Supporting Information

Unravelling the Prebiotic Origins of the Simplest α-Ketoacids in Cometary Ices: A Computational Investigation

Soumya Ranjan Dash,[†a,b] Rinu Pandya,[†a,b] Geetika Singh,[†a] Himanshu Sharma,[†a,b] Tamal Das,[†a,b] Hritwik Haldar,[c] Srinivas Hotha,[*c] Kumar Vanka[*a,b]

[a]*Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Pune 411008, India. *E-mail: k.vanka@ncl.res.in*

[b]*Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India.*

[c]*Department of Chemistry, Indian Institute of Science Education and Research, Pune 411008, India. *E-mail: s.hotha@iiserpune.ac.in*

[†]These authors contributed equally.

Table of Contents

Computational Details

Quantum Mechanical Calculations: In order to determine the reaction free energy (ΔG) and energy barriers ($\Delta G^{\#}$), we have done the minimum energy pathway (MEP) search by full quantum mechanical calculations, including zero point energy, internal energy, and entropic contributions, with the temperature taken to be 298.15 K. All structural calculations were carried out using density functional theory (DFT) within the Gaussian09 software package[S1] employing the 6-311++G(d,p) basis set[S2] and the B3LYP hybrid density functional[S3] with dispersion corrections (D3). We have modelled the solvent with the polarizable continuum model (PCM)[S4] with water as the solvent. Frequency calculations were performed for all the stationary points to confirm them as a local minima or transition state structures. The absence of imaginary frequency confirmed the minima, while the presence of a single imaginary frequency verified the transition states. We have further done intrinsic reaction coordinate (IRC) calculations to confirm that the obtained transition states connect with the correct reactants and products.

Details of the *ab initio* Nanoreactor (AINR) Simulations: The nanoreactor simulations utilized the TeraChem 1.9 software package[S5-S11] with the Hartree-Fock (HF)[S12] electronic wave function and the 3-21G basis set[S13] to calculate the Born-Oppenheimer potential energy surface. This methodology was developed by Martinez and co-workers.[S14] This level of theory was chosen due to the HF method's reputation for predicting chemically reasonable structures.[S15] It is important to clarify that HF was not used for determining barrier heights and reaction rates; rather, its role was limited to the discovery phase. This approach mirrors Martinez and co-workers' original AINR paper, where they successfully replicated results from the Urey-Miller experiment and the interaction of acetylene molecules using HF/3-21G.[S14] While attempts were made previously to explore discovery in the AINR simulations with DFT using the B3LYP density functional and the 3-21g(d) basis set, preliminary intermediates were identified at a higher computational cost without yielding significantly different results compared to the HF/3-21g approach (Reference 12 of the main manuscript). Consequently, the discovery process in the AINR simulations was limited to HF/3-21g.

AINR Spherical Boundary Conditions[14]: Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the “evaporation” event. The

spherical boundary conditions were provided in the form of a sum of two harmonic terms. The molecules were restricted to move inside a spherical volume by a boundary potential, with a time-dependent component:

$$V(r, t)= f(t)U(r,r_1,k_1) +(1− f(t)) U(r,r_2,k_2)$$

$$U(r,r_0,k) = mk/2\ (r – r_0)^2\theta(r – r_0);\ f(t) = \theta(\lfloor t/T \rfloor − t/T + \tau/T)$$

where k_1 = 1.0 kcal $mol^{-1}Å^{-2}$, r_1 = 13.0 Å, k_2 = 0.5 kcal $mol^{-1}Å^{-2}$, r_2 = 5.0 Å, τ = 0.75ps, T = 1.0 ps, ⌊⌋ is the floor function and θ is the heaviside step function. The function f(t) is a rectangular wave that oscillates between one (duration τ) and zero (duration T–τ), and U(r, r_0, k) is a radial potential that is zero inside the prescribed radius r_0 and harmonic outside. The force constant is multiplied by the atomic mass (in a.m.u) such that all the atoms at the same radial coordinate have attained equal acceleration. The rectangular waveform switches the restraint potential between U(r, r_1, k_1) and U(r, r_2, k_2), which forces the atoms with a radial position 10.0 Å to 3.5 Å towards the centre of the sphere and allows them to collide. When the sphere is expanded again, the molecules present in the smaller volume diffuse rapidly (because of the high simulation temperature) to occupy the larger volume. Due to the repeating compression and expansion of spherical volume, the molecules collide and relax. Therefore, throughout the simulation, new molecules are formed and then break again to form other new molecules. We have run simulations with different composition of CH_3OH, H_2O, NH_3, HCHO, CO and CO_2 with a timestep of 0.5 fs.

Optimizations of the *ab initio* Nanoreactor (AINR) Simulations: We have done several AINR simulations for optimizing the initial conditions of the simulations, based on the different parameters that can affect the results. The parameters are the ratio of the reactant species, the number of molecules taken in the simulation box, spherical boundary conditions, temperature and the total time of the AIMD simulations.

We now briefly discuss the results obtained from our several simulations based on the tested nanoreactor parameters:

(i) <u>Formation of key intermediates</u>: Throughout the diverse pathways from initial reactants to key species, intermediates including formaldehyde, singlet carbenes, hydrogen cyanide (HCN), carbonic acid, formic acid and formaldoxime were consistently identified. These findings indicate the pivotal role of these intermediates in the synthesis of the target molecules.

Additionally, a diverse array of organic species was also detected to have formed during the simulations. The intermediates like glycolaldehyde, ethylene glycol, oxalic acid, ketene and amino acetaldehyde were also formed, which turn out to be important substrate in subsequent reactions.

The intermediate stage of formaldehyde formation can occur through several routes, depicted in the **Figure S1**. One pathway involves a series of elementary steps leading to carbene formation, with an energy barrier of 7.5 kcal/mol. An alternative route involves the reaction between carbon monoxide and water, presenting a barrier of 48.2 kcal/mol. It has been observed that glycolaldehyde **10**, a crucial intermediate in numerous pathways, is formed through a carbene-mediated pathway, exhibiting a low energy barrier as shown in figure. Alternatively, the AINR has identified an alternative pathway involving the formation of glycolaldehyde **10** from **3** and **8**.

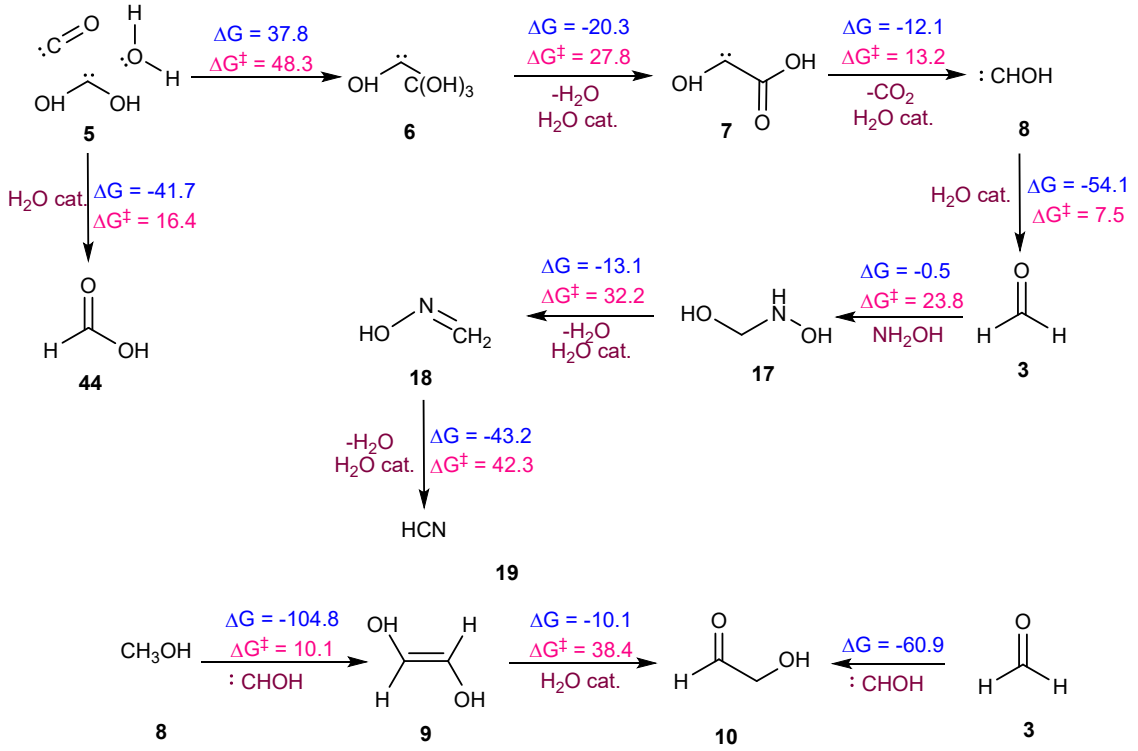


Figure S1: The sequence of elementary reaction steps derived from the AINR: the formation of formaldehyde, hydrogen cyanide and glycoaldehyde. Molecules labeled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/6-311++G (d,p) PCM (water). The values are in kcal/mol.

(ii) <u>AINR1 (12H2O+6CH3OH+6NH3+6CO+6CO2)</u>: Here, we observed the formation of both the Gly and Pyr in our simulation at 22 and 65 ps respectively, while setting the other

parameters to be k1 = 1.0 kcal mol^{-1}Å^{-2} (the force constant at the outer boundary), r1 = 13.0 Å, k2 = 0.5 kcal mol^{-1}Å^{-2} (the force constant at the inner boundary), r2 = 5.0 Å, τ = 0.75 ps, Total time between collisions = 1.0 ps. The details are discussed in main manuscript (**Figure 2**).

