad (3)$$

For the system we are considering, where glycines and waters collide to produce small polymers, step-wise reactions need to

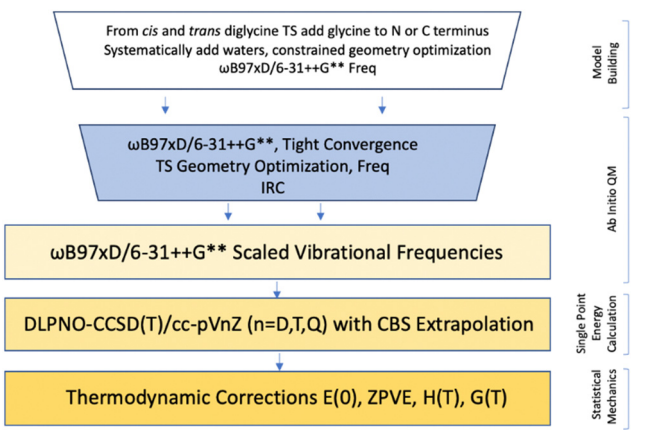


Fig. 1 Computational methodology used to determine the structure and thermodynamics of oligopeptides in gaseous water nanoclusters.

be considered, where we have omitted the collision with M, such as the unhydrated reactions,

$$\mathrm{Gly} + \mathrm{Gly} \rightarrow \mathrm{diGly} \quad (4)$$

$$\mathrm{diGly} + \mathrm{Gly} \rightarrow \mathrm{triGly} \quad (5)$$

and hydrated reactions such as:

$$\mathrm{Gly} + \mathrm{H_2O} \rightarrow \mathrm{Gly(H_2O)} \quad (6)$$

$$\mathrm{Gly(H_2O)_5} + \mathrm{Gly(H_2O)} \rightarrow \mathrm{diGly(H_2O)_6\ TS} \rightarrow \mathrm{diGly(H_2O)_7} \quad (7)$$

Previous work has demonstrated that hydration of water molecules results in negative values for the stepwise formation of pure water clusters from $n = 2$ to $n = 6$ at 273 K, thus predicting that measurable concentrations of $(H_2O)_{n=2-6}$ will be present in the prebiotic atmosphere.[55,56] In addition the $Gly(H_2O)$ concentration (reaction (6)) should also be appreciable at 273 K.[16] Thus we anticipate that reaction (7) would occur in the prebiotic atmosphere given enough time, since thermodynamics will drive this reaction. We will return to reaction (7) later in this paper.

Temperature and phase considerations

It is interesting to consider the differences that temperature has on bulk quantities of water compared to water on the nano scale. While bulk water, including in large aerosol droplets, would freeze at 273 K and 1 atm pressure, nanoclusters of water in the atmosphere might not. Pressure drops as the temperature decreases in the atmosphere, the effects of these conditions are different in the nanoworld than in the macroworld. This is illustrated by low temperature spectroscopy experiments of water nanoclusters. For instance, water clusters produced in a pulsed supersonic expansion have been studied using chirped-pulse Fourier transform microwave spectroscopy. These broadband rotational spectroscopy experiments, where the beam is at a low temperature estimated to be in the 20–30 K range from the supersonic expansion, reveal that there are three different isomers of the water hexamer at these low temperatures.[57] The $(H_2O)_6$ isomers are the Cage, Prism, and Book structures. Waters in liquid water move constantly, swapping hydrogen bonding partners, which are only locked in place when liquid water crystallizes. It isn't clear when a handful of waters should be considered in a phase other than "cluster." But clusters at low temperatures and pressures do exhibit motion. Broadband rotational spectroscopy and quantum simulations have revealed that the Prism hexamer exhibits quantum tunneling. This process breaks two different hydrogen bonds and two molecules in the cluster exhibit coupled motion during this transition.[58] This geared rotation of a pair of water molecules involves the concerted breaking of two hydrogen bonds.

Impact of vibrational anharmonicity on predicted Gibbs free energies

The harmonic oscillator approximation is used to obtain frequencies which are then used with statistical mechanics to convert the quantum mechanical electronic energy of one molecular entity to energies, enthalpies, and entropies for an ensemble of entities. As harmonic frequencies are always higher than the experimental frequencies, an empirical correction is required to account for vibrational anharmonicity.[59–61] As the low frequency modes contribute the most to the entropy, small deviations in these modes can lead to large errors in Gibbs free energies. This is a real problem for scientists studying intermolecular forces with hydrogen bonded structures that contain many low modes.[62] One way to confront this problem is to calculate anharmonic frequencies directly using second-order vibrational perturbation theory (VPT2),[60,61] where anharmonic corrections are calculated from higher-order derivatives of the PES. Two main drawbacks of VPT2 are the time it takes to carry out the calculation on larger hydrogen-bonded clusters, and the near-degeneracies (also known as resonances), that plaque perturbation methods. Because of the importance of calculating accurate ΔG values for cluster formation, much work has been done to assess the ability of various methods to compute reliable ΔG values. The water dimer is the only water cluster where the complete experimental vibrational spectrum is known. Work on water clusters has allowed for some conclusions to be drawn on the accuracy of using VPT2 anharmonic or scaled harmonic frequencies for these hydrogen-bonded systems. Fortunately, scaling Hartree–Fock (HF) harmonic frequencies with the 6-31G* basis set by 0.8928 results in good agreement with the experimental system, with a root-mean-square deviation (RMSD) of 24 cm^{-1}.[63,64] The scaling is surprisingly effective. Combining scaled harmonic frequencies for the intramolecular modes with VPT2 anharmonic frequencies for the intermolecular modes reduces the RMSD to 20 cm^{-1}. Larger water clusters, $(H_2O)_{3-9}$, have been thoroughly investigated with the frozen-core resolution-of-the-identity Moller–Plesset (RI-MP2) method to understand the differences between harmonic, scaled harmonic, and VPT2 anharmonic frequencies.[65] By separating water clusters into three distinct regions, intermolecular modes below 1100 cm^{-1}, bending modes between 1100 and 1800 cm^{-1}, and stretching modes above 3000 cm^{-1}, three different scaling factors were derived and the scaled frequencies were compared to the anharmonic frequencies.[56,65] Using VPT2 or a three-split scaling model gave the most accurate results, a single scaling factor that worked best for the intermolecular modes also was in good agreement, as the bending and stretching modes are less anharmonic relative to the intermolecular modes.[56,65,66] The experimental ΔG° for the formation of the $H_2SO_4(H_2O)$ dimer and the $H_2SO_4(H_2O)_2$ trimer at 298 K are -3.6 ± 1 and -5.9 ± 1 kcal mol^{-1}.[67] The same methodology used for the water cluster results applied to the sulfuric acid/water system using scaled harmonic frequencies yields ΔG° (298 K) values of -2.9 and -4.1 kcal mol^{-1} for the dimer and trimer respectively. The VPT2 anharmonic frequencies are within 0.1 kcal mol^{-1} of the scaled results, -2.8 and -4.2 kcal mol^{-1} for the dimer and trimer, and in reasonable agreement with the experiments.[68] For the $(H_2SO_4)_2(H_2O)_{n=0-6}$ system, a similar methodology revealed that the mean absolute error for the scaled harmonic frequencies relative to eight anharmonic frequencies was 0.25 kcal mol^{-1}.[69] Much work

has been completed on the appropriate scaling factors for many different model chemistries, and the ωb97X-D scaling factor is 0.971,[70] which we used for all of the structures in this paper. Anharmonic frequency calculations on Gly(H_2O) complexes with the ωb97X-D/6-31++G** model chemistry suffer from large resonances,[71] so we rely on the scaled frequencies for this work. Finally, we note that the largest contributor to ΔG is the electronic energy component, so that the errors in relying on DFT for structures and energies *versus* using a CCSD(T)/CBS//DFT approach are most likely larger than the error in the entropies and thermal energy factors.[72]

Gibbs free energy of reaction

The experimental Gibbs free energy change for the dimerization of glycine in aqueous solution at 298 K is 3.6 kcal mol^{-1}.[73] The zwitterionic form of amino acids inhibits nucleophilic attack on the carboxyl carbon of the subsequent amino acid to form the peptide bond.[74] In the gas phase, highly accurate quantum chemical calculations have revealed that the thermodynamics and kinetics for formation of diglycine from two canonical glycine molecules slightly favors a concerted process leading to a *trans* configuration about the peptide bond.[75] Building on this result, in our previous work we used the PW91/6-311++G** model chemistry to predict that the gas-phase ΔG° for the reaction of two canonical glycines to form diglycine and a water at 298 K is −2.9 kcal mol^{-1}.[16] For the reaction of n waters and two glycines to produce diglycine and n + 1 waters, with the reactants and products separated at an infinite distance, the DFT value for ΔG° will of course remain the same. If one instead calculates ΔG° assuming that the product diglycine is fully hydrated with waters at the end of the reaction, the ΔG° values steadily decrease from −7.2 kcal mol^{-1} to −18 kcal mol^{-1} as n increases from one to seven.[16] Either way, the glycine dimerization reaction is thermodynamically favored. In this work we have updated the value for the gas-phase reaction of n waters and two glycines to form the products diglycine and waters separated at an infinite distance with the DLPNO-CCSD(T)/CBS//ωB97X-D/6-31++G(d,p) model chemistry, and obtained values of 0.3 to 0.6 kcal mol^{-1} over the temperature range of the prebiotic atmosphere (217 to 273 K). For the reaction of diglycine and glycine to form triglycine and water over this temperature range, ΔG° for this reaction encompasses −5.6 to −5.4 kcal mol^{-1}. These results are presented in Table 1. Thus, the thermodynamics for formation of diglycine and triglycine in the gas-phase, when amino acids are not zwitterions, but exist in their canonical forms, favor the formation of the peptide bond. This is in stark contrast to aqueous solution, where ΔG° is positive.[73] We turn now to the kinetics for polymerization of glycine in the gas phase.

