# A common precursor to both prebiotic and biological pathways to RNA nucleotides

Andrea Pérez-Villa, Thomas Georgelin, Jean-François Lambert, Marie-Christine Maurel, François Guyot, A. Marco Saitta, Fabio Pietrucci

# A common precursor to both prebiotic and biological pathways to RNA nucleotides

Andrea Pérez-Villa[1], Thomas Georgelin[2,3], Jean-François Lambert[2], Marie-Christine Maurel[4], François Guyot[1], A. Marco Saitta[1*], Fabio Pietrucci[1*]

[1]Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IRD, Institut de Minéralogie, de Physique des Matriaux et de Cosmochimie, IMPMC, F-75005 Paris, France

[2]Sorbonne Université, CNRS UMR 7197, Laboratoire de Réactivité de Surface, LRS, F-75005 Paris, France

[3]Centre de Biophysique Moléculaire, CNRS UPR 4301, F-45071 Orléans, France

[4]Sorbonne Université, Muséum National d'Histoire Naturelle, CNRS MNHN UMR 7205, Institut de Systématique, Evolution, Biodiversité, ISyEB, F-75005 Paris, France

*Corresponding authors. E-mail: marco.saitta@sorbonne-universite.fr, fabio.pietrucci@sorbonne-universite.fr

**Abstract**

Understanding the mechanism of spontaneous formation of ribonucleotides under realistic prebiotic conditions is a key open issue of origins-of-life research. In cells, *de novo* and salvage nucleotide enzymatic synthesis combines 5-phospho-$\alpha$-D-ribose-1-diphosphate ($\alpha$-PRPP) and nucleobases. Interestingly, these reactants are also known as prebiotically plausible compounds. Combining ab initio simulations with mass spectrometry experiments, we compellingly demonstrate that nucleobases and $\alpha$-PRPP spontaneously combine, through the same facile mechanism, forming both purine and pyrimidine ribonucleotides, under mild hydrothermal conditions. Surprisingly, this mechanism is very similar to the biological one, and yields ribonucleotides with the same anomeric carbon chirality as in biological systems. These results suggest that natural selection might have optimized – through enzymes – a pre-existing ribonucleotide formation mechanism, carrying it forward to modern life forms.

# Introduction

Exploring the synthesis and functionality of nucleic acid molecules is one of the key issues in the origins-of-life research. Since the late 80's, scientists have proposed (*1*) the possibility of a 'RNA world' as the precursor of current biochemistry, a hypothesis strongly supported by the high versatility of RNA molecules to perform different biochemical processes, such as catalysis, gene regulation, storage of information and self-replication, that are fundamental features involved in the development and propagation of life. A plausible 'RNA World' implies sufficiently successful synthesis of the RNA monomers, namely ribonucleotides (or simply nucleotides), followed by their accumulation, their subsequent polymerization and their further propagation and persistence in the environment of occurrence (*2, 3*). All these steps are affected by a great number of factors (temperature, presence of catalytic surfaces, pH, ionic strength, and so forth), and their plausibility is mainly determined by the thermodynamics and kinetics of the chemical reactions involved.

In the last decades, many efforts have been made in addressing some of these questions. Nucleotides are composed of three chemical subunits: a ribose in its $\beta-$furanose conformation, a nucleobase (purine or pyrimidine) and a phosphate group at the 5' position of the sugar. One of the most challenging issues has been the synthesis of both purine and pyrimidine nucleotides. Different prebiotic setups have successfully yielded nucleobases (*4–6*), sugars and their phosphate derivatives (*7, 8*); however, the assembly of these subunits to constitute the first nucleotides has presented several obstacles (*2, 9*).

Several strategies to produce nucleotides under prebiotic conditions have been assessed. Among them, the nucleoside phosphorylation has been one of the most widely studied. From previous works, pyrimidine and purine nucleosides were phosphorylated after a treatment with different phosphate minerals (*10*), or in phosphate solutions with the addition of phosphorylating agents such as urea (*11*), another prebiotically relevant substrate. In 2009, Powner *et al.* (*12*) proposed an elegant synthetic pathway that involved the reaction between prebiotically relevant non-canonical subunits of the nucleotides that led to pyrimidine nucleotides. Along very different synthetic routes, experiments performed by Becker *et al.* (*13*) in 2016, successfully explored purine nucleoside synthesis following the formamidopyrimidine (FaPy) pathway. A coherent prebiotic picture, encompassing the formation of both purine and pyrimidine nucleotides in similar conditions has been one of the largest current challenges in the origin of life field. In this regard, two recent works by Stairs *et al.* (*14*) and Kim and Benner (*15*) addressed the synthesis

of both types of ribonucleotides along different chemical routes.