(iii) AINR2 ($6H_2O+8CH_3OH+12NH_3+7CO$): In this AINR simulation, we observed the formation of cyanide precursor (CH_3COCN) of pyruvic acid at 14 ps while setting the other parameters to be k1 = 1.0 kcal mol^{-1}Å^{-2} (the force constant at the outer boundary), r1 = 15.0 Å, k2 = 0.5 kcal mol^{-1}Å^{-2} (the force constant at the inner boundary), r2 = 4.9 Å, τ = 0.75 ps, Total time between collisions = 1.0 ps.

However, our investigation into pyruvic acid formation revealed an alternative pathway with different intermediates. Specifically, we observed the formation of intermediate **21** through the highly thermodynamically favourable reaction (ΔG=-144.2 kcal/mol) of singlet carbene **8** with methylidene. Intermediate **21** subsequently undergoes dehydration to form **23** via a six-membered, water-catalyzed transition state with a barrier of 7.5 kcal/mol. Intermediate **23** then reacts with **19** to form **24**, which hydrolyzes to yield pyruvic acid. The energy barriers for the hydrolysis of **23** are reasonable, as depicted in the **Figure S2**, and the reactions are facilitated by water through a six-membered transition state.

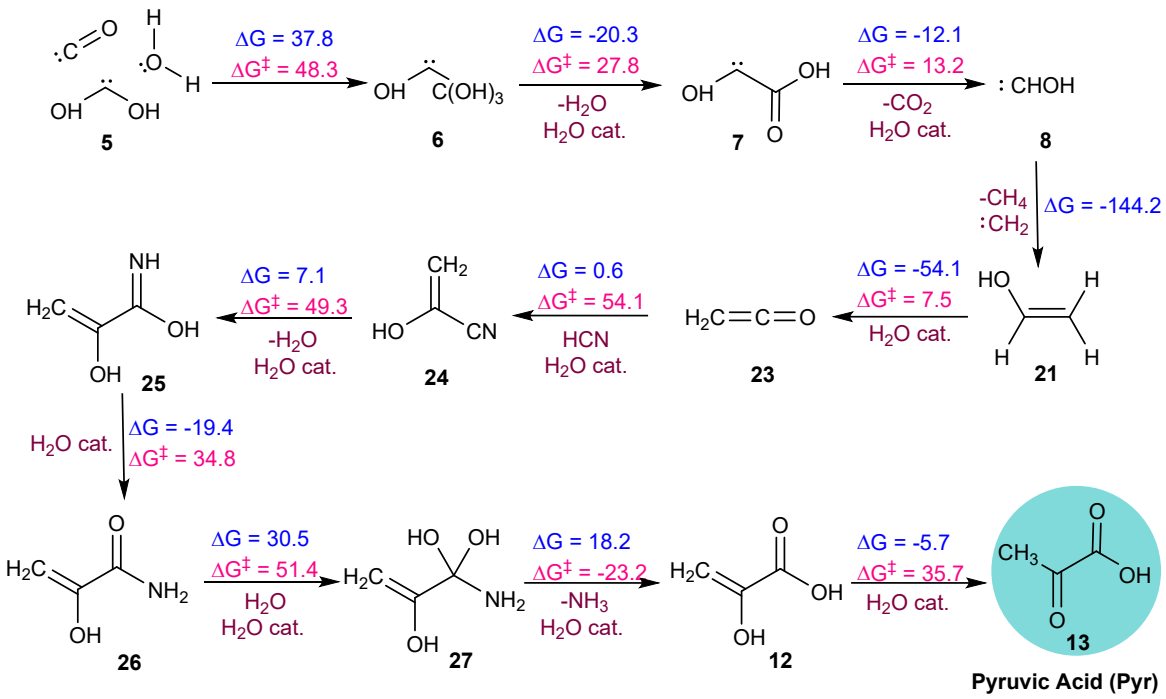


Figure S2: Sequence of elementary reaction steps derived from the AINR: the formation of Pyr from composition $6H_2O+8CH_3OH+12NH_3+7CO$. Molecules labelled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-

D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

(iv) AINR3 ($6H_2O+8CH_3OH+12NH_3+7CO+6CO_2$): In this AINR simulation, we employed a different composition of molecules resembling those found in interstellar ice. Our investigation revealed the exclusive formation of glyoxylic acid at **29** ps while setting the other parameters to be k1 = 1.0 kcal $mol^{-1}Å^{-2}$ (the force constant at the outer boundary), r1 = 14.0 Å, k2 = 0.5 kcal $mol^{-1}Å^{-2}$ (the force constant at the inner boundary), r2 = 5.4 Å, τ = 0.75 ps, Total time between collisions = 1.0 ps.

The pathway includes the formation of the carbonic acid intermediate **46**, which reacts with HCN to produce intermediate **47**, with an energy barrier of 46.4 kcal/mol. The hydrolysis of **47**, catalysed by water, results in intermediate **51**, with the concurrent release of ammonia. The transition state energy barriers, depicted in the figure 2, are reasonable. Subsequent dehydration of **51** yields intermediate **52**, a key participant in many biological reactions. Finally, the dehydrogenation of **52** leads to the formation of glyoxylic acid, encountering a slightly higher energy barrier of 60.0 kcal/mol.

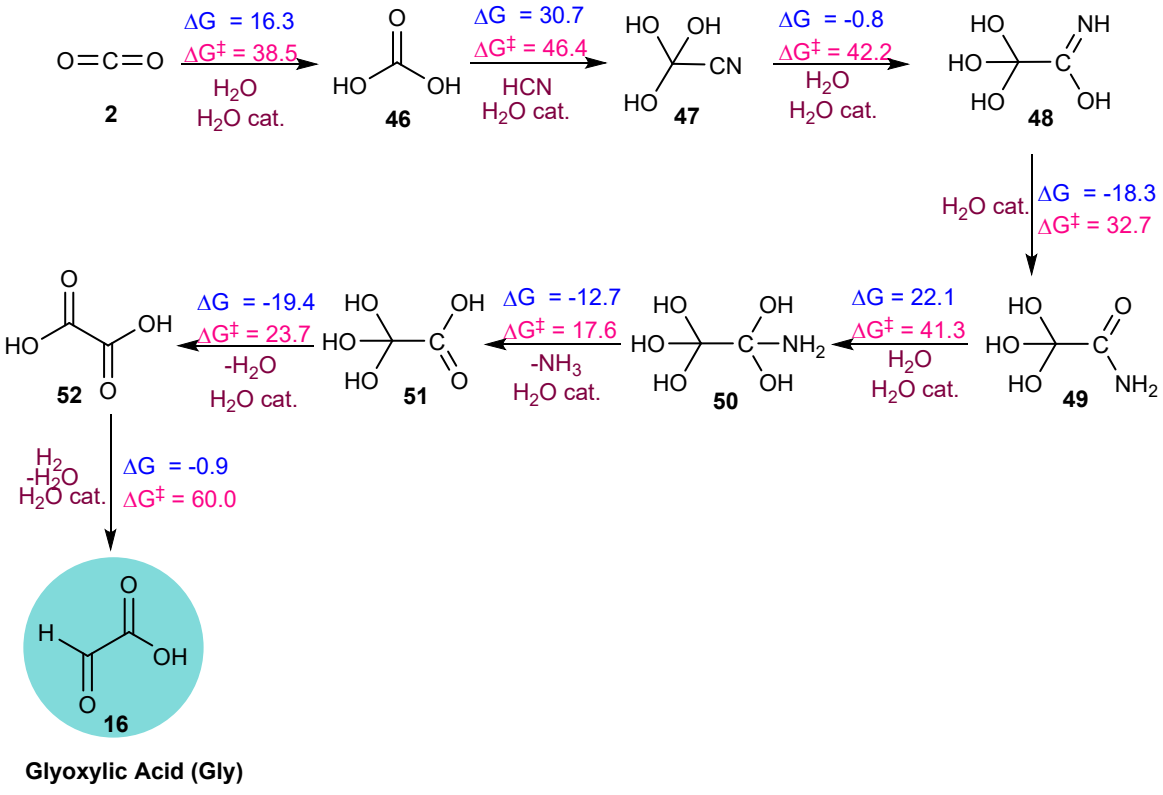


Figure S3: Sequence of elementary reaction steps derived from the AINR: the formation of Gly from composition $6H_2O+8CH_3OH+12NH_3+7CO+6CO_2$. Molecules labelled "cat.", shown

in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

(v) AINR4 ($9H_2O+9HCHO+10NH_3+7CO_2$): In this AINR simulation, we utilized a different composition of interstellar ice, incorporating additional molecules suggested by Robert et al. Specifically, we included HCHO in our composition to explore its effects on the simulation outcomes. Our findings indicate the exclusive formation of glyoxylic acid at 31 ps while setting the other parameters to be k1 = 1.0 kcal $mol^{-1}Å^{-2}$ (the force constant at the outer boundary), r1 = 13.0 Å, k2 = 0.5 kcal $mol^{-1}Å^{-2}$ (the force constant at the inner boundary), r2 = 4.9 Å, τ = 0.75 ps, Total time between collisions = 1.0 ps, following pathways involving various intermediates.