Gibbs free energy of activation

Our previous work demonstrated that the formation of *cis* diglycine formation from two glycine monomers in the gas phase is stabilized in the presence of one to five water molecules.[16] The *cis* reaction pathway of diglycine formation is enhanced at low temperatures for the addition of up to four waters, while the effect is lessened due to the entropic penalty of hydrogen bond formation at higher temperatures.[36] In this work we calculated the *trans* diglycine pathway. The non-hydrated *cis* and *trans* diglycine transition states are compared in Fig. 2. We start by calculating the Gibbs free energy activation barrier from the infinitely separated reactants, and we will discuss pathways along the PES in the next section. The *trans* diglycine TS has a free energy of activation of 42.6 kcal mol^{-1} at 217 K (45.2 kcal mol^{-1} at 273 K) while the *cis* diglycine TS has a free energy of activation of 49.5 kcal mol^{-1} at 217 K (53.1 kcal mol^{-1} at 273 K). The *trans* pathway for diglycine formation is the lowest free energy pathway with addition of waters.

The *trans* diglycine and triglycine transition state free energies of activation as a function of water nanocluster size are compared in Table 2. The addition of a third glycine leads to a reduction in $\Delta G^\ddagger$, suggesting that a peptide chain in gas-phase nanoclusters of water may grow at a lower energetic cost with the successive addition of amino acids. Based on this comparison, we examined the four different pathways leading to triglycine formation as well as the catalytic effect of water. All pathways involve the cleavage of the C–OH bond in the active carboxyl group and the N–H bond in the active amine group. The deprotonated amine forms the C–N amide bond, while the proton from the amine group joins with the hydroxyl group to form a water molecule. Biologically, the ribosome forms peptide bonds from the N-terminus of the monomer to the C-terminus of the growing peptide chain[76]

We gain insights into the evolution of this process by comparing the energetics for forming one triglycine from the N-terminus of a glycine monomer to the C-terminus of the diglycine dimer to the energetics for forming a triglycine from the C-terminus of the glycine monomer to the N-terminus of the diglycine dimer.

The mechanistic difference between these two pathways is illustrated in Fig. 3. In addition to the N-to-C directionality of peptide growth, *trans* amino acid structures are predominantly found in nature. This is a result of the strain caused by the R-group and peptide carbon-nitrogen chain interactions. To study these energetics, we include both the *cis* and *trans* triglycine transition states when finding the lowest energy pathway for triglycine to form prebiotically.

Fig. 4 shows the four non-hydrated triglycine transition states, *trans* N-to-C, *trans* C-to-N, *cis* N-to-C, and *cis* C-to-N. As shown in Table 3, the structure with the lowest $\Delta G^{\circ\#}$ at all three

Table 1 Gibbs free energies (kcal mol^{-1}) for the reactions of glycine and diglycine to form triglycine at temperatures spanning the prebiotic atmosphere

Reaction	217 K	259 K	273 K
Diglycine + glycine + $(H_2O)_n$ → triglycine + $(H_2O)_{n+1}$	−5.59	−5.41	−5.36

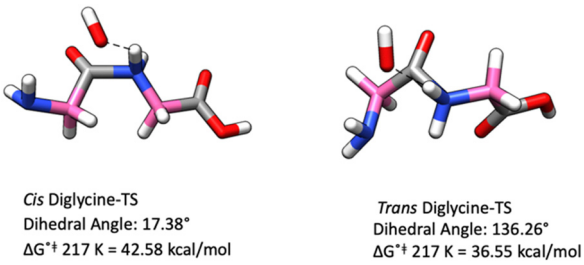


Fig. 2 Comparison of non-hydrated diglycine *cis* and *trans* transition states for the reaction of two glycines to form diglycine and a water. The dihedral angles are measured between the alpha carbons highlighted in pink. The *trans* TS is 6 kcal mol^{-1} more negative than the *cis* TS, so that it is six orders of magnitude faster at 217 K.

Table 2 Comparison of DLPNO-CCSD(T)/CBS//ωB97X-D/6-31++G** $\Delta G^{\circ\#}$ values in kcal mol^{-1} for the hydration of a diglycine transition state and triglycine transition state with 0–5 waters using ωB97X-D DFT at 217 K. The TS $\Delta G^{\circ\#}$ values are calculated relative to the separated reactants, *n* H_2O molecules and a glycine to form the *trans*-diglycine TS, and *n* H_2O molecules, glycine, and diglycine to form the *trans*-triglycine T

Number waters	*trans*-Diglycine TS	*trans*-Triglycine TS
0	42.58	36.55
1	39.87	35.28
2	35.19	34.34
3	31.70	29.15
4	30.60	28.38
5	29.36	27.42

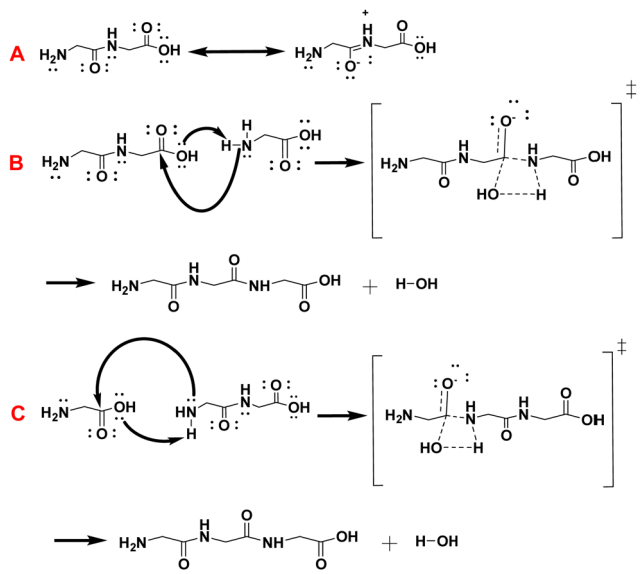


Fig. 3 Section A shows the relevant resonance structures of diglycine. Section B shows the mechanism, transition state and products of N-to-C terminus triglycine formation. Section C shows the mechanism, transition state and products of C-to-N triglycine formation.

temperatures is the *trans* TS formed N-to-C. Interestingly, these results align with the way in which peptide chains have evolved

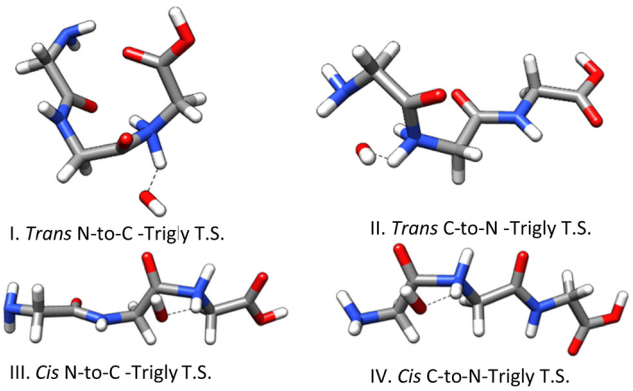


Fig. 4 The structures of non-hydrated transition states for the reaction of diglycine with glycine on the reaction path to form triglycine. The $\Delta G^{\circ\#}$ values are presented in Table 3.

in biology and indicate kinetic favorability for this pathway relative to the other three. As shown in Fig. 4, *trans* transition state structures have greater bending in the peptide backbones while the *cis* structures remain more linear. Formation of the triglycine from the N-terminus of glycine to the C-terminus of diglycine has a more severe bend (Fig. 4I) because of increased van der Waals forces, making it the most thermodynamically favorable pathway. From the free energies of activation in Table 3, the non-hydrated *trans* N-to-C TS has the lowest activation energy and is the most favorable pathway to form triglycine. Below, we compare the *cis* N-to-C and *trans* N-to-C pathways to analyze how the two pathways compare when complexed with nanoclusters of water (n = 0–9).

Catalytic role of water

To examine the role of the explicit addition of water molecules to the reaction pathway of triglycine formation in the gas-phase, waters were sequentially added to each TS. The results for the other pathways can be found in the ESI† (Fig. S1–S3; energy Tables). Fig. 5 shows the sequential addition of water for the *trans* triglycine N-to-C terminus transition state, which was the lowest energy non-hydrated TS, while Table 4 contains the $\Delta G^{\circ\#}$ values computed from the TS and the infinitely separated reactants.