In current living organisms and metabolic systems, the *de novo* and salvage synthesis of purine and pyrimidine nucleotides implies a direct reaction between the nucleobase and a molecule of PRPP, a biological metabolite highly activated in C1' due to the presence of the pyrophosphate group. This reaction is mediated by the enzymatic catalysis of phosphoribosyltransferases specific of each nucleobase (*16, 17*). Primitive biochemical systems have evolved to modern ones improving their performance and selectivity: from this point of view, an interesting question is to understand whether features observed in contemporary enzymatic reactions have been conserved from prebiotic non-enzymatic pathways (chemiomimesis). In particular, one can wonder whether PRPP also played a crucial role as precursor of nucleotides in the context of a RNA world (*18*). In this work we examined this hypothesis (Fig. 1), which is supported by the possibility to generate PRPP as a ribose phosphorylation product (*19, 20*), and by its plausibility to operate as a reactant toward nucleotides (*20*) and nucleotide mimics (ribose-pyrazole derivatives) (*21*). We considered mild hydrothermal conditions, that are compatible with different proposed origin-of-life scenarios, including hot surface lagoons/lakes and submarine or continental hydrothermal systems (*22*). We underline that we deliberately choose the simplest possible setting that allows to prove the general thermodynamic viability of the ribonucleotide synthesis reactions.

To address this crucial question, a combination of quantum-based computer simulations and experiments was carried out. *Ab initio* computer simulations in prebiotic chemistry under aqueous solutions (*23*) are very challenging due to the large system size (several hundred atoms), the bulk conditions that have to be taken into account (such as temperature, pH, ionic strength, and so forth), and especially because of the long time scale of chemical reactions. To overcome the latter obstacle, we exploited a new general approach able to simulate complex reaction mechanisms in solution, combining enhanced sampling techniques (such as metadynamics and umbrella sampling (*24–26*)) with topological coordinates capable of tracking changes in the chemical bond network (*27*). After elucidating the reaction mechanism, we could extract useful information about the environmental conditions (temperature and pH) and free energy barriers, which were further exploited to successfully test – by means of Mass Spectrometry (MS) – the formation of nucleotides in experimental conditions matching those of simulations.

# Results and discussion

## *Ab initio* simulation of nucleotide synthesis

We have assessed the formation of uridine monophosphate (UMP) and adenosine monophosphate (AMP) starting from $\alpha$-PRPP and uracil or adenine, respectively, simulating easily-accessible hydrothermal/hot lagoon conditions (T = 400 K, neutral pH), in pure water without any catalyst. Our protocol employs a preliminary exploration of the reaction pathways with metadynamics, followed by the collection of extended statistics exploiting committor analysis and umbrella sampling (see Materials and Methods).

We have discovered a simple pathway for the association of PRPP with the nucleobase, where the nitrogen forming the glycosidic bond is deprotonated. In both purine and pyrimidine cases, the reaction follows a nucleophile substitution $S_N2$ mechanism, with an inversion of chirality of the anomeric carbon. The free energy profile of the reaction is reported in Fig. 2, together with the detailed atomic configurations at key moments. The transition state exhibits a trigonal bipyramidal geometry, where C1' becomes planar with O and N nucleophiles from pyrophosphate and nucleobase molecules at a distance of 2.08 and 2.40 Å, respectively, for the pyrimidine, and 2.12 and 2.84 Å for the purine. The reaction connecting the deprotonated nucleobase with the nucleotide follows Hammond's postulate, *i.e.*, the transition state is more similar to the reactants than to the products, since the forward barrier is smaller than the reverse one. When comparing the biological synthesis mediated by enzymes with the simulated prebiotic reactions, it is observed that some mechanistic details are featured in both cases: for instance the nucleobase has to be deprotonated so it can perform the nucleophilic attack on the C1' (*28, 29*); moreover around the transition state region, the C1' adopts a planar geometry during the nucleophilic attack by the base, necessary to undergo the inversion of the chirality when forming the $\beta$-nucleotide. Overall, our results suggest that a mildly basic pH (below 9, to avoid degradation of PRPP), and an excess of the nucleobase could promote the reaction, increasing the probability of reactive collisions between a deprotonated nucleobase and the PRPP. These are the conditions we therefore adopted in our experiments.