During the simulation, we observed the formation of hydrogen peroxide, which facilitates the oxidation process. The formaldehyde **3** reacts to form intermediate **44** via a peroxide intermediate **43**, with an energy barrier of 30.1 kcal/mol. This peroxide intermediate **43** then loses water to form **44** through a six-membered transition state, with an energy barrier of 36.8 kcal/mol, using water as a catalyst. Subsequently, the singlet carbene **5** reacts with intermediate **44**, leading to the formation of glyoxylic acid via the dehydration of intermediate **45**. The energy barriers for the dehydration process are reasonable, as shown in the **Figure S4**. This pathway highlights the critical role of hydrogen peroxide and formaldehyde in the prebiotic synthesis of glyoxylic acid under simulated interstellar conditions.

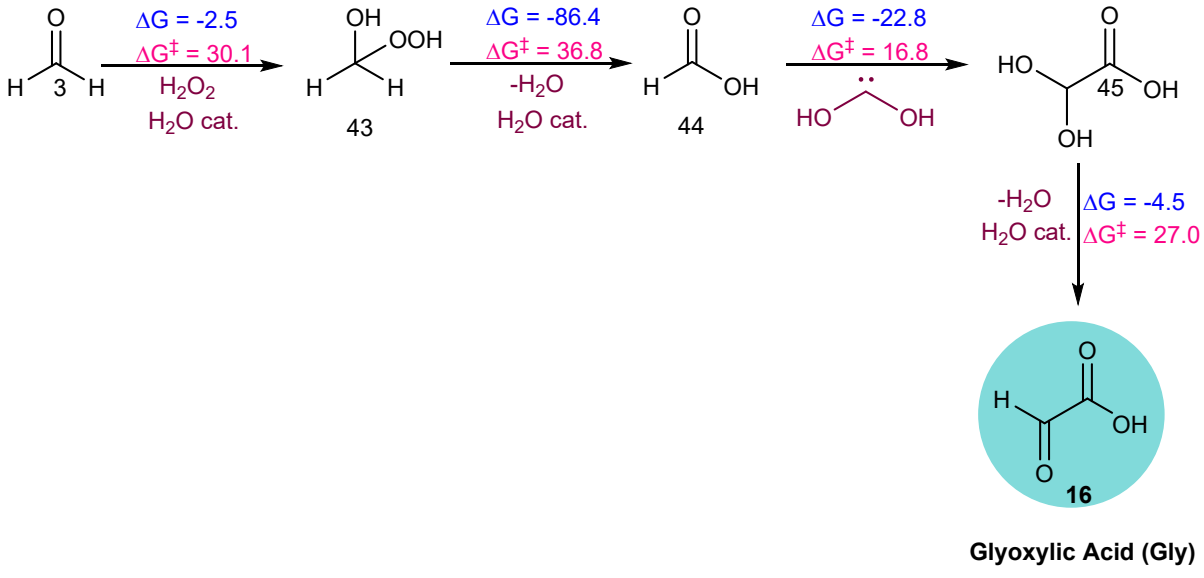


Figure S4: Sequence of elementary reaction steps derived from the AINR: the formation of Gly from composition $9H_2O+9HCHO+10NH_3+7CO_2$. Molecules labelled "cat.", shown in

brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

(vi) AINR5 ($4H_2O+8CH_3OH+10NH_3+8CO$): In this AINR simulation, our findings identified pyruvic acid at 29 ps, while setting the other parameters to be k1 = 1.0 kcal $mol^{-1}Å^{-2}$ (the force constant at the outer boundary), r1 = 12.0 Å, k2 = 0.5 kcal $mol^{-1}Å^{-2}$ (the force constant at the inner boundary), r2 = 4.1 Å, τ = 0.75 ps, total time between collisions = 1.0 ps accompanied by pathways featuring new important intermediates. The formation of amide **37** occurred through a series of elementary steps, originating from the generation of intermediate **19**. The hydrolysis of intermediate **19**, facilitated by water, led to the formation of intermediate **36** with a barrier energy of 48.3 kcal/mol. Subsequently, tautomerization processes ensued, resulting in the formation of amide **37** with an energy barrier of 36.5 kcal/mol.

Additionally, a notable reaction pathway involved the formation of acetaldehyde **39**, from singlet carbene **38**, which exhibited high feasibility (ΔG = -51.3 kcal/mol) and a barrier energy of 7.4 kcal/mol. This acetaldehyde then engaged with amide **37**, undergoing successive hydrolysis and dehydration reactions leading to the formation of pyruvic acid (**Figure S5**). Notably, the final step in this pathway involved a dehydrogenation reaction characterized by a relatively higher energy barrier.

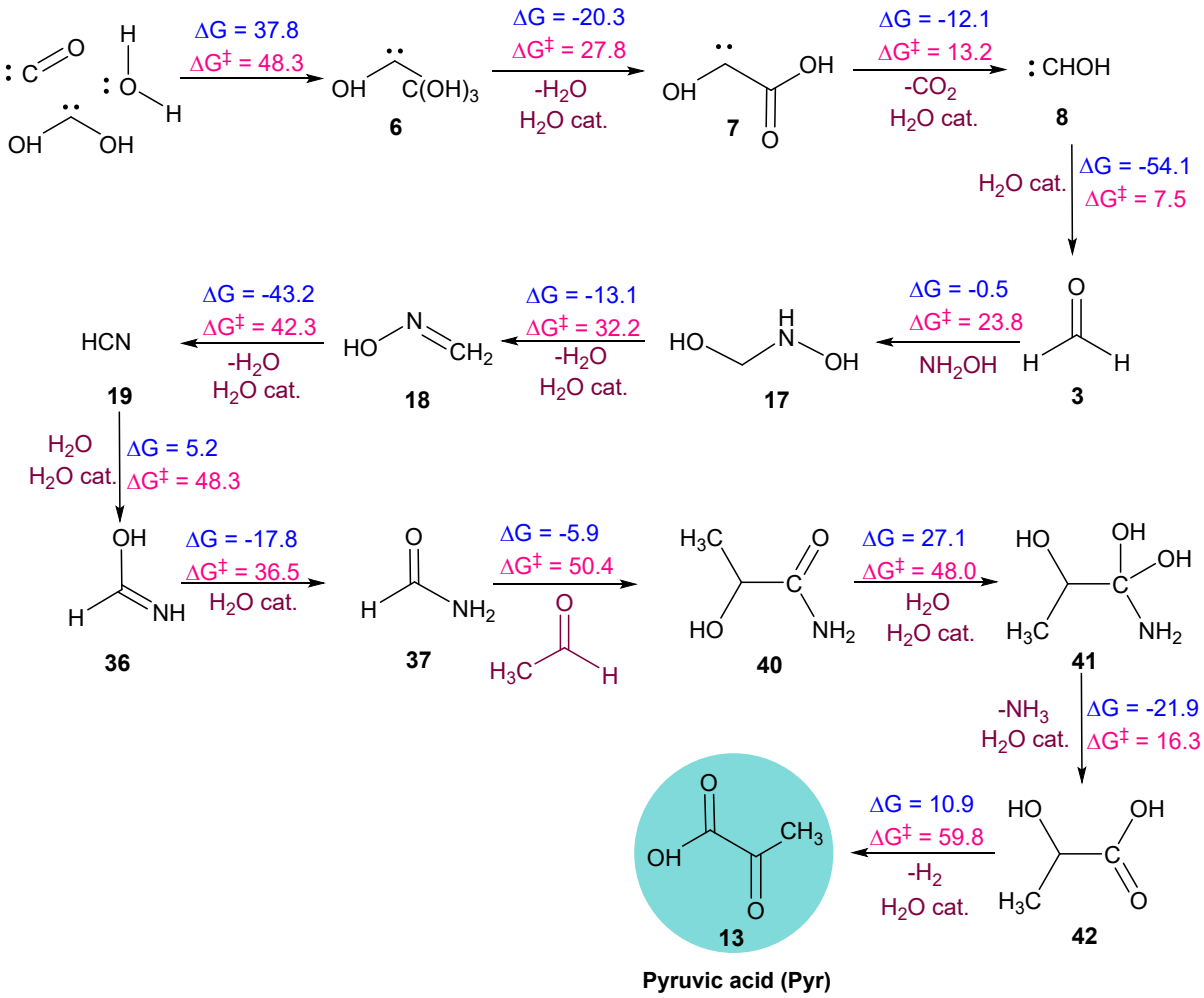


Figure S5: Sequence of elementary reaction steps derived from the AINR: the formation of Pyr from composition $4H_2O+8CH_3OH+10NH_3+8CO$. Molecules labelled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

(vii) <u>AINR6 ($11CH_3OH+12NH_3+8CO$):</u> In this AINR simulation, we observed the formation of pyruvic acid and its diol isomer. The formation of pyruvic acid is detected at 14 ps while setting the other parameters to be k1 = 1.0 kcal $mol^{-1}Å^{-2}$ (the force constant at the outer boundary), r1 = 13.0 Å, k2 = 0.5 kcal $mol^{-1}Å^{-2}$ (the force constant at the inner boundary), r2 = 4.7 Å, τ = 0.75 ps, Total time between collisions = 1.0 ps. The details are discussed in the manuscript (**Figure 4**).