Examining the sequentially hydrated transition states in Fig. 5, a fully developed water spine of hydration is formed for the structure hydrated with six waters, I-6W. With the sequential addition of waters, a hydration ring/spine motif forms and expands to envelope the peptide backbone which maximizes the amount of hydrogen bonding on the peptide backbone surface. At the two lower temperatures, each addition of water provides stabilizing hydrogen bonds that reduce the energy barrier of the reaction, as shown in Table 4. *trans*-Triglycine-N-to-C hydrated TSs with 7–9 waters are the lowest energy transition states at 217 K. At 259 K the best stabilized TS has nine water molecules, and at 273 K the best stabilization is found for 3, 6, and 7 waters. The results at 273 K can be attributed to the entropic cost ($T\Delta S$) of adding more waters at a higher temperature. In all of these nanoclusters, the stabilizing

Table 3 The DLPNO-CCSD(T)/CBS//wb97X-D/6-31++G** $\Delta G^{\circ\#}$ values in kcal mol^{-1} for four different non-hydrated transition states for the reaction of diglycine with glycine on the reaction path to form triglycine at temperatures relative to the prebiotic atmosphere. The TS $\Delta G^{\circ\#}$ values are calculated relative to the separated reactants

TS conformation	TS number	217 K	259 K	273 K
trans-N-to-C	1	36.55	38.72	39.43
trans-C-to-N	II	40.25	42.17	42.80
cis-N-to-C	III	44.88	46.70	47.29
cis-C-to-N	IV	47.22	48.42	49.87

effect of the water molecules is a result of favorable hydrogen bonding networks involving the carboxyl and amino functional groups within the peptide as well as with the leaving hydroxyl group. From the energies reported above, the six water TS structure is most notable. The I-6W TS has an energy reduction of 2–4 kcal mol^{-1} relative to the I-5W TS, which corresponds to a reaction rate increase of roughly two to four orders of magnitude over the temperature range of the prebiotic atmosphere. As shown in Fig. 5, it is striking that six waters complete the spine from one end of the C-terminus hydroxyl group of the incoming glycine to the OH that is being abstracted from the carbon during the formation of the peptide bond. The seventh water is inserted into the water spine, elongating it. The eighth and ninth waters are attached to the water spine, and do not interact with the peptide. For I-9W, the classic water dimer[55,77] is attached to the outside edge of the water network. In the examination of the sequential hydration of the *cis* pathway, more localized water networks form compared to the *trans* pathway. These waters do not interact with all of the hydrogen bonding sites on the peptide chain or envelope the backbone as in the *trans* pathway. This results in the *trans* pathway being kinetically favored over the *cis* pathway. In response to hydration, the original *cis* transition states sequentially rotate into a *trans*-like conformation with each addition of water. From the energies associated with the *cis* transition states, the first three

Table 4 $\Delta G^{\circ\#}$ values in kcal mol^{-1} for the *trans* N-to-C TS structures as waters are added for the reaction of diglycine with glycine on the reaction path to form triglycine at three temperatures spanning the prebiotic troposphere. The TS $\Delta G^{\circ\#}$ values are calculated relative to the separated reactants

	Temperature		
TS structure	217 K	259 K	273 K
I	36.55	38.72	40.19
I-1W	35.28	38.67	39.78
I-2W	31.35	36.38	38.06
I-3W	29.15	35.44	37.52
I-4W	28.38	35.98	38.51
I-5W	27.42	36.53	39.55
I-6W	23.25	34.04	37.63
I-7W	21.75	33.72	37.72
I-8W	21.83	35.05	49.46
I-9W	21.46	32.11	41.93

structures (n = 1–3) are most impactful. These structures have an energy reduction of $\sim$5 kcal mol^{-1} which corresponds to a reaction rate increase of roughly four to five orders of magnitude over this temperature range. These figures are shown in the SI along with their energetics.

Reaction pathway for formation of the I-6W TS

The I-6W TS would not form from eight separated reactants, but rather from two-body collisions as in eqn (1)–(3), so we have computed and plotted the PES for formation of this pathway. To determine the prereactive complex involved in the formation of the I-6W TS, we followed the reaction path by integrating the intrinsic reaction coordinate (IRC)[78–80] from the TS towards both reactants and products. The IRC pathway toward the reactants revealed that the reactant clusters consist of Digly$(H_2O)_5$ and Gly(H_2O) enroute to making the Trigly$(H_2O)_7$ product. Fig. 6 illustrates the thermodynamics and predicted sequential pathways for (1) formation of a water pentamer, $(H_2O)_5$, from the sequential addition of waters, (2) formation of

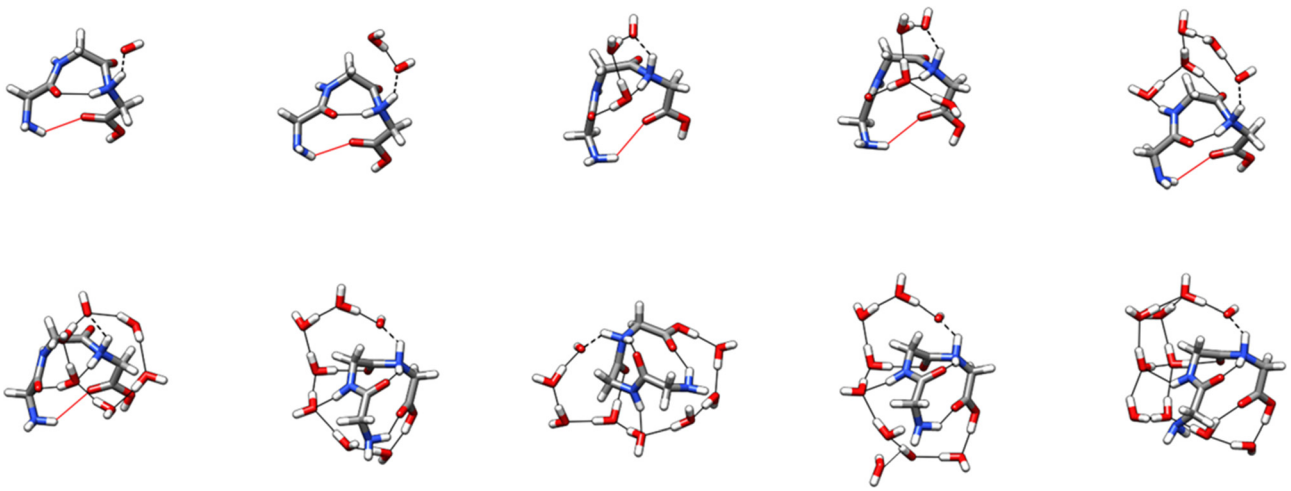


Fig. 5 The structures of *trans* N-to-C triglycine transition states in nanoclusters of water (w = 0–9). These TSs have the $\Delta G^{\circ\#}$ values presented in Table 4. Hydrogen bonds are illustrated with black lines and van der Waals interactions (defined as O–H–O, N–H–O, and O–H–N bond angles less than 140°) are illustrated with red lines.

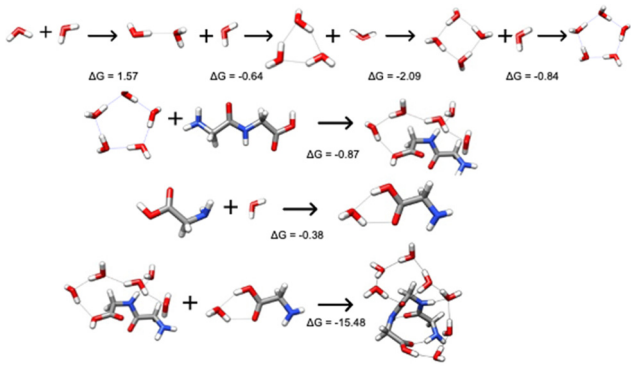


Fig. 6 Step-wise reaction pathway of plausible two-body gas phase reactions to form (Trigly)$(W)_7$. ΔG° values reported in kcal mol^{-1} at 217 K.

the prereactive complex Digly$(H_2O)_5$ from the water pentamer and Digly, (3) formation of the prereactive complex Gly(H_2O) from the Gly and water monomers, and (4) the reaction of Digly$(H_2O)_5$ and Gly(H_2O) to form the Trigly$(H_2O)_7$ product, all at 217 K. All of the standard state Gibbs free energies for these reactions are negative, and the concentration of $(H_2O)_5$ is on the order of 10^9 cm^{-3} and the concentration of Gly$(H_2O)_5$ is on the order of 10^5 cm^{-3} over our temperature ranges,[16,55] so Fig. 6 shows a plausible pathway for formation of the Trigly$(H_2O)_7$ postreactive complex. The PES for this reaction at 217, 259, and 273 K is displayed in Fig. 7. The figure shows that ΔG° for formation of the water pentamer and Gly(H_2O) ranges from −1.1 kcal mol^{-1} at 217 K to 7.3 kcal mol^{-1} at 273 K. Formation of diGly$(H_2O)_5$ from the water pentamer and DiGly ranges from −0.25 kcal mol^{-1} at 217 K to 10.7 kcal mol^{-1} at 273 K. The next step is the formation of the prereactive complex, and the Gibbs free energy difference relative to the separated reactants is 7.13, 17.84, and 21.42 kcal mol^{-1} at 217, 259, and 273 K. The Gibbs free energy activation barriers from these two prereactive complexes is 16.12 kcal mol^{-1} at 217 K, 16.24 kcal mol^{-1} at 259 K, and 16.21 kcal mol^{-1} at 273 K. These results illustrate that when calculating $\Delta G^{\circ\#}$ from the prereactive complexes that the Gibbs activation energies decrease compared to the activation energies calculated from the separated reactants (Table 4 and Fig. 7; 23.25, 34.04, 37.63 kcal mol^{-1}). As the entropy cost is built into the prereactive complexes, $\Delta G^{\circ\#}$ calculated from the prereactive complex is about 16 kcal mol^{-1} at all three temperatures. Thus, all the activation energies determined from the infinitely separated reactants in Tables 2–4 will be lowered.

We note that in the prebiotic atmosphere, the long arc of evolutionary time will ensure that thermodynamically favored reactions will occur. As the ratio of the forward and reverse rate constants equal the equilibrium constant, collisions of the reacting species will eventually produce the lowest Gibbs free energy products. Therefore, the PES presented in Fig. 7 is just one potential reaction mechanism, and one could construct others with different numbers of hydrating waters on the reactants. Water clusters, $(H_2O)_{2-6}$, are ubiquitous in the atmosphere.[5,56] We can imagine a variety of ways for two Glys and six waters to form the products in reaction (7), all of which are thermodynamically favored (Table 1 and Fig. 7).