## Mass spectrometry of PRPP and nucleobase samples

Experiments were carried out in order to confirm the formation of nucleotides from nucleobases and $\alpha$-PRPP, as suggested by calculations. Samples of uracil or adenine in presence of PRPP were prepared under conditions similar to the *ab initio* molecular dynamics simulations and

subsequently analyzed by Electrospray Ionization (ESI) MS. Fig. 3 and 4 show the negative-mode ESI spectra of solutions containing PRPP + U, and PRPP + A, respectively, after heating at 120°C in a closed vessel for 15 minutes. The spectra were recorded after one day aging at room temperature; reducing this waiting period did not cause significant changes in the ESI results.

PRPP is a highly activated molecule, that in water solution gives rise to degradation products well-visible in the spectra: ribose-diphosphate, 1-2 cyclic PRPP, ribose-monophosphate, cyclic ribose- monophosphate, and free phosphates. The respective amounts of these species changed with time, making it difficult to record a generally-valid PRPP reference. First, we consider the reaction of PRPP with the pyrimidine uracil. In order to determine if any UMP had been formed, we recorded the ESI spectrum of a genuine UMP solution (Fig. 3A). The spectrum is dominated by the molecular peak $(\text{UMP–1H})^-$ at 322.9 atomic mass unit (amu). Fragmentation (MSMS, Fig. 3B) of this species yielded cyclic ribose-phosphate, and then phosphate ions, confirming the assignment (see also supplementary material, Fig. S5A-D).

Zooming in on the 323 amu region of the PRPP + U solution showed that a peak was indeed present, although with a low intensity (about 1% of the maximum intensity peak, Fig. 3D). Since the intensity of this signal was only about three times the noise level (see Fig. S7), we carried out second fragmentation of the species at 323 amu. This yielded cyclic ribose-phosphate and a small amount of phosphate ions, as in the genuine UMP reference (Fig. 3B,E), thus confirming the presence of UMP in the analyzed solution.

Clearer conclusions could be drawn for the PRPP + A solution. Here, the diagnostic peak of genuine AMP is molecular $(\text{AMP–1H})^-$ at 345.9 amu (Fig. 4A). Indeed, a peak at this value did appear in the PRPP + A solution (Fig. 4C,D) – in this case representing about 5% of the maximum intensity peak – and its assignment was confirmed by fragmentation of the peak at 345.9 amu, Fig. 4B,E and supplementary material, Fig. S6A-C.

The relative intensity is given only as an order of magnitude, since ESI is difficult to quantify. However, it does seem that AMP is formed from PRPP + A in higher yields than UMP from PRPP + U, all other things being equal. This is consistent with the result of our simulations: the free energy difference between the products of the reaction (including the nucleotide) and the initial reagents is negative in both cases, but higher in absolute value for AMP formation, as compared to UMP formation (Fig. 2). Thus, the formation of AMP was expected to be more favored.

We applied another mass spectrometry technique, Matrix Assisted Laser Desorption/Ionization

(MALDI), to check the conclusions reached by ESI. In fact, the results were not conclusive for experiments with uracil, perhaps due to sensitivity limitations. For instance, the negative-mode reference spectrum of genuine UMP has a peak at 323 amu corresponding to the singly deprotonated $(UMP–1H)^-$ anion (supplementary material, Fig. S8), but the spectrum of a heated PRPP + U solution did not show this signal (although, puzzlingly, there was a peak at 324 amu, reported on supplementary material, Fig. S9). Experiments with adenine were more conlusive. The genuine AMP reference (supplementary material, Fig. S10) in the positive mode shows two peaks at 348 and 370 amu that can be assigned to $AMPH^+$ and $AMPNa^+$. For the PRPP + A solution, these two peaks are clearly visible in the blown-up region shown in supplementary material, Fig. S11. In addition, and quite unexpectedly, there are also significant signals that could be attributed to species bearing more phosphoryl groups: $ADPH^+$ (or one of its isomers) at 428 amu, with $ADPNa^+$ at 450 amu and $(ADPNa_2–1H)^+$ at 472 amu, and $ATPH^+$ at 508 amu, with $ATPNa^+$ at 530 amu (see supplementary material, Fig. S12-S13). These data were confirmed by negative-mode MALDI where $(AMP–1H)^-$ was detected at 346 amu, as well as $(ADP–1H)^-$ at 426 amu and $(ATP–1H)^-$ at 506 amu as reported in supporting information (see supplementary material, Fig. S14-S15).