(viii) Pathway for the formation of carbamic acid **53** (TS 51), glyceraldehyde **55** (TS 49) and isocyanic acid **54** (TS 50) in AINR simulations as shown in **Figure S6.**

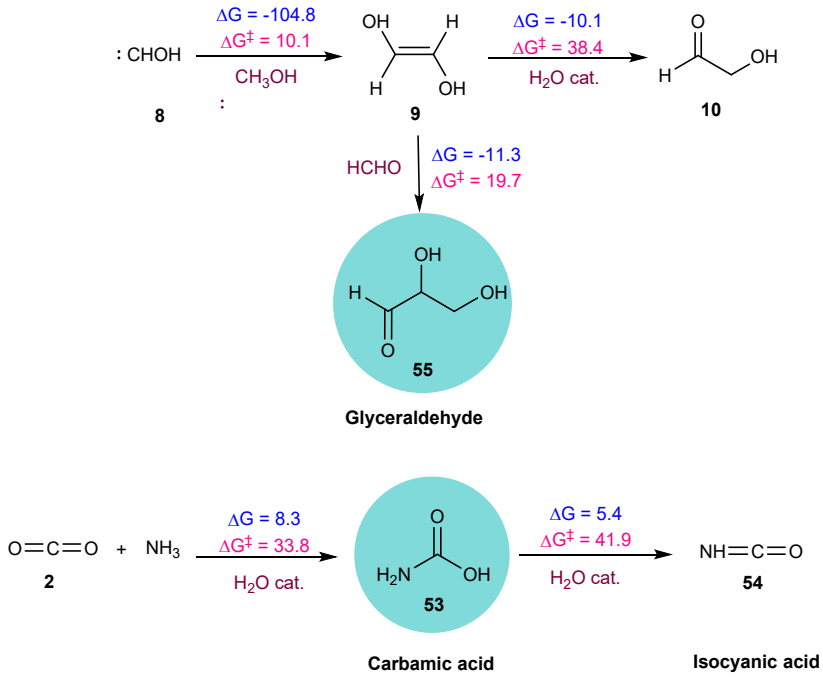


Figure S6: Sequence of elementary reaction steps derived from the AINR for the formation of carbamic acid, glyceraldehyde and isocyanic acid. Molecules labelled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

(ix) Another important parameter in the AINR simulations is the spherical boundary condition, which we have discussed in the "AINR Spherical Boundary Conditions" subsection in the Computational Details section here in the SI. Optimizing the boundary in the AINR is a trial and error process. Our objective was to establish the boundary conditions to ensure effective collisions among molecules within the inner sphere and provide ample space for relaxation in the outer sphere. It was crucial to prevent collisions from occurring at velocities high enough to dissociate molecules into their elemental forms.

(x) Temperature is another important parameter in the AINR simulations. The temperature that we have used in our simulations (2000 K) is not the actual reaction temperature at which the reactions would occur. The reason that we have provided such a high temperature is to avoid noncovalent interactions such as hydrogen bonding in our AINR simulation and also to provide enough kinetic energy to the molecules so that they could collide with each other and cross the activation barriers, leading to the products. We reiterate that the purpose of the AINR is to act

as a tool for discovery of chemical reactions, the feasibility of which could then be determined with careful, high level QM (DFT, MP2 and CC2) studies of the thermodynamics and kinetics of the discovered reactions. Therefore, the parameters (ratio of the reactants, spherical boundary conditions and temperature) that would lead to the best possibility of discovering new processes have to be employed, regardless of whether they necessarily represent the actual experimental conditions or not.

(xi) The goal of the AINR simulations is to find new reactions and mechanistic pathways, and not to equilibrate the systems. We have run most of the simulations at a ~1 ns timescale, which sufficed to yield different interesting intermediates and products, as well as the corresponding mechanistic pathways.

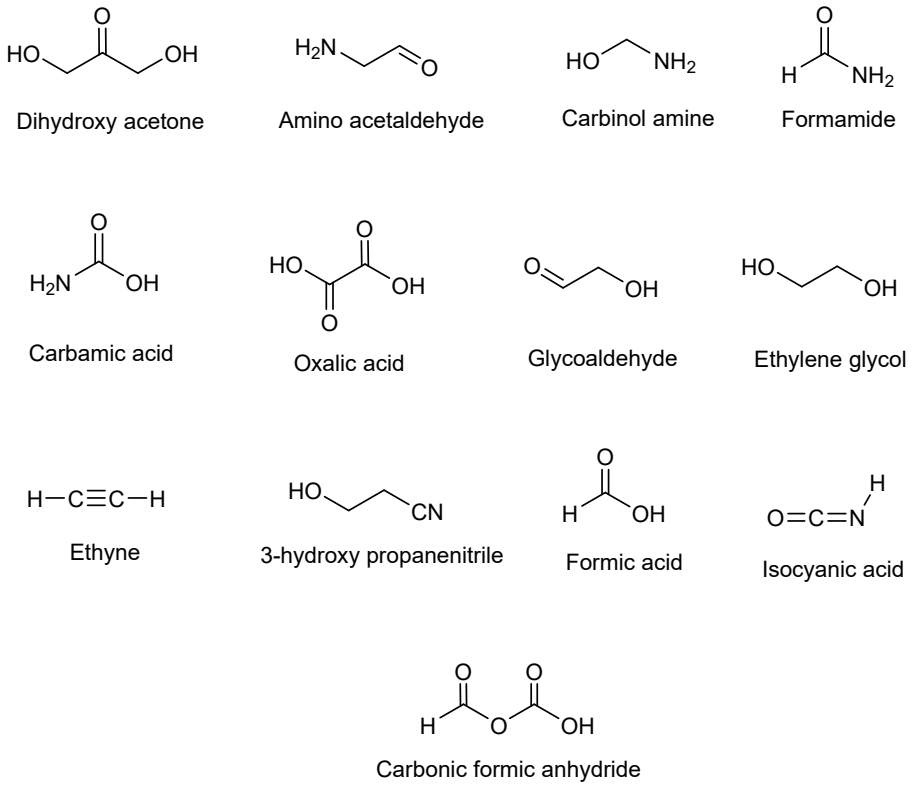


Figure S7: A selection of the products that were discovered from the *ab initio* nanoreactor simulations, including ribonucleotide and amino acid precursors and other intermediate compounds were also formed during the simulations.

Energy profile of the pathways for formation of glyoxylic and pyruvic acid:

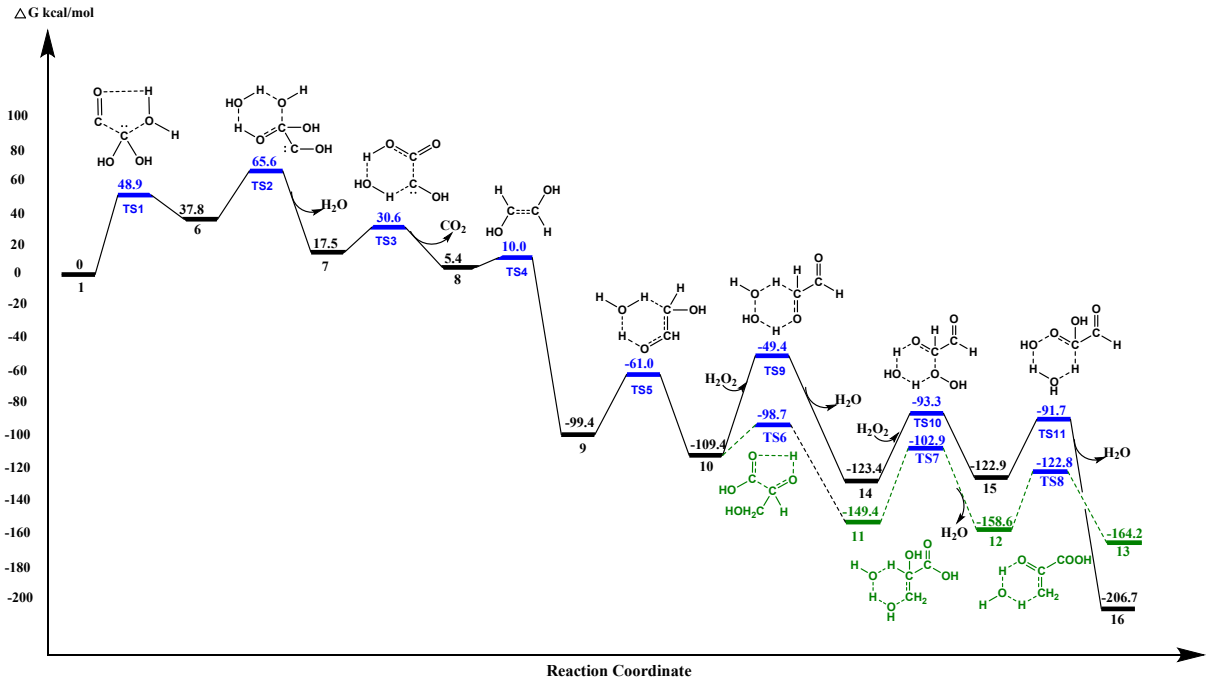


Figure S8: The reaction free energy profile diagram for the formation of Gly (represented by black lines) and Pyr (represented by green lines) for the pathway $12H_2O+6CH_3OH+6NH_3+6CO+6CO_2$ (AINR1). Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using the Gaussian09 software package.[1] The values are in kcal/mol.