Zwitterions *versus* canonical glycine, diGlycine, and triglycine

In aqueous solution, the polymerization of amino acids is thermodynamically unfavorable under biologically relevant conditions as free amino acids are zwitterions in solution and because of the condensation reaction that produces a water

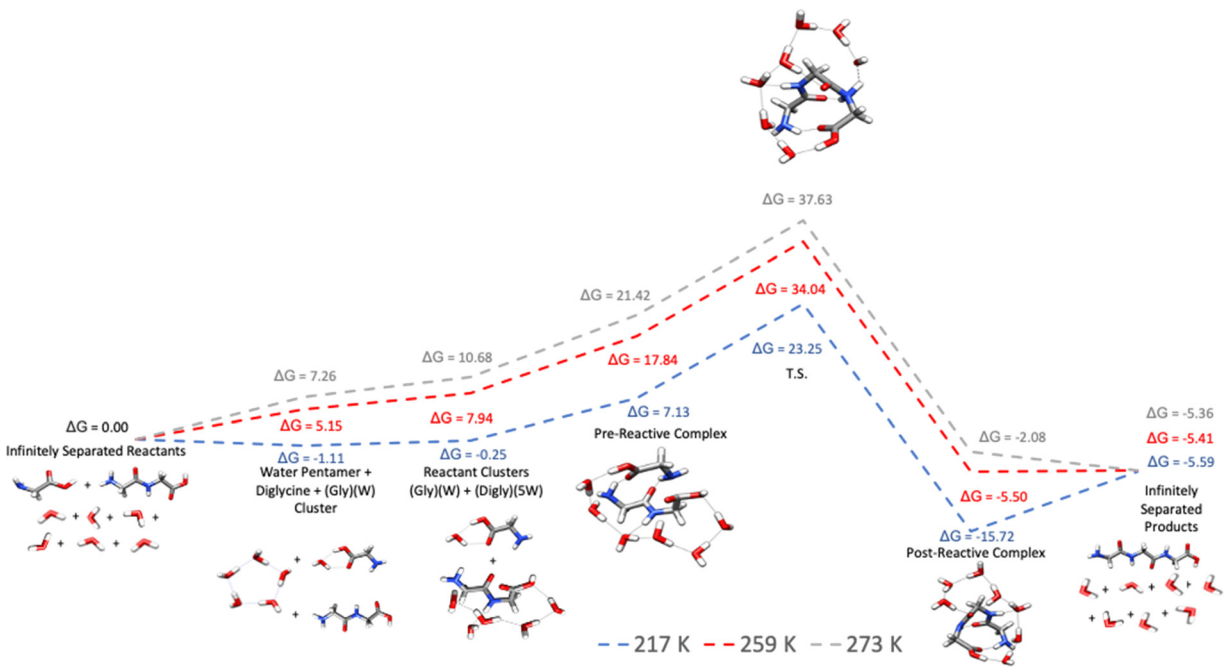
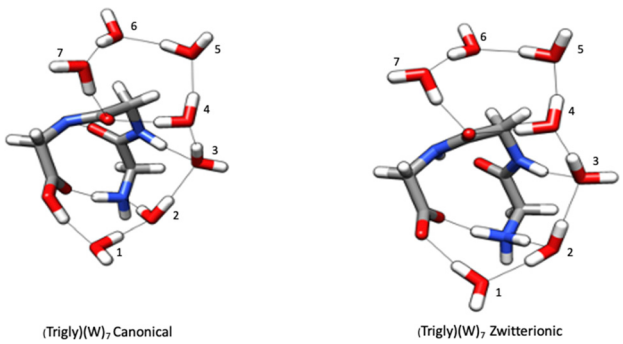


Fig. 7 The Gibbs free energy surface for the reaction of glycine and diglycine with six waters to produce triglycine and seven waters along the reaction pathway illustrated in Fig. 6, relative to the infinitely separate reactants.

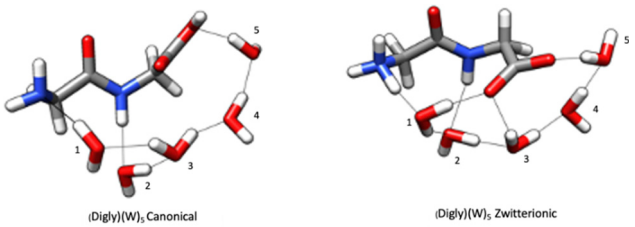
molecule.[16] In the gas phase, however, if one uses a zwitterionic Gly and performs a geometry optimization using quantum chemistry, either through solving the Schrödinger equation using electronic structure theory, or by solving for the electron density using density functional theory, the zwitterion optimizes to the canonical (neutral) Gly: the ammonium cation returns to an amino group and the carboxylate anion returns to a carboxyl group as the hydrogen moves from NH_3^+ to COO^- during geometry optimization. The Gly zwitterion, the stable form in solution, is unstable in the gas phase. The local environment is critical for determining whether Gly is in its neutral canonical form or in its zwitterionic form. For instance, a canonical Gly sticks to a silica surface and its zwitterion is unstable, yet microsolvation with just two water molecules per Gly on this surface changes the population of glycines to approximately 50% of each form, and after microsolvation with four waters per Gly only the zwitterionic form remains.[81] In the gas phase, the lower limit for stabilizing the Gly zwitterion is five water molecules, based on photoelectron spectroscopy of $(Gly)(H_2O)_5^-$.[82] DFT calculations also find five waters as the minimum number to turn the neutral Gly into a zwitterion.[16,83] DLPNO-CCSD(T)/CBS calculations on MP2 geometry optimized structures suggest that 10 waters are required before Gly is unambiguously in its zwitterionic form.[84] We have optimized every structure reported in this paper in both the zwitterionic and canonical forms, and find that in every case the DLPNO/CCSD(T)/CBS//ωb97X-D/6-31++G** calculations predict that the neutral forms are the dominant structures. An illustrative example is provided in Fig. 8. The Intrinsic Reaction Coordinate (IRC) calculation back to the prereactive complex (Fig. 7) reveals that the formation of $(Trigly)(W)_6$ *via* the *trans* N-to-C pathway occurs from the dehydration synthesis reaction of a diglycine cluster with five water molecules and a glycine cluster with one water molecule. We optimized these two reactant clusters both in their canonical form and as zwitterions. The (Gly)(W) cluster returned to the canonical form when optimized as a zwitterion. Fig. 8 shows the results for the optimizations of the canonical and zwitterionic forms of $(Digly)(W)_5$. The canonical structure has seven hydrogen bonds to waters encompassing the entire structure of the dipeptide. Water #1 is donating a hydrogen bond to the amino terminus and accepting a hydrogen bond from water #3. Water #2 is accepting a hydrogen bond from the amino group of the dipeptide and donating one to water #3. Waters #4 and #5 form homodromic hydrogen bonds in a spine connecting to the carboxyl terminus of the dipeptide. The zwitterionic form has nine hydrogen bonds. Water #1 is accepting a hydrogen bond from the protonated amino terminus of the dipeptide and donating hydrogen bonds to water #2 and the negatively charged carboxyl terminus. Water #2 is accepting a hydrogen bond from water #1 and from the amino group of the dipeptide and is donating a hydrogen bond to water #3. Similar to the canonical form, waters #4 and #5 are forming a water spine to the carboxyl terminus. Despite this network of hydrogen bonds, the zwitterionic form is higher in Gibbs free energy at all relevant temperatures. This indicates the stability of the canonical form of diglycine in the gas phase which allows for peptide bond formation to occur. The electronic energy difference between these two forms at the ωb97-XD level, E_{el} (Zwitterion) − E_{el} (Canonical), slightly favors the zwitterion (−1.29 kcal mol^{-1}), while at the DLPNO-CCSD(T)/CBS level the electronic energy difference favors the canonical structure (3.64 kcal mol^{-1}). Adding zero-point-vibrational energy raises the ωb97-XD difference to −0.67 kcal mol^{-1} and the DLPNO-CCSD(T)/CBS difference to 4.37 kcal mol^{-1}. The Gibbs free energy differences at all temperatures favor the canonical form.

The forward path of the IRC also revealed the structure of the postreactive complex as a triglycine peptide with seven waters (Fig. 7). Using the coordinates of this postreactive complex, we optimized it in both its zwitterionic and canonical forms, which are displayed in Fig. 9. The energy of the canonical form is lower at all temperatures, indicating that this is the form that results from the reaction of the $(Digly)(W)_5$ and (Gly)(W) reactant clusters. The canonical form has twelve total hydrogen bonds. Water #1 is donating a hydrogen bond to water #2 and accepting one from the carboxyl group. Water #2 is donating a hydrogen bond from the amino nitrogen as well as to water #3. Water #3 also accepts a hydrogen bond from the amino functional group and donates one to water #4. Water #4 donates a hydrogen bond to the carboxyl oxygen of the central glycine. Water 5, 6, and 7 continue the homodromic water network to the central carboxyl oxygen. Notably, this water spine creates a large amount of bending, allowing for two intramolecular hydrogen bonds to form between functional groups of the tripeptide. The zwitterionic form of the tripeptide forms an almost identical water spine. However, waters #1 and water #2 are flipped in their orientation to accept and donate hydrogen bonds to the charged functional groups. There is also less bending within the tripeptide, as one of the intramolecular hydrogen bonds in the canonical form is a van der Waals interaction in the zwitterionic form. Despite the high level of similarity between these two structures, the zwitterionic form is higher in energy by 5 to 6 kcal mol^{-1}.



Temperature	217 K	259 K	273 K	298 K
ΔG (Zwitterion-Canonical)	5.56 kcal/mol	5.68 kcal/mol	5.72 kcal/mol	5.80 kcal/mol

Fig. 8 Diglycine hydrated with five water molecules in both the canonical (neutral) and zwitterionic forms. The zwitterion is 5 kcal mol^{-1} higher in Gibbs free energy relative to the canonical form over the temperature range of the prebiotic atmosphere.



Temperature	216.65 K	259.20 K	273.15 K	298.15 K
ΔG (Zwitterion-Canonical)	5.04 kcal/mol	5.41 kcal/mol	5.54 kcal/mol	5.78 kcal/mol

Fig. 9 Triglycine hydrated with seven water molecules in both the canonical (neutral) and zwitterionic forms. The zwitterion is 5 kcal mol^{-1} higher in Gibbs free energy relative to the canonical form over the temperature range of the prebiotic atmosphere.