Following glycosylation, nucleobases could in principle bind to ribose through other sites than N9 for purines and N1 for pyrimidines. We note, however, that in deprotonated uracil the most stable resonance structure features the negative charge on N1 (*30*), indicating that the glycosidic bond is more likely to be formed on that site. In the case of adenine, a specific regioselectivity study should be performed, to distinguish the different possible sites for the formation of the glycosidic bond.

In summary, mass spectroscopic data show the formation of purine and pyrimidine nucleotides in PRPP + nucleobase hydrothermal solutions. In the case of PRPP + U, although MALDI, likely because of sensitivity reasons, does not detect the UMP nucleotide, ESI strongly indicates its formation. We explicitly verified using $^{31}$P NMR that the ESI-detected UMP (as well as AMP) was not present as a contaminant in the PRPP reagent employed for the synthesis reactions (see supplementary material, Fig. S16). In the case of PRPP + A, both ESI and MALDI clearly detect the AMP nucleotide, probably with larger yields than UMP, consistently with free-energy calculations. Furthermore, *ab initio* simulations predict that the largest barrier for nucleotide synthesis (Fig. 2) is of about 22 $k_BT$ ($T = 400$ K), pointing to a relatively fast kinetics that is compatible with the time scale of nucleotide synthesis in our experiments. The relatively low yields seem to be due mainly to the competition between nucleotide formation

and PRPP degradation reactions, by hydrolysis and cyclic phosphate formation. Further work is required to better understand this competition and determine what settings favor most nucleotide formation. For instance, in plausible prebiotic conditions, mineral surfaces are known to play a crucial role in favoring both the synthesis and stability of PRPP (*20*), while wet-dry (hydration-dehydration) cycles could favor nucleotide accumulation and further polymerization (*31*).

# Conclusions

In conclusion, we present hereby two main discoveries. First, we identify a new synthetic route spontaneously leading to *both* purine and pyrimidine ribonucleotides, starting from precursors of *both* prebiotic and biological relevance. This route is compatible with several hydrothermal environments, from submarine vents to surface lagoons and lakes, most likely present in the early Earth as well as on other planets. In this context, mineral surfaces could have played an important role in the synthesis and stabilization of PRPP, as well as in improving nucleotide condensation yields by reducing water activity (*20*).

Second, we demonstrate, with atomic-level precision, that this spontaneous prebiotic synthetic pathway is very similar to the biological one, to the point of sharing the same precursors, the same steps, and the same chirality-inverting character (Fig. 1). This correspondence is highly suggestive of an evolution process towards complex life forms conserving the basic features of prebiotic ribonucleotide synthesis. Building upon these findings, further studies will be able to assess under which conditions the reactions can proceed up the ladder of complexity, starting from the oligomerization of nucleotides, until the possible inception of a fully-functional RNA world.

# Materials and Methods

## Born-Oppenheimer Molecular Dynamics (BOMD)

*Ab initio* calculations were based on the Density Functional Theory with the exchange-correlation Perdew-Burke-Ernzerhof functional (*32*), including Grimme's dispersion corrections (*33*) and the Martins-Troullier pseudopotentials (*34*) for C, N, O, Na, P and H atoms, as implemented in the code CPMD 4.1 (*35*). BOMD simulations were performed in the canonical ensemble (NVT) at a temperature of 400 K, with a time step of $\sim$0.48 fs, in a periodically repeated cubic box of approximately 4.296 and 4.292 $nm^3$, and a density of 0.9888 and 0.9984 $g/cm^3$ for

UMP and AMP respectively. Temperature was controlled by the Nosé-Hoover thermostat with a chain length of four and with a frequency of 3000 $cm^{-1}$. The electronic convergence criterion has been set to $10^{-5}$ a.u. based on benchmark NVE simulations. The simulation boxes comprise a PRPP molecule with charge 5- and a uracil/adenine molecule (reactants) or a nucleoside monophosphate molecule (UMP or AMP) and a pyrophosphate ion in its $HP_2O_7^{3-}$ form (products), with additional sodium counterions and 108 explicit water molecules, for a total of 371 atoms in the pyrimidine setup and 374 for the purine.