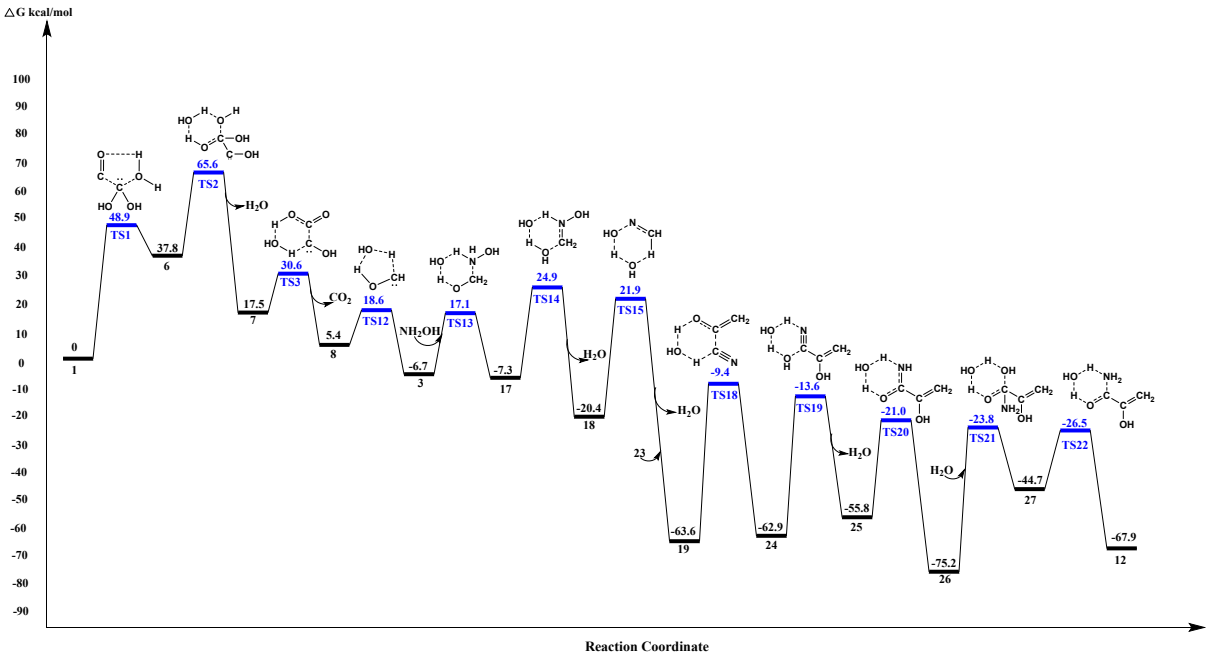


Figure S9: The reaction free energy profile diagram for the formation of pyruvic acid enol for the pathway $6H_2O+8CH_3OH+12NH_3+7CO$ (AINR2). Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using the Gaussian09 software package.[1] The values are in kcal/mol.

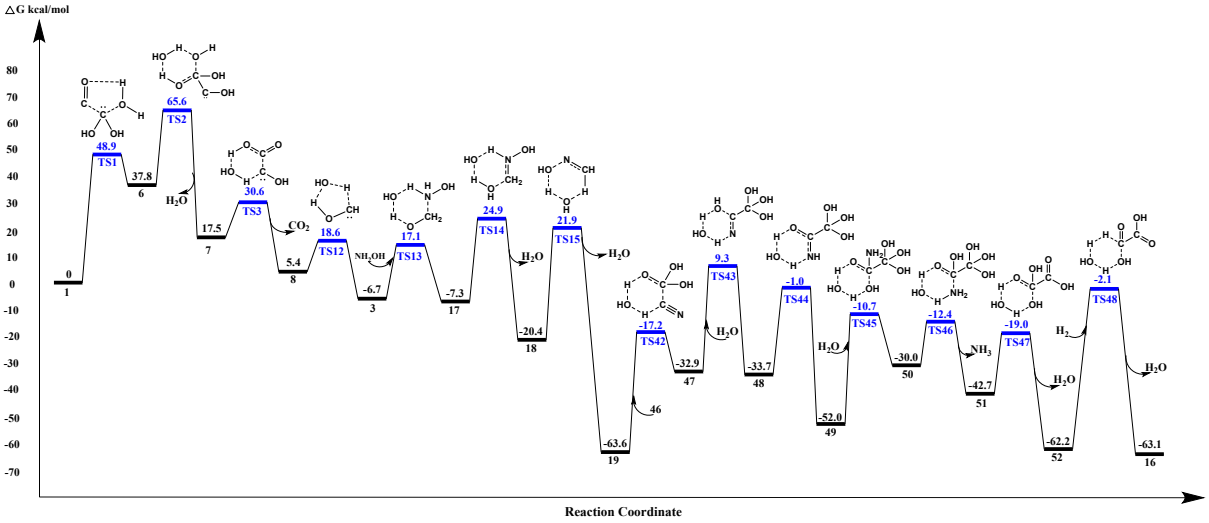


Figure S10: The reaction free energy profile diagram for the formation of Gly for the pathway $6H_2O+8CH_3OH+12NH_3+7CO+6CO_2$ (AINR3). Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

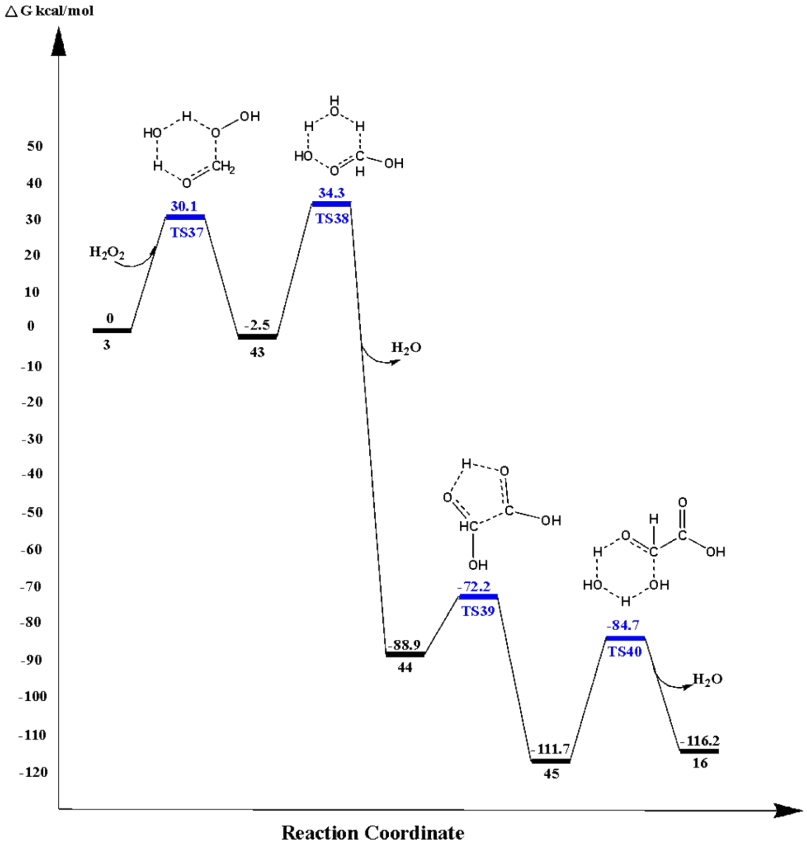


Figure S11: The reaction free energy profile diagram for the formation of Gly for the pathway $9H_2O+9HCHO+10NH_3+7CO_2$ (AINR4). Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

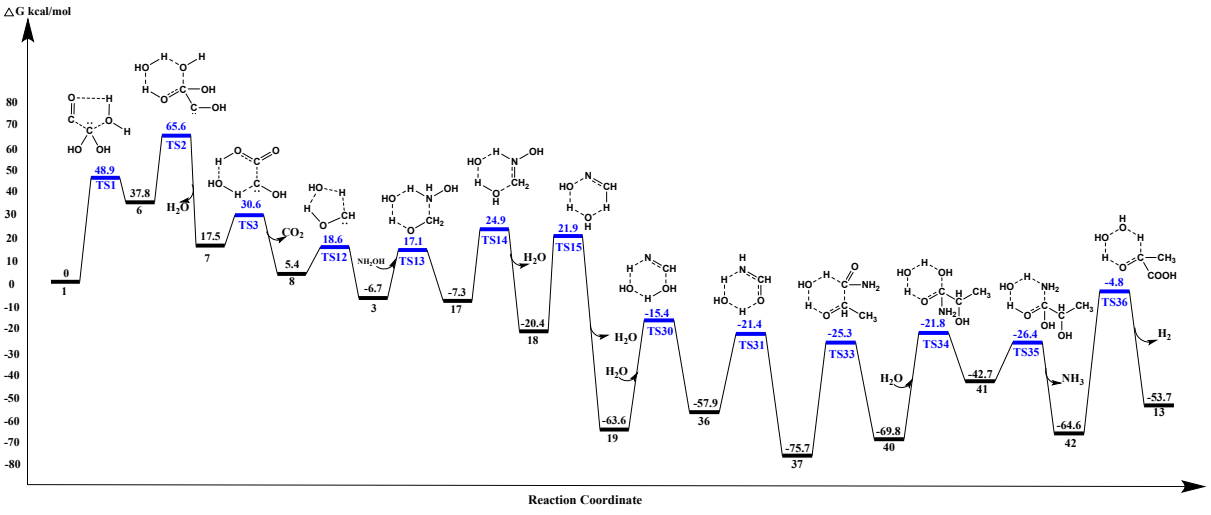


Figure S12: The reaction free energy profile diagram for the formation of Pyr for the pathway $4H_2O+8CH_3OH+10NH_3+8CO$ (AINR5). Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