Hydrogen bonding network comparison: the specific case of six waters illustrates the special stability of the water hydrogen bonding network for the *trans* I-6W TS structure

Hydrogen bonding cooperativity is key to understanding the special stability of the *trans* I-6W TS relative to the *cis* III-6W TS. In the top half of Fig. 10 the interaction between the triglycine transition states and six waters is shown, one transition state being *cis* (III-6W) and one *trans* (I-6W). The *trans* transition state I-6W has a $\Delta G^{\circ\#}$ of 23.25 kcal mol^{-1} and the *cis* transition state III-6W has a $\Delta G^{\circ\#}$ value of 28.19 kcal mol^{-1} at 217 K relative to the infinitely separated reactants. The *trans* transition state has ten hydrogen bonds and one van der Waals interaction while the *cis* transition state has nine hydrogen bonds and one van der Waals interaction. We define hydrogen bonds as having hydrogen bonding angles greater than 140° and van der Waals interactions as having N–H–O, O–H–O, and O–H–N angles less than 140°. The van der Waals interactions are represented by the red lines and hydrogen bonds by black lines in the figures. The nature of these water interactions is not the same between the two TSs. In the *trans* I-6W TS structure, the water network encompasses the TS in a much more concerted way. As seen in Fig. 10, key distances of water molecules are positioned to lower the free energy of activation. Water #1 accepts a very strong hydrogen bond from the C-terminus hydroxyl group of the glycine, 1.571 Å, and then this water donates to water #2 (1.663 Å), which donates to water #3 (1.919 Å), which donates to water #4 (1.649 Å), which donates to water #5 (1.773 Å), which donates to water #6 (1.705 Å), and then water #6 donates a very strong hydrogen bond of 1.569 Å to the OH that is being abstracted from the carbon during the formation of the peptide bond. The oxygen on the OH is 2.171 Å from the carbon it is leaving and is 1.612 Å from the hydrogen on the nitrogen (1.071 N–H bond distance) that is part of the formation of the nascent water molecule. In the *cis* TS structure III-6W, synergy of the water hydrogen bonding network does not exist, and the OH oxygen is 2.804 Å from the carbon it is leaving and 2.048 Å away from the hydrogen it is abstracting from the nitrogen. The 4.94 kcal mol^{-1} difference in free energy of activation between these two pathways can be attributed to

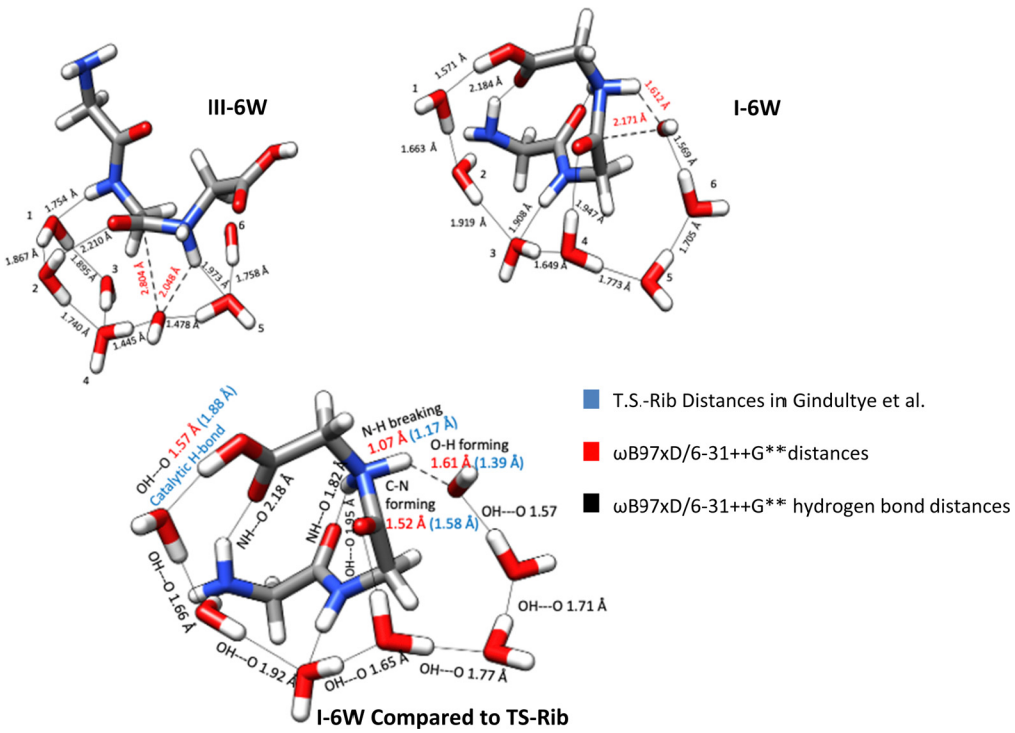


Fig. 10 Comparison of *cis* N-to-C triglycine transition state with six waters (III-6W), *trans* N-to-C triglycine transition state with six waters (I-6W), and the TS calculated for formation of the peptide bond in the ribosome (TS-Rib).[78] Hydrogen bonds are illustrated with black lines and van der Waals interactions (defined as hydrogen bond angles less than 140°) are illustrated with red lines. Covalent bonds that are breaking or forming are illustrated with dashed lines.

the differences in the concerted hydrogen bonding network of I-6W. As Xantheas has demonstrated, the cooperativity of the hydrogen-bonding network in water clusters explains the stability of different water arrangements in gas-phase water clusters.[85] Homodromic hydrogen-bonding networks, which consist of sequential donor–acceptor arrangements between all water molecules, are the most stable. So, for instance, a gas-phase water cluster consisting of four water molecules arranged such that each water molecule accepts one hydrogen bond from the previous water and donates one hydrogen bond to the next water, is more stable than an arrangement where two double donor water molecules are bridged by two double acceptor molecules.[64,85] As seen in Fig. 10, the *trans* TS, I-6W, has six water molecules involved in accepting and donating hydrogen bonds in a mostly homodromic fashion, starting from water #1 accepting a strong hydrogen bond from the hydroxyl of the added glycine (1.571 Å) and ending with water #6 donating a strong hydrogen bond to the dissociating water molecule (1.569 Å). Water #3 also accepts a hydrogen bond from the peptide backbone of the original diglycine (1.849 Å) while water #4 donates a hydrogen bond to the forming carbonyl oxygen in the TS (1.947 Å). Comparing the two networks between the *cis* III-6W TS and the *trans* I-6W TS, the hydrogen bonds are much stronger in *trans* TS. The water molecules are broken into a non-homodromic tetramer (waters 1–4) and a dimer, which are on opposite sides of the OH. bond is 2.171 Å Because glycine has a more flexible back bone, these structures are not strictly *cis* or *trans* (dihedral angles of 0° and 180° respectively). The *trans* TS I-6W has a dihedral angle of 154° and the *cis* TS III-6W has a dihedral angle of 60.4°. Most importantly, the water spine of hydration in the *trans* TS completely encompasses the entire tripeptide, while in the *cis* TS, water only interacts with the glycine monomer and one of the glycines of the dimer.

Comparison of the *trans* I-6W TS with the TS of the ribosome

Ribosomes are the site of protein synthesis in living organisms and are complex molecular machines containing RNA and proteins.[86–89] In the large subunit of the ribosome a peptide bond is formed by the nucleophilic displacement of the P-site tRNA by the amino group of the 3′-linked aminoacyl-tRNA in the A site, requiring that the tRNAs in the A and P sites are arranged with acceptor stems close together.[90] The TS for the formation of the peptide bond in the ribosome has been constructed by using the crystallographic coordinates of the tRNA substrate located in the ribosome peptidyl transferase center combined with quantum mechanical calculations.[91] In this insightful study, 50 atoms were optimized to calculate the TS that is formed simultaneously with the rotary motion that enables the translocation of the A-site tRNA 3′ end into the P site. The calculated TS energy, based on the DFT model chemistry, B3LYP/6-31+G**,[92,93] was 35.5 kcal mol^{-1}. (This is an energy, without the entropic corrections necessary to compute the free energy of activation and calculated at a lower level of theory than our work.) Strikingly, the calculated TS structure, TS-Rib in the lower half of Fig. 10, shows a surprising amount of synergy with the *trans* I-6W TS. In TS-Rib, the C-N forming bond is 1.58 Å, the N–H breaking bond is 1.17 Å, and the C–OH breaking distance is 1.91 Å. These same values for *trans* I-6W TS are 1.517, 1.071, and 2.171 Å. In the ribosome, there is a catalytic hydrogen-bond from the 2′OH of A76 of the P-site tRNA with the carboxyl oxygen of the amino acid bound to A-site RNA. In I-6W, the corresponding hydrogen bond is the 1.57 Å hydrogen bond between the carboxyl oxygen of the glycine to water #1. This strong hydrogen bond is absent in the *cis* TS. Also, in the TS-Rib structure, there is a 1.91 Å breaking bond between the P-site tRNA and the growing peptide. This bond is replaced in I-6W by the 2.171 Å breaking bond between the diglycine carbon and the leaving OH group. Notably, the OH group is stabilized by water #6 and the homodromic water spine of hydration that connects this network to the added glycine through water #1. The striking resemblance of the I-6W TS to TS-Rib leads to the idea that perhaps prebiotic synthesis of small peptides in the gas phase share a similar pathway as that in the first primitive ribosome, because it is kinetically favored in both cases through hydrogen bonding.