## Enhanced Sampling

BOMD simulations were coupled with enhanced sampling techniques. In the first place we performed metadynamics simulations (*24*), employing as set of reaction coordinates the path collective variables (pathCV) $s$ and $z$ by Branduardi et al (*36*). In general, $s$ indicates the progress along a putative pathway composed by a discrete sequence of atomic configurations, while $z$ represents the distance from the putative pathway (complete definitions are provided in supplementary material). In this work, we employed only two reference configurations, corresponding to reactants and products, in this way avoiding to bias the observed reaction mechanisms with hypotheses on the path. In addition, we used a topological metric particularly suited to chemical reactions in solution (*27*), accurately tracking changes in the chemical bond network passing from reactants, through intermediates, until the products. This metric comparing structure $R(t)$ with reference structure $R_k$ is defined as $D(R, R_k) = \sum_{IS} \left[C_{IS} - C_{IS}^k\right]^2$, where coordination numbers of a specific atom $I$ with the atoms of species $S$ appear.

In the second stage, we carried out a committor analysisto identify the transition state of each reaction and to get more insights about its mechanism. We extracted a set of about 10 atomic configurations from reactive metadynamics trajectories, and we performed up to 20 independent unbiased BOMD trajectories of 0.5 ps starting from each configuration, verifying whether they fell into reactants or products basins. A transition state configuration is identified as committed to both basins with a probability of $50\pm10\%$.

In the final stage, a series of umbrella sampling (*25*) simulations were performed, systematically restraining the $s$ coordinate at different locations along the reaction pathway obtained by committor analysis using an harmonic potential (all details are reported in supplementary material, Table S5). The trajectory length of each window was 10 ps, for a total simulation time of 250 and 220 ps for the pyrimidine and purine cases, respectively. The weighted histogram analysis method (*37*) finally yielded the free energy profile of the reactions. All enhanced sampling

simulations employed a modified version of the plugin Plumed 1.3 (*38*).

## Sample preparation

All data presented in this work were recorded with controlled and pure PRPP purchased from Sigma Aldrich. We further verified the lack of any nucleotide contaminant by $^{31}$P NMR (see supplementary material, Fig. S16). 0.1 mol/L solutions of PRPP, UMP, and AMP were prepared in pure water as reference. Mixtures of PRPP and bases (uracil or adenine) were prepared by dissolving 12.5 mg of PRPP (0.05 mol/L) with 11.5 mg of uracil (0.2 mol/L) or 13.5 mg of adenine (0.2 mol/L) in 500 $\mu$L of pure water. When needed, the pH was adjusted to 8.5 by controlled addition of NaOH. Before mass spectrometry analysis, solutions are filtered with 2$\mu$m paper. The reactions are carried on in a glass vial with milliQ water. Then, the reactants are thermally activated at 120°C during 15 minutes.

## ElectroSpray Ionization - Ion Trap - Mass Spectrometry (ESI-IT-MS)

Samples were diluted to the micromolar range in a solution of 50% acetonitrile, 49.8% water and analyzed by direct infusion in a HCT Ultra PTM Discovery mass spectrometer from Bruker Daltonics (Germany) equipped with an electrospray ion source. The nebulizer gas pressure was 11 psi and the spray voltage was $\sim$3.5 kV. The drying gas flow was 5 L/min and the temperature was 300°C. Spectra were acquired in negative ion MS mode over a 50–2000 m/z range until the ion charge control target had reached 90000. The MS acquisitions were carried out at enhanced resolution (8100 m/z units per second). External calibration was performed with the ESI-T Tuning Mix (Agilent Technologies). The instrument was controlled using EsquireControl 6.2 software (Bruker Daltonics) and mass spectra were processed using DataAnalysis 4.0 SP2 software (Bruker Daltonics).

## MALDI-TOF/TOF

Mass spectra were generated using a 4700 Proteomic Analyzer MALDI-TOF/TOF (Applied Biosystems) instrument fitted with a Nd:YAG laser ($\lambda$ = 355 nm; pulse duration = 4 ns; repetition rate = 200 Hz). A dilute solution was vortexed for 1 min at high speed prior to deposition on the MALDI plates. All MALDI-TOF spectra, resulting from the average of a few tens to a hundred laser shoots, were obtained in positive and negative ions using the reflector mode in the m/z range of (10–4000).