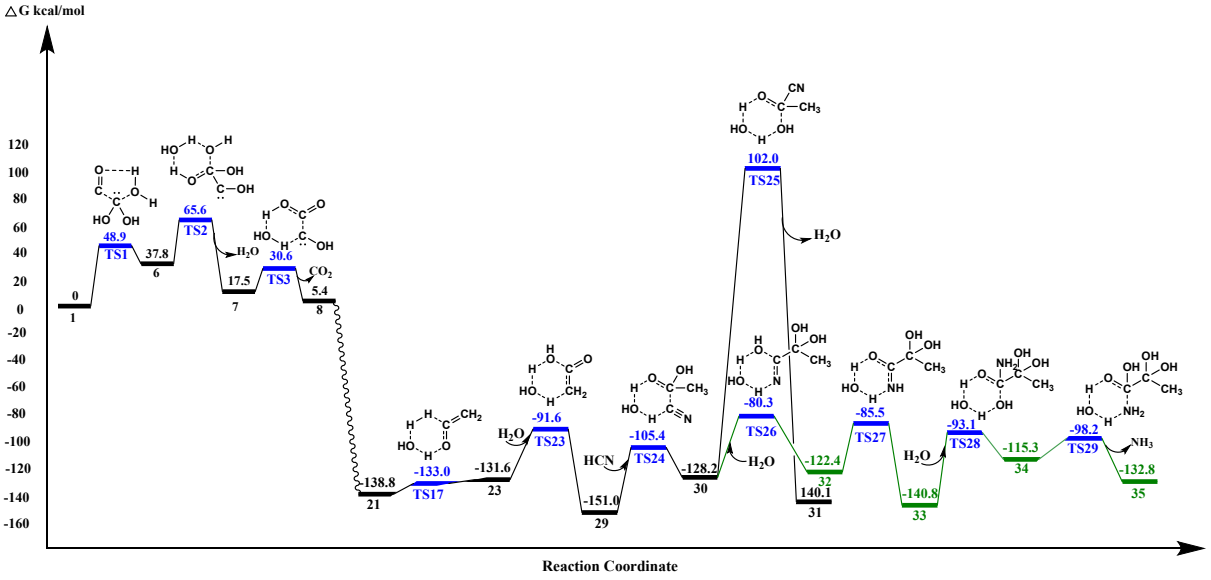
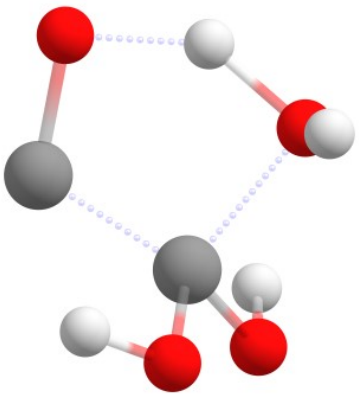


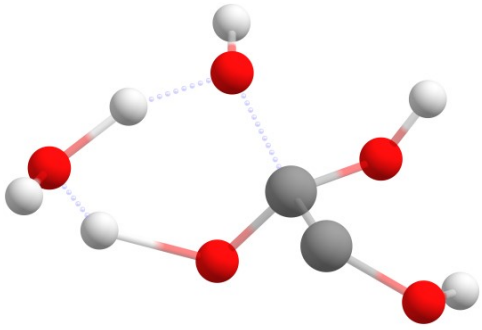
Figure S13: The reaction free energy profile diagram for the formation of Pyr for the pathway $11CH_3OH+12NH_3+8CO$ (AINR6). Values have been calculated at the B3LYP-D3/6-311++G(d,p)+PCM(water) using Gaussian09 software package.[1] The values are in kcal/mol.