Serine peptides and chirality

There are ten amino acids identified to be prebiotically relevant.[94] Based on this identification, we have built and optimized dipeptide TS structures involving serine, glycine, and alanine. Structures containing serine and alanine were built with stereochemistry in mind, as there is still much unknown about the origins of homochirality. This is an essential problem in prebiotic chemistry, as how would life have evolved without homochirality? There are three main classes of molecules wherein homochirality may have evolved: sugars,[95] RNA/DNA,[96] and amino acids.[97] As shown in Fig. 11, there are energetic differences for unhydrated diserine TS structures with different stereochemistry. The lowest energy TS involves the N terminus of a serine with D stereochemistry adding to the C terminus of a serine with L stereochemistry.

The energetic differences of these TS structures likely result from steric hindrance or strain between the serine R groups. The gly-ser-D TS structures show more intramolecular bending than the ser-gly-L TS structures (not shown). In all the gly-ser-D TS structures (Fig. 12) there is a hydrogen bond between the amino Fig. 12. The structures of gly-ser-D transition states.

In all the gly-ser-D TS structures (Fig. 12) there is a hydrogen bond between the amino terminus and the R group hydroxyl. The structure with one water contains five hydrogen bonds. Both amino functional groups are donating hydrogen bonds to the leaving hydroxyl group. The water is donating an additional hydrogen bond and accepting one from the central amino group. There is a remarkable decrease in energy with the addition of the second water molecule of 10.59 kcal mol^{-1} at 217 K. This is the largest drop in energy with the addition of a water molecule in any of the peptides examined. This structure is like the one water structure, with the second water donating a fourth hydrogen bond to the leaving hydroxyl group. The stability of this transition state is likely due to the way in which the four hydrogen bonds are positioning the leaving hydroxyl

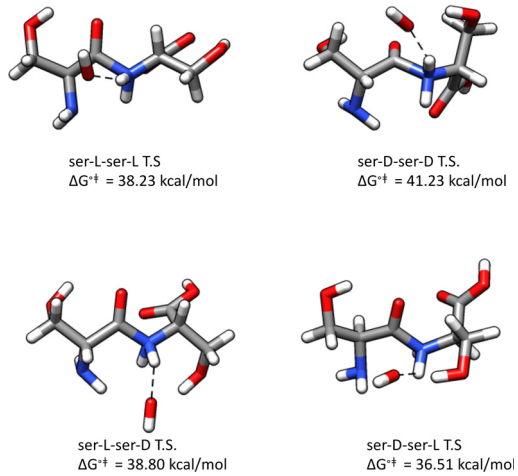


Fig. 11 The structures of diserine transition states. These TSs have the $\Delta G^{\circ\#}$ values presented in kcal mol^{-1} at 217 K.

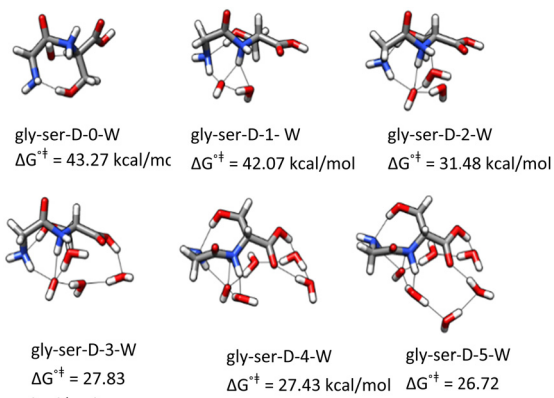


Fig. 12 The structures of gly-ser-D transition states. These TSs have the $\Delta G^{\circ\#}$ values presented in kcal mol^{-1} at 217 K.

group to form a water molecule. This transition state structure highlights the capacity of hydrogen bonding topology to lower the activation energy of a reaction. The three, four, and five water structures form the water spine motif that has been observed in many of the peptides. Fig. 13 shows the gly-gly-ser-L TS tripeptide structures. These transition states were modelled by adding a glycine *via* the N-to-C *tra*ns pathway in accordance with the results of the triglycine study. The unhydrated TS structure has an intramolecular hydrogen bond between the amino functional group of the glycine monomer and the carboxyl functional group of the serine. The energy of this structure is 39.74 kcal mol^{-1} at 217 K, making it lower in energy than trigly TS structures II, III, and IV (Table 3). The one and two water structures begin to form a hydrogen network around the leaving hydroxyl group. The addition of the third water begins a bridging motif where the intramolecular hydrogen bond exists in the zero, one, and two water structures. The four, five, six, and seven water structures builds on this bridging network. The energy values appear to remain consistent around 29 kcal mol^{-1} once the fourth water is added, whereas in the comparable triglycine TS structures, the sequential addition of water continuously lowers the energy of activation. Due to serine's polar R group, there might be a lower upper limit of waters that reduces the energy barrier of peptide formation. This phenomenon should be further examined with other peptide combinations involving polar or charged R groups in nanoclusters of water.

Conclusions

By examining four different diglycine and triglycine transition states, we conclude that the formation of diglycine and triglycine in atmospheric nanoclusters of water in the prebiotic atmosphere kinetically favors adding through an N-to-C

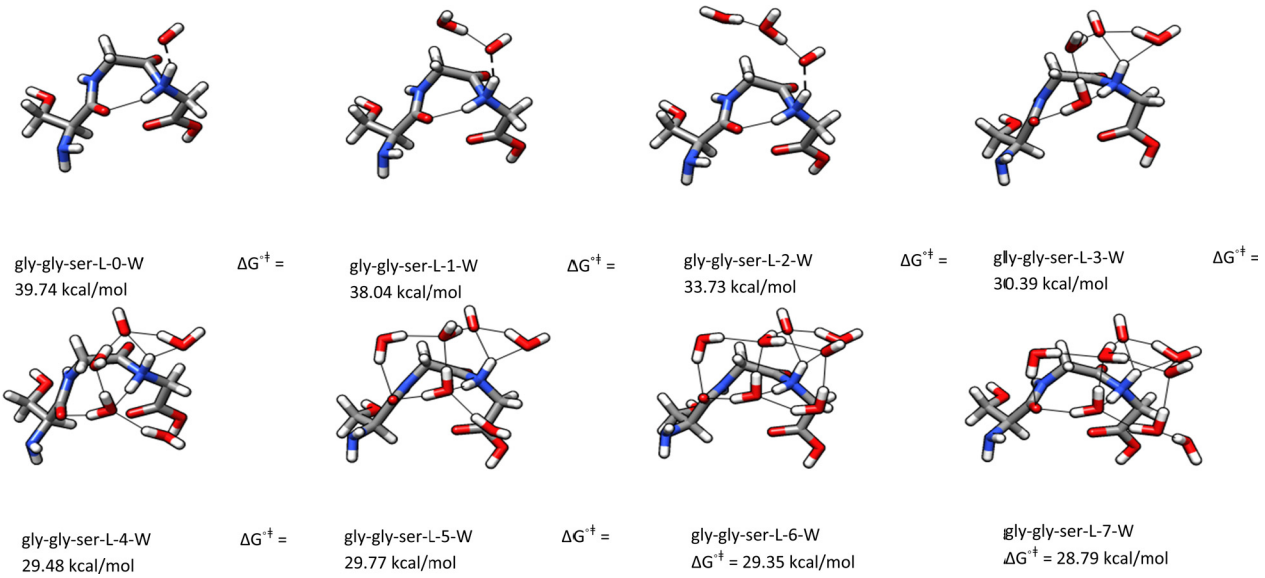


Fig. 13 The structures of gly-gly-ser-L transition states hydrated with waters (0–7). These TSs have the $\Delta G^{\circ\#}$ values presented in kcal mol^{-1} at 217 K.

synthesis of glycines through a *trans* conformation. The *trans* N-to-C free energies of activation values, $\Delta G^{\circ\#}$, are consistently lower than *trans* C-to-N and both *cis* pathways by three to eleven kcal mol^{-1} over the estimated temperature range of the prebiotic atmosphere. Water stabilizes the transition state structures and lowers the activation energy in the *trans* N-to-C transition states. By adding six waters to stabilize the TS structure, the Gibbs free activation energy calculated relative to the infinitely separated reactants is reduced by 13 kcal mol^{-1} at 217 K, which is equivalent to a rate speed up of roughly 13 orders of magnitude! As temperature increases, the rate decreases, and the value of $\Delta G^{\circ\#}$ is reduced by 4.7 kcal mol^{-1} at 259 K and 2.6 kcal mol^{-1} at 273 K for the six water nanocluster. The Gibbs activation energies of nonhydrated diglycine and triglycine range from 36–40 kcal mol^{-1} over the temperature range of the prebiotic atmosphere, and these values are reduced dramatically when the Gibbs activation energies are calculated from the prereactive complexes: for the diglycine-glycine-six-water prereactive cluster, the Gibbs activation energies are reduced to 16 kcal mol^{-1}. Examination of the *trans vs. cis* six water transition states reveals that a homodromic water network maximizes the acceptor/donor nature of the six waters. These form a spine of hydration extending from the added glycine to the leaving water molecule and the homodromic water network is responsible for the kinetic favorability of the *trans* N-to-C pathway. Nature uses the *trans* N-to-C pathway to synthesize proteins in the ribosome, and we note the similarities in hydrogen bonding stabilization between the transition state for peptide synthesis in the ribosome and the transition states formed in nanoclusters of water in the same pathway.

Our comparison of serine dipeptides with different stereochemistries reveals that there does not appear to be a significant energetic preference for the L configuration. Thus, it is likely that a driving force (polarized light, catalyst, *etc.*) was needed to produce the homochirality present in extant amino acids. The serine-glycine dipeptide TS structures show structural similarities to the diglycine TS structures. However, the positioning of the R group appears to be key in the catalytic role of waters for these reactions. These results support the hypothesis that small oligomers formed in a prebiotic atmosphere and rained down onto earth's surface.