# Supplementary Materials

Supplementary Text: 1.1 Enhanced sampling
Table S1. Coordination matrix for PRPP + U solution
Table S2. Coordination matrix for UMP + PPi
Table S3. Coordination matrix for PRPP + A solution
Table S4. Coordination matrix for AMP + PPi
Table S5. Umbrella sampling parameters
Supplementary Text: 1.2 Assignment of MS peaks
Fig S1. Free-energy as a function of pathCV (*s,z*) for UMP formation
Fig S2. Free-energy as a function of pathCV (*s,z*) for AMP formation
Fig S3. Molecular representation UMP synthesis
Fig S4. Molecular representation AMP synthesis
Fig S5. ESI negative mode UMP reference
Fig S6. ESI negative mode AMP reference
Fig S7. ESI negative mode for PRPP + U and PRPP
Fig S8. MALDI positive mode UMP reference
Fig S9. MALDI positive mode PRPP + U
Fig S10. MALDI positive mode AMP reference
Fig S11. MALDI positive mode PRPP + A, formation of AAP
Fig S12. MALDI positive mode PRPP + A, formation of ADP
Fig S13. MALDI positive mode PRPP + A, formation of ATP
Fig S14. MALDI negative mode PRPP + A, zoomed region 304–397 amu
Fig S15. MALDI negative mode PRPP + A, zoomed region 400–520 amu
Fig S16. $^{31}$P NMR for genuine AMP, UMP and PRPP
Other supplementary material: Movie S1. reaction PRPP + U
Other supplementary material: Movie S2. reaction PRPP + A

# References

1. W. Gilbert, *Nature* **319**, 618 (1986).

2. A. Lazcano, S. L. Miller, *Cell* **85**, 793 (1996).

3. S. A. Benner, H.-J. Kim, Z. Yang, *Cold Spring Harb. Perspect. Biol.* **4** (2012).

4. J. Oró, *Biochem. Biophys. Res. Commun.* **2**, 407 (1960).

5. R. A. Sanchez, J. P. Ferris, L. E. Orgel, *J. Mol. Biol.* **38**, 121 (1968).

6. M. P. Robertson, S. L. Miller, *Nature* **375**, 772 (1995).

7. S. Pitsch, A. Eschenmoser, B. Gedulin, S. Hui, G. Arrhenius, *Orig. Life. Evol. Biosph.* **25**, 297 (1995).

8. C. Gibard, S. Bhowmik, M. Karki, E.-K. Kim, R. Krishnamurthy, *Nat Chem* **10**, 212 (2017).

9. B. J. Cafferty, N. V. Hud, *Curr. Opin. Chem. Biol.* **22**, 146 (2014).

10. A. W. Schwartz, M. van der Veen, T. Bisseling, G. J. F. Chittenden, *Orig. Life Evol. Biospheres* **6**, 163 (1975).

11. R. Reimann, G. Zubay, *Orig. Life Evol. Biospheres* **29**, 229 (1999).

12. M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **459**, 239 (2009).

13. S. Becker, *et al.*, *Science* **352**, 833 (2016).

14. S. Stairs, *et al.*, *Nat. Commun.* **8**, 15270 (2017).

15. H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11315 (2017).

16. R. Zrenner, M. Stitt, U. Sonnewald, R. Boldt, *Annu. Rev. Plant Biol.* **57**, 805 (2006).

17. W. L. Nyhan, *Nucleotide Synthesis via Salvage Pathway. eLS.* (John Wiley & Sons, 2001).

18. K. Kawamura, M.-C. Maurel, *Orig. Life. Evol. Biosph.* **47**, 281 (2017).

19. A. L. Dennis, M. Puskas, S. Stasaitis, R. K. Sandwick, *J. Inorg. Biochem.* **81**, 73 (2000).

20. M. Akouche, M. Jaber, M.-C. Maurel, J.-F. Lambert, T. Georgelin, *Angew. Chem.* **129**, 8028 (2017).

21. P. Gopinath, V. Ramalingam, R. Breslow, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12011 (2015).

22. W. Martin, J. Baross, D. Kelley, M. J. Russell, *Nat. Rev. Micro* **6**, 805814 (2008). 10.1038/nrmicro1991.

23. A. M. Saitta, F. Saija, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13768 (2014).

24. A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).

25. G. Torrie, J. Valleau, *J. Comput. Phys.* **23**, 187 (1977).

26. F. Pietrucci, *Rev. Phys.* **2**, 32 (2017).

27. F. Pietrucci, A. M. Saitta, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15030 (2015).

28. C. Grubmeyer, M. R. Hansen, A. A. Fedorov, S. C. Almo, *Biochemistry* **51**, 4397 (2012).

29. W. Shi, *et al.*, *J. Biol. Chem.* **277**, 39981 (2002).

30. C. A. Cole, Z.-C. Wang, T. P. Snow, V. M. Bierbaum, *Phys. Chem. Chem. Phys.* **16**, 