3D structures of the transition states

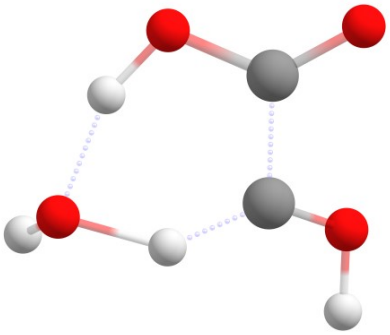
TS1



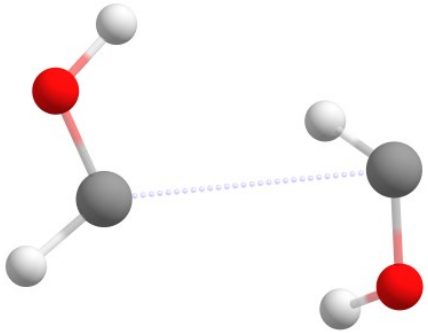
TS2



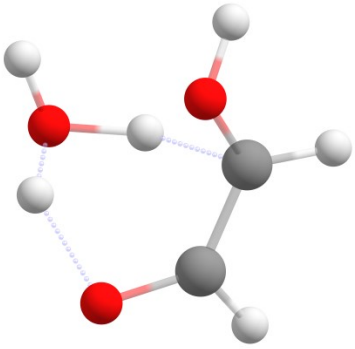
TS3



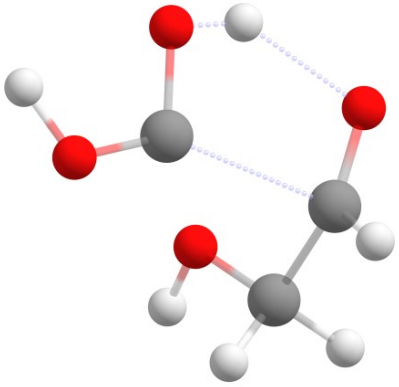
TS4



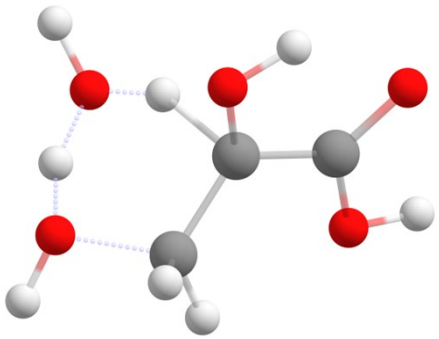
TS5



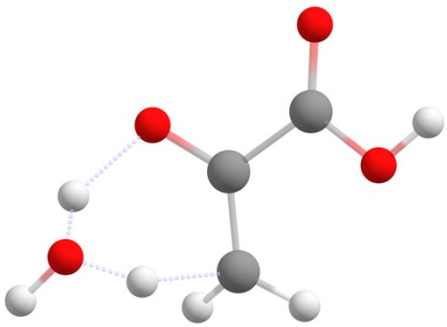
TS6



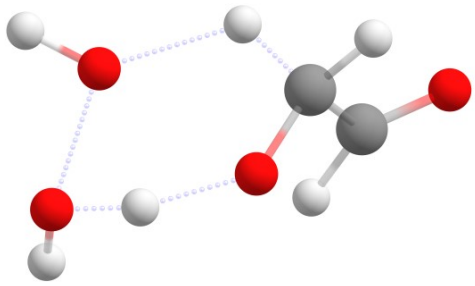
TS7



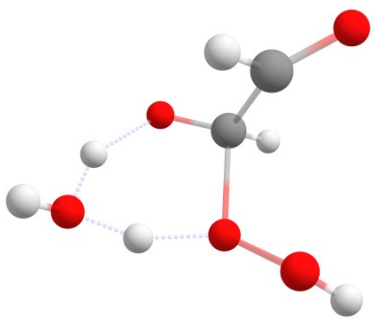
TS8



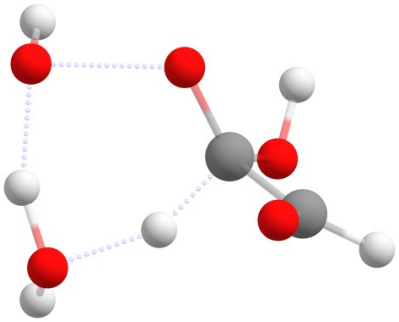
TS9



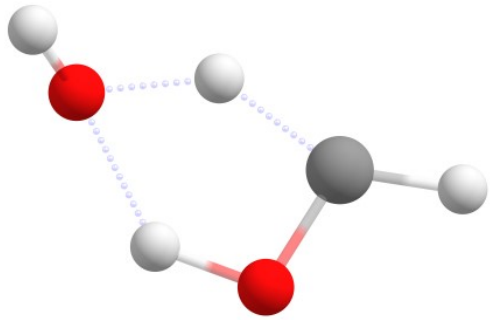
TS10



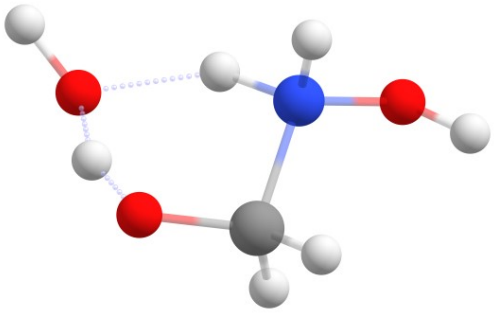
TS11



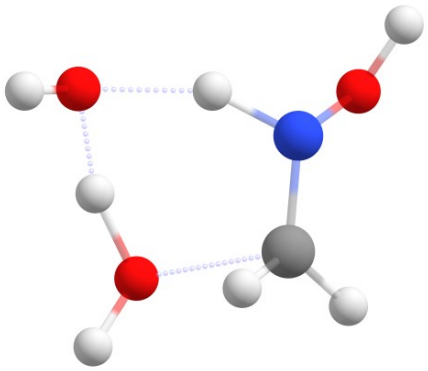
TS12



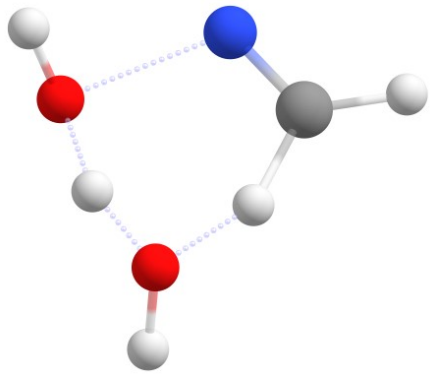
TS13



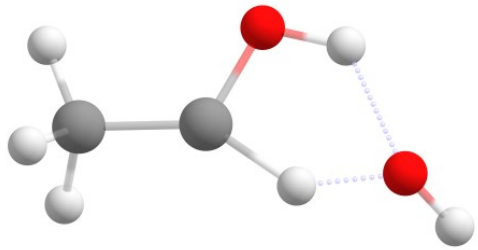
TS14



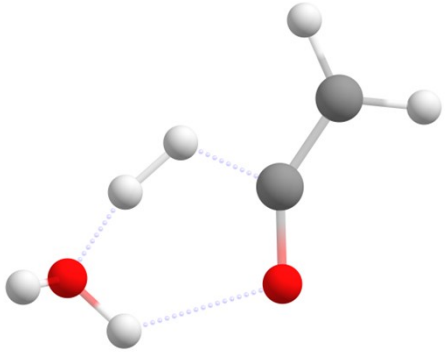
TS15



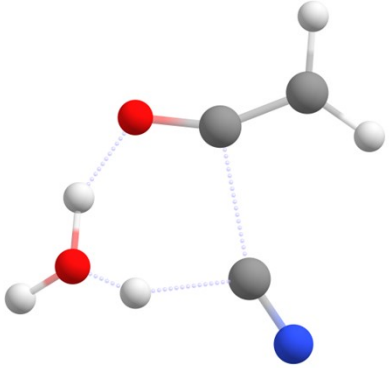
TS16



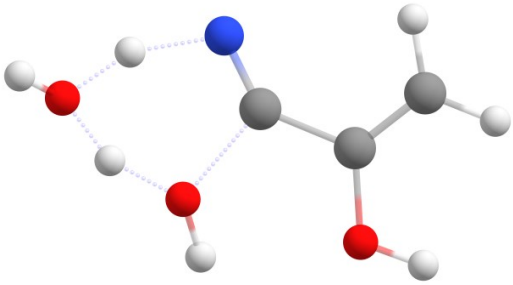
TS17



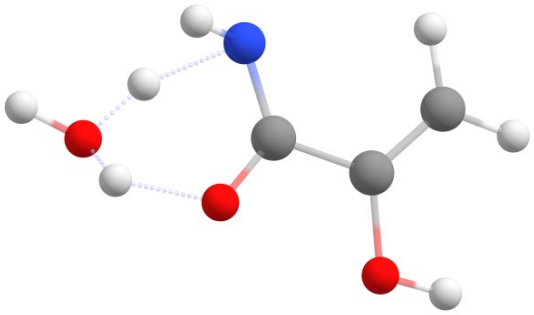
TS18



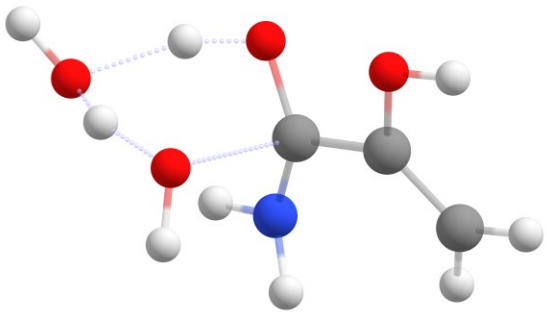
TS19



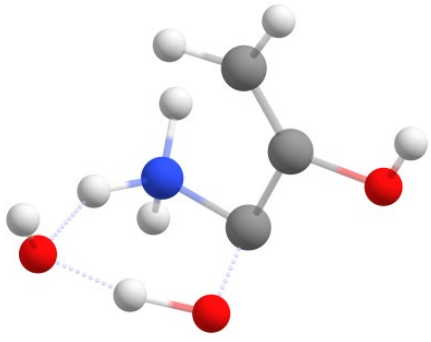
TS20



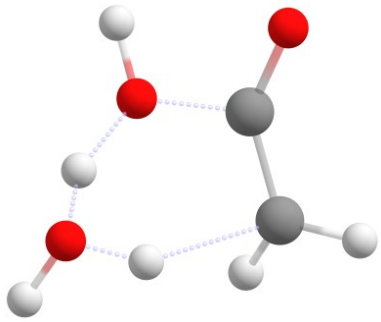
TS21



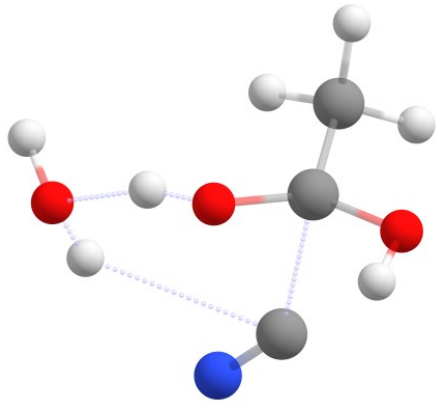
TS22



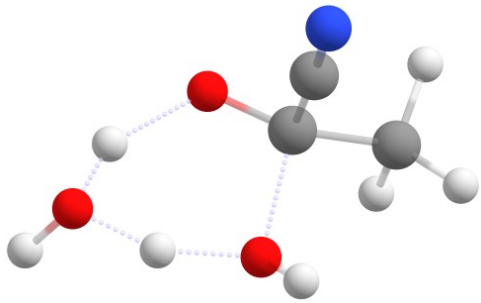
TS23



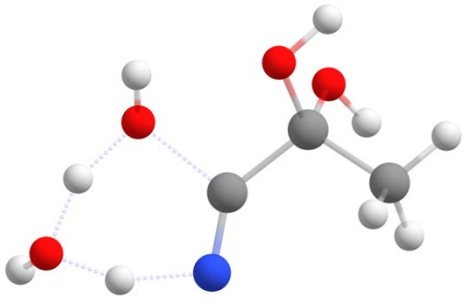
TS24



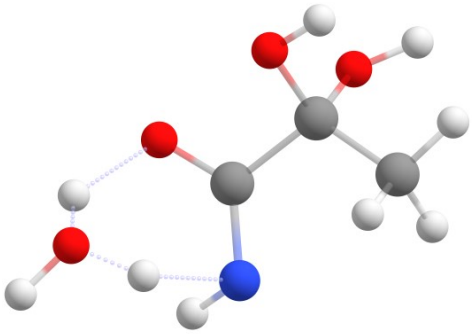
TS25



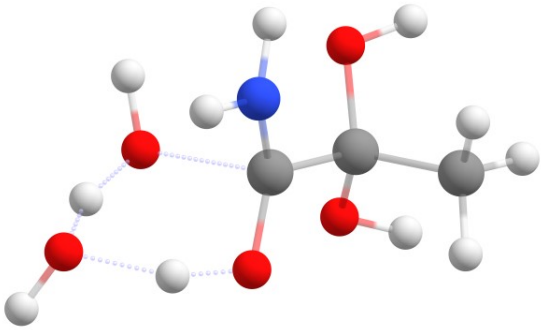
TS26



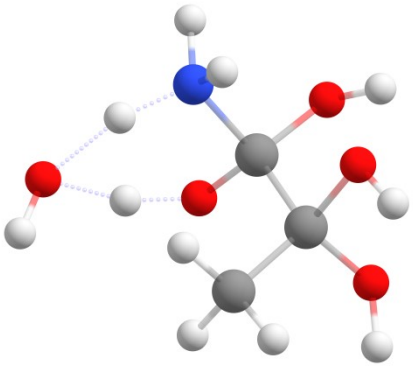
TS27



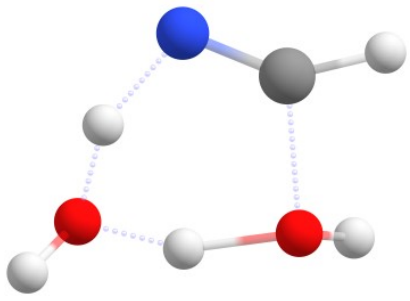
TS28



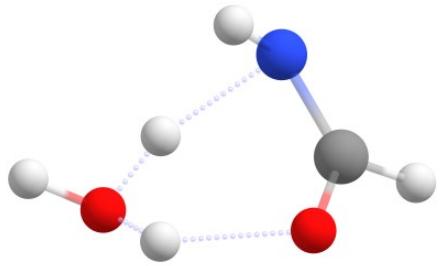
TS29



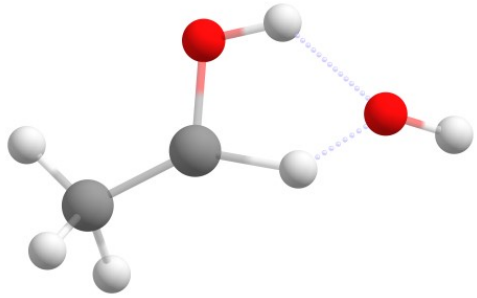
TS30



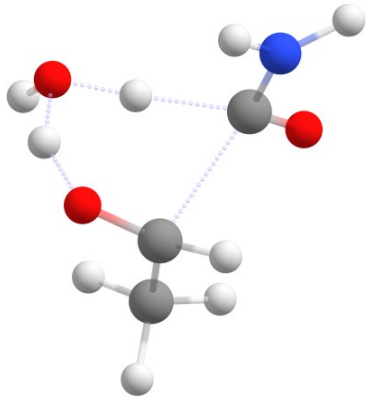
TS31



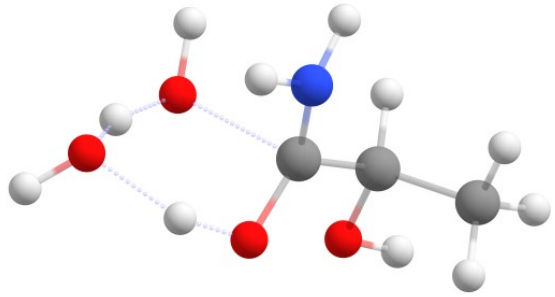
TS32



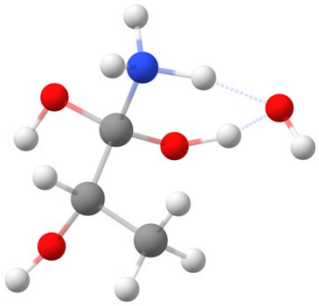
TS33



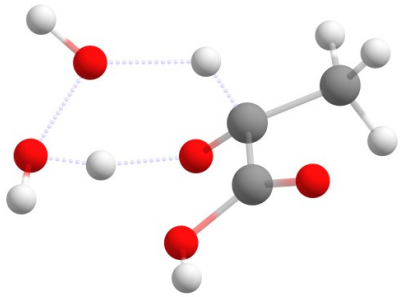
TS34



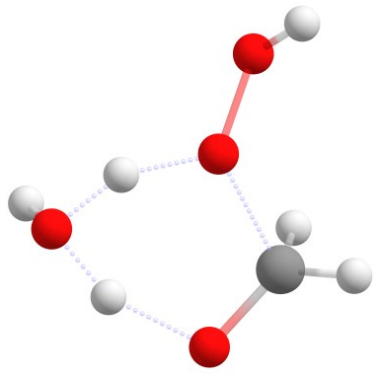
TS35



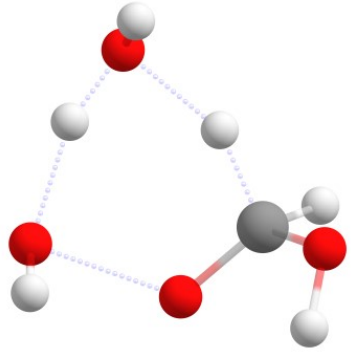
TS36



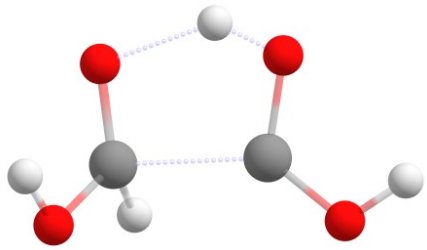
TS37



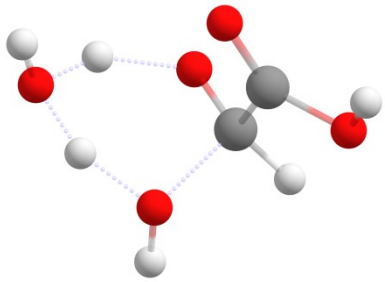
TS38



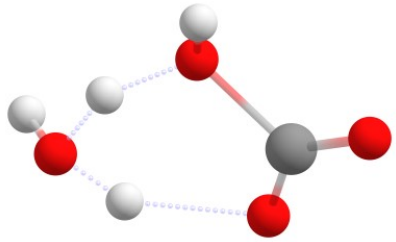
TS39



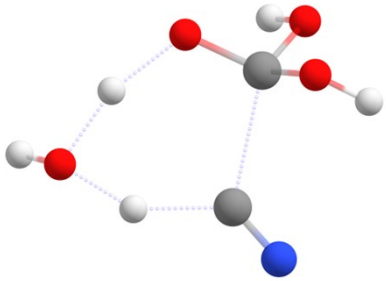
TS40



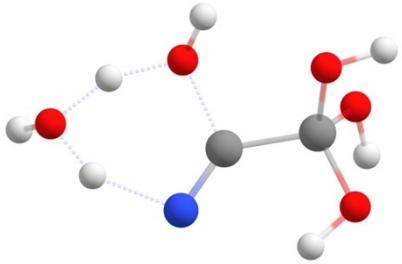
TS41



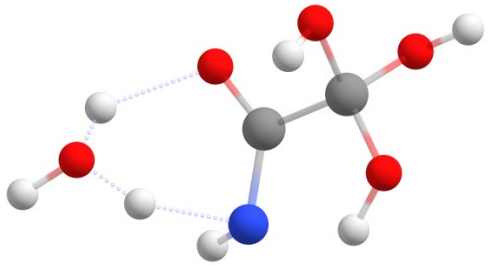
TS42



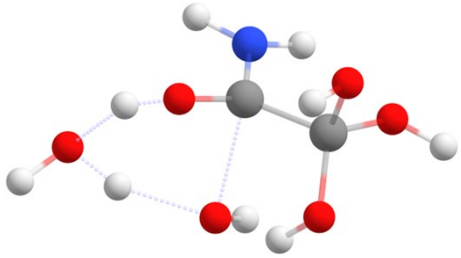
TS43



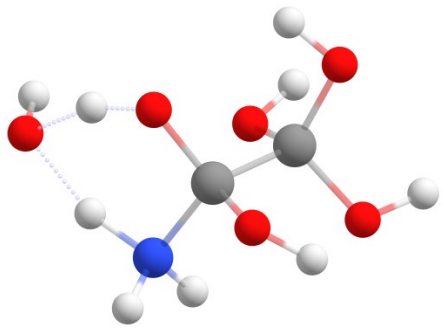
TS44



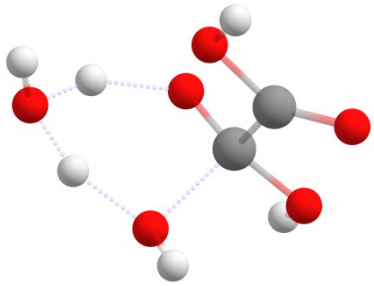
TS45



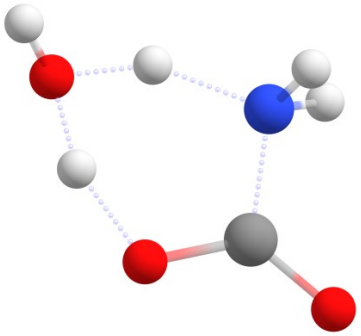
TS46



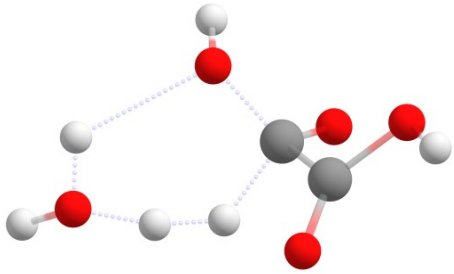
TS47



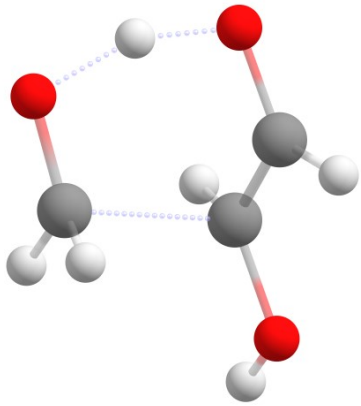
TS51



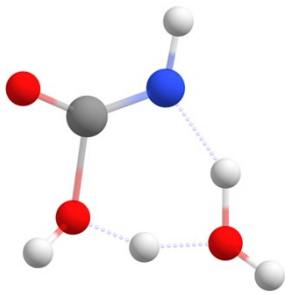