Author contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Funding for this work was provided by grants CHE-1229354, CHE-16626238, CHE-1903871, and CHE-2018427 from the National Science Foundation. High-performance computing resources were provided by the Research Corporation for Science Advancement (27446) and the MERCURY Consortium (**https://www.mercuryconsortium.org**)[98,99]

References

1 K. Zahnle, N. Arndt, C. Cockell, A. Halliday, E. Nisbet, F. Selsis and N. H. Sleep, Emergence of a Habitable Planet, *Space Sci. Rev.*, 2007, **129**, 35–78.
2 S. Marchi, W. F. Bottke, L. T. Elkins-Tanton, M. Bierhaus, K. Wuennemann, A. Morbidelli and D. A. Kring, Widespread mixing and burial of Earth's Hadean crust by asteroid impacts, *Nature*, 2014, **511**, 578–582.
3 M. Frenkel-Pinter, M. Samanta, G. Ashkenasy and L. J. Leman, Prebiotic Peptides: Molecular Hubs in the Origin of Life, *Chem. Rev.*, 2020, **120**, 4707–4765.
4 S. Peters, D. A. Semenov, R. Hochleitner and O. Trapp, Synthesis of prebiotic organics from CO_2 by catalysis with meteoritic and volcanic particles, *Sci. Rep.*, 2023, **13**, 6843.
5 M. A. Sephton, Organic compounds in carbonaceous meteorites, *Nat. Prod. Rep.*, 2002, **19**, 292–311.
6 C. Woese, *The Genetic Code: The Molecular Basis for Genetic Expression*, Harper and Row, New York, 1967.
7 F. H. Crick, The origin of the genetic code, *J. Mol. Biol.*, 1968, **38**, 367–379.
8 L. E. Orgel, Evolution of the genetic apparatus, *J. Mol. Biol.*, 1968, **38**, 381–393.
9 M. P. Robertson and G. F. Joyce, The origins of the RNA world, *Cold Spring Harb Perspect. Biol.*, 2012, **4**, 01003608.
10 S. Islam and M. W. Powner, Prebiotic Systems Chemistry: Complexity Overcoming Clutter, *Chem*, 2017, **2**, 470–501.
11 J. G. Forsythe, A. S. Petrov, W. C. Millar, S.-S. Yu, R. Krishnamurthy, M. A. Grover, N. V. Hud and F. M. Fernández, Surveying the sequence diversity of model prebiotic peptides by mass spectrometry, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E7652–E7659.
12 M. Frenkel-Pinter, M. Bouza, F. M. Fernández, L. J. Leman, L. D. Williams, N. V. Hud and A. Guzman-Martinez, Thioesters provide a plausible prebiotic path to proto-peptides, *Nat. Commun.*, 2022, **13**, 2569.
13 F. Sauer, M. Haas, C. Sydow, A. F. Siegle, C. A. Lauer and O. Trapp, From amino acid mixtures to peptides in liquid sulphur dioxide on early Earth, *Nat. Commun.*, 2021, **12**, 7182.
14 M. Frenkel-Pinter, J. W. Haynes, M. C. , A. S. Petrov, B. T. Burcar, R. Krishnamurthy, N. V. Hud, L. J. Leman and L. D. Williams, Selective incorporation of proteinaceous over nonproteinaceous cationic amino acids in model prebiotic oligomerization reactions, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 16338–16346.
15 I.-S. Jeon, D.-S. Ahn, S.-W. Park, S. Lee and B. Kim, Structures and isomerization of neutral and zwitterion serine–water clusters: computational study, *Int. J. Quantum Chem.*, 2005, **101**, 55–66.

16 A. G. Gale, T. T. Odbadrakh, B. T. Ball and G. C. Shields, Water-Mediated Peptide Bond Formation in the Gas Phase: A Model Prebiotic Reaction, *J. Phys. Chem. A*, 2020, **124**, 4150–4159.
17 D. Comte, L. Lavy, P. Bertier, F. Calvo, I. Daniel, B. Farizon, M. Farizon and T. D. Märk, Glycine Peptide Chain Formation in the Gas Phase via Unimolecular Reactions, *J. Phys. Chem. A*, 2023, **127**, 775–780.
18 R. I. Kaiser, A. M. Stockton, Y. S. Kim, E. C. Jensen and R. A. Mathies, On the formation of dipeptides in interstellar model ices, *Astrophys. J.*, 2013, **765**, 111.
19 D. T. Holden, N. M. Morato and R. G. Cooks, Aqueous microdroplets enable abiotic synthesis and chain extension of unique peptide isomers from free amino acids, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2212642119.
20 E. C. Griffith and V. Vaida, In situ observation of peptide bond formation at the water-air interface, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 15697–15701.
21 I. Nam, H. G. Nam and R. N. Zare, Abiotic synthesis of purine and pyrimidine ribonucleosides in aqueous microdroplets, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 36–40.
22 V. Vaida, Prebiotic phosphorylation enabled by microdroplets, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 12359–12361.
23 W. Wang, L. Qiao, J. He, Y. Ju, K. Yu, G. Kan, C. Guo, H. Zhang and J. Jiang, Water Microdroplets Allow Spontaneously Abiotic Production of Peptides, *J. Phys. Chem. Lett.*, 2021, **12**, 5774–5780.
24 Q. Zhao, S. S. Garimella and B. M. Savoie, Thermally Accessible Prebiotic Pathways for Forming Ribonucleic Acid and Protein Precursors from Aqueous Hydrogen Cyanide, *J. Am. Chem. Soc.*, 2023, **24**, 6135–6143.
25 M. T. C. Martins-Costa and M. F. Ruiz-López, Electrostatics and Chemical Reactivity at the Air–Water Interface, *J. Am. Chem. Soc.*, 2023, **145**, 1400–1406.
26 G. B. Ellison, A. F. Tuck and V. Vaida, Atmospheric processing of organic aerosols, *J. Geophys. Res.: Atmos.*, 1999, **104**, 11633–11641.
27 C. M. Dobson, G. B. Ellison, A. F. Tuck and V. Vaida, Atmospheric aerosols as prebiotic chemical reactors, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11864–11868.
28 S. M. Horst, R. V. Yelle, A. Buch, N. Carrasco, G. Cernogora, O. Dutuit, E. Quirico, E. Sciamma-O'Brien, M. A. Smith, A. Somogyi, C. Szopa, R. Thissen and V. Vuitton, Formation of Amino Acids and Nucleotide Bases in a Titan Atmosphere Simulation Experiment, *Astrobiology*, 2012, **12**, 809–817.
29 S. Kadoya, J. Krissansen-Totton and D. C. Catling, Probable Cold and Alkaline Surface Environment of the Hadean Earth Caused by Impact Ejecta Weathering, *Geochem., Geophys., Geosyst.*, 2020, **21**, e2019GC008734.
30 Z. R. Todd, A. C. Fahrenbach, C. J. Magnani, S. Ranjan, A. Björkbom, J. W. Szostak and D. D. Sasselov, Solvated-electron production using cyanocuprates is compatible with the UV-environment on a Hadean–Archaean Earth, *Chem. Commun.*, 2018, **54**, 1121–1124.
31 B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy and J. D. Sutherland, Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism, *Nat. Chem.*, 2015, **7**, 301–307.
32 S. Ranjan, Z. R. Todd, J. D. Sutherland and D. D. Sasselov, Sulfidic Anion Concentrations on Early Earth for Surficial Origins-of-Life Chemistry, *Astrobiology*, 2018, **18**, 1023–1040.
33 S. Kempe and E. T. Degens, An early soda ocean?, *Chem. Geol.*, 1985, **53**, 95–108.
34 J. L. Bada and A. Lazcano, Some Like It Hot, But Not the First Biomolecules, *Science*, 2002, **296**, 1982–1983.
35 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminf.*, 2012, **4**, 17.
36 A. G. Gale, T. T. Odbadrakh and G. C. Shields, Catalytic activity of water molecules in gas-phase glycine dimerization, *Int. J. Quantum Chem.*, 2020, **120**, e26469.
37 J. D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, *J. Chem. Phys.*, 2008, **128**, 084106.
38 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions, *J. Chem. Phys.*, 1980, **72**, 650–654.
39 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F, *J. Comput. Chem.*, 1983, **4**, 294–301.
40 M. J. Frisch, J. A. Pople and J. S. Binkley, Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets, *J. Chem. Phys.*, 1984, **80**, 3265–3269.
41 L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov and J. A. Pople, Gaussian-3 (G3) theory for molecules containing first and second-row atoms, *J. Chem. Phys.*, 1998, **109**, 7764–7766.
42 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. L. 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 B.01*, Wallingford, CT, 2016.
43 K. Fukui, The path of chemical reactions – the IRC approach, *Acc. Chem. Res.*, 2002, **14**, 363–368.
44 K. K. Irikura, *THERMO.PL*, NIST, 2002.
45 C. Riplinger, P. Pinski, U. Becker, E. F. Valeev and F. Neese, Sparse maps--A systematic infrastructure for reduced-

scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory, *J. Chem. Phys.*, 2016, **144**, 024109.
46 C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, Natural triple excitations in local coupled cluster calculations with pair natural orbitals, *J. Chem. Phys.*, 2013, **139**, 134101.
47 C. Riplinger and F. Neese, An efficient and near linear scaling pair natural orbital based local coupled cluster method, *J. Chem. Phys.*, 2013, **138**, 034106.
48 F. Neese, A. Hansen and D. G. Liakos, Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis, *J. Chem. Phys.*, 2009, **131**, 064103.
49 F. Neese, The ORCA program system, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.
50 F. Neese, Software update: the ORCA program system, version 4.0, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2017, **8**, e1327.
51 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, The ORCA quantum chemistry program package, *J. Chem. Phys.*, 2020, **152**, 224108.
52 T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
53 T. Helgaker, W. Klopper, H. Koch and J. Noga, Basis-set convergence of correlated calculations on water, *J. Chem. Phys.*, 1997, **106**, 9639–9646.
54 S. S. Xantheas, On the importance of the fragment relaxation energy terms in the estimation of the basis set superposition error correction to the intermolecular interaction energy, *J. Chem. Phys.*, 1996, **104**, 8821–8824.
55 M. E. Dunn, E. K. Pokon and G. C. Shields, Thermodynamics of forming water clusters at various temperatures and pressures by Gaussian-2, Gaussian-3, Complete Basis Set-QB3, and Complete Basis Set-APNO model chemistries; implications for atmospheric chemistry, *J. Am. Chem. Soc.*, 2004, **126**, 2647–2653.
56 B. Temelso, K. A. Archer and G. C. Shields, Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections, *J. Phys. Chem. A*, 2011, **115**, 12034–12046.
57 C. Pérez, M. T. Muckle, D. P. Zaleski, N. A. Seifert, B. Temelso, G. C. Shields, Z. Kisiel and B. H. Pate, Structures of Cage, Prism, and Book Isomers of Water Hexamer from Broadband Rotational Spectroscopy, *Science*, 2012, **336**, 897–901.
58 J. O. Richardson, C. Perez, S. Lobsiger, A. A. Reid, B. Temelso, G. C. Shields, Z. Kisiel, D. J. Wales, B. H. Pate and S. C. Althorpe, Concerted hydrogen-bond breaking by quantum tunneling in the water hexamer prism, *Science*, 2016, **351**, 1310–1313.
59 K. S. Pitzer and W. D. Gwinn, Energy Levels and Thermodynamic Functions for Molecules with Internal Rotation I. Rigid Frame with Attached Tops, *J. Chem. Phys.*, 1942, **10**, 428–440.
60 V. Barone, Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation, *J. Chem. Phys.*, 2004, **120**, 3059–3065.
61 V. Barone, Anharmonic vibrational properties by a fully automated second-order perturbative approach, *J. Chem. Phys.*, 2005, **122**, 014108.
62 O. Isayev, L. Gorb and J. Leszczynski, Theoretical calculations: Can Gibbs free energy for intermolecular complexes be predicted efficiently and accurately?, *J. Comput. Chem.*, 2007, **28**, 1598–1609.
63 M. E. Dunn, T. M. Evans, K. N. Kirschner and G. C. Shields, Prediction of accurate anharmonic experimental vibrational frequencies for water clusters, $(H_2O)(n)$, n = 2–5, *J. Phys. Chem. A*, 2006, **110**, 303–309.
64 M. B. Day, K. N. Kirschner and G. C. Shields, Global search for minimum energy $(H_2O)(n)$ clusters, n = 3–5, *J. Phys. Chem. A*, 2005, **109**, 6773–6778.
65 B. Temelso and G. C. Shields, The Role of Anharmonicity in Hydrogen-Bonded Systems: The Case of Water Clusters, *J. Chem. Theory Comput.*, 2011, **7**, 2804–2817.
66 T. E. Morrell and G. C. Shields, Atmospheric Implications for Formation of Clusters of Ammonium and 1-10 Water Molecules, *J. Phys. Chem. A*, 2010, **114**, 4266–4271.
67 D. R. Hanson and F. L. Eisele, Diffusion of H_2SO_4 in Humidified Nitrogen: Hydrated H_2SO_4, *J. Phys. Chem. A*, 2000, **104**, 1715–1719.
68 B. Temelso, T. E. Morrell, R. M. Shields, M. A. Allodi, E. K. Wood, K. N. Kirschner, T. C. Castonguay, K. A. Archer and G. C. Shields, Quantum Mechanical Study of Sulfuric Acid Hydration: Atmospheric Implications, *J. Phys. Chem. A*, 2012, **116**, 2209–2224.
69 B. Temelso, T. N. Phan and G. C. Shields, Computational Study of the Hydration of Sulfuric Acid Dimers: Implications for Acid Dissociation and Aerosol Formation, *J. Phys. Chem. A*, 2012, **116**, 9745–9758.
70 I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries, *J. Chem. Theory Comput.*, 2010, **6**, 2872–2887.
71 C. J. Bready, S. Vanovac, T. T. Odbadrakh and G. C. Shields, Amino Acids Compete with Ammonia in Sulfuric Acid-Based Atmospheric Aerosol Prenucleation: The Case of Glycine and Serine, *J. Phys. Chem. A*, 2022, **126**, 5195–5206.
72 J. Elm, D. Ayoubi, M. Engsvang, A. B. Jensen, Y. Knattrup, J. Kubečka, C. J. Bready, V. R. Fowler, S. E. Harold, O. M. Longsworth and G. C. Shields, Quantum chemical modeling of organic enhanced atmospheric nucleation: a critical review, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2023, e1662.
73 R. B. Martin, Free energies and equilibria of peptide bond hydrolysis and formation, *Biopolymers*, 1998, **45**, 351–353.
74 G. Danger, R. Plasson and R. Pascal, Pathways for the formation and evolution of peptides in prebiotic environments, *Chem. Soc. Rev.*, 2012, **41**, 5416–5429.
75 E. Van Dornshuld, R. A. Vergenz and G. S. Tschumper, Peptide bond formation via glycine condensation in the gas phase, *J. Phys. Chem. B*, 2014, **118**, 8583–8590.
76 H. M. Dintzis, Assembly of peptide chains of hemoglobin, *Proc. Natl. Acad. Sci. U. S. A.*, 1961, **47**, 247–261.

77 R. M. Shields, B. Temelso, K. A. Archer, T. E. Morrell and G. C. Shields, Accurate Predictions of Water Cluster Formation, $(H_{(2)}O)(n = 2–10)$, *J. Phys. Chem. A*, 2010, **114**, 11725–11737.
78 K. Fukui, The path of chemical reactions – the IRC approach, *Acc. Chem. Res.*, 1981, **14**, 363–368.
79 H. P. Hratchian and H. B. Schlegel, Accurate reaction paths using a Hessian based predictor–corrector integrator, *J. Chem. Phys.*, 2004, **120**, 9918–9924.
80 H. P. Hratchian and H. B. Schlegel, Using Hessian Updating To Increase the Efficiency of a Hessian Based Predictor-Corrector Reaction Path Following Method, *J. Chem. Theory Comput.*, 2005, **1**, 61–69.
81 A. Rimola, B. Civalleri and P. Ugliengo, Neutral *vs.* Zwitterionic Glycine Forms at the Water/Silica Interface: Structure, Energies, and Vibrational Features from B3LYP Periodic Simulations, *Langmuir*, 2008, **24**, 14027–14034.
82 S. Xu, J. M. Nilles and K. H. Bowen, Jr., Zwitterion formation in hydrated amino acid, dipole bound anions: How many water molecules are required?, *J. Chem. Phys.*, 2003, **119**, 10696–10701.
83 R. Pérez de Tudela and D. Marx, Water-Induced Zwitterionization of Glycine: Stabilization Mechanism and Spectral Signatures, *J. Phys. Chem. Lett.*, 2016, **7**, 5137–5142.
84 R. Tripathi, L. D. Caballero, R. P. de Tudela, C. Holzl and D. Marx, Unveiling Zwitterionization of Glycine in the Microhydration Limit, *ACS Omega*, 2021, **6**, 12676–12683.
85 S. S. Xantheas, Cooperativity and hydrogen bonding network in water clusters, *Chem. Phys.*, 2000, **258**, 225–231.
86 N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution, *Science*, 2000, **289**, 905–920.
87 B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch and V. Ramakrishnan, Structure of the 30S ribosomal subunit, *Nature*, 2000, **407**, 327–339.
88 F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi and A. Yonath, Structure of functionally activated small ribosomal subunit at 3.3 angstrom resolution, *Cell*, 2000, **102**, 615–623.
89 J. Harms, F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi and A. Yonath, High resolution structure of the large ribosomal subunit from a mesophilic Eubacterium, *Cell*, 2001, **107**, 679–688.
90 D. Voet, J. G. Voet and C. W. Pratt, *Fundamentals of Biochemistry: Life at the Molecular Level*, John Wiley & Sons, Inc., New York, 5th edn, 2002.
91 A. Gindulyte, A. Bashan, I. Agmon, L. Massa, A. Yonath and J. Karle, The transition state for formation of the peptide bond in the ribosome, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 13327–13332.
92 A. D. Becke, Density–functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
93 C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
94 C. E. Cornell, R. A. Black, M. Xue, H. E. Litz, A. Ramsay, M. Gordon, A. Mileant, Z. R. Cohen, J. A. Williams, K. K. Lee, G. P. Drobny and S. L. Keller, Prebiotic amino acids bind to and stabilize prebiotic fatty acid membranes, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 17239–17244.
95 M. Mauksch, Spontaneous emergence of enantioenriched chiral aldol reaction products from Achiral precursors in solution and origin of biological homochirality of sugars: a first-principles study, *Phys. Chem. Chem. Phys.*, 2023, **25**, 1734–1754.
96 K. Michaelian, Homochirality through Photon-Induced Denaturing of RNA/DNA at the Origin of Life, *Life*, 2018, **8**, 21.
97 Q. Sallembien, L. Bouteiller, J. Crassous and M. Raynal, Possible chemical and physical scenarios towards biological homochirality, *Chem. Soc. Rev.*, 2022, **51**, 3436–3476.
98 G. C. Shields, The Molecular Education and Research Consortium in Undergraduate Computational Chemistry (MERCURY): Twenty Years of Exceptional Success Supporting Undergraduate Research and Inclusive Excellence, *SPUR-Scholarship and Practice of Undergraduate Research*, 2019, **3**, 5–15.
99 G. C. Shields, Twenty years of exceptional success: the molecular education and research consortium in undergraduate computational chemistry (MERCURY), *Int. J. Quantum Chem.*, 2020, **120**, e26274.