TS48



TS49



TS50



References

S1. Gaussian 09, Revision E.01; Frisch , M. J.; Trucks, G. W.; Schlegel, H. B.;Scuseria,G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.;Jr.,Peralta, J. E.; Ogliaro, F.,Bearpark, M.; Heyd, J. J. E.; Brothers,Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Raghavachari, J.K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

S2. McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18 J. Chem. Phys. 1980, 72, 5639-5648.

S3. Becke, A. D. Density functional thermochemistry. III. The role of exact exchange J. Chem. Phys. 1993, 98, 5648–5652.

S4. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models Chem. Rev.2005, 105, 2999-3094.

S5. Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. J. Chem. Theory Comput. 2009, 5, 2619−2628.

S6. Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge transfer and polarization in solvated proteins from *ab initio* molecular dynamics. J. Phys. Chem. Lett. 2011, 2, 1789−1793.

S7. Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Excited-state electronic structure with configuration interaction singles and Tamm−Dancoff time dependent density functional theory on graphical processing units. J. Chem. Theory Comput. 2011, 7, 1814−1823.

S8. Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating efficient quantum chemistry codes for novel architectures. J. Chem. Theory Comput. 2013, 9, 213−221.

S9. Ufimtsev, I. S.; Martinez, T. J. Graphical processing units for quantum chemistry. Comput. Sci. Eng. 2008, 10, 26−34.

S10. Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. J. Chem. Theory Comput. 2008, 4, 222−231.

S11. Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation. J. Chem. Theory Comput. 2009, 5, 1004−1015.

S12. Froese, F. C. General Hartree-Fock program. Comput. Phys. Commun. 1987, 43, 355−365.

S13. Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-Consistent molecular orbital methods. 21. Small split-valence basis sets for firstrow elements. J. Am. Chem. Soc. 1980, 102, 939−947.

S14. Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. Discovering chemistry with an *ab initio* nanoreactor. Nat. Chem 2014, 6, 1044-1048.

S15. Feller, D.; Peterson, K. A. An examination of intrinsic errors in electronic structure methods using the Environmental Molecular Sciences Laboratory computational results database and the Gaussian2 set. J. Chem. Phys. 1998, 108, 154−176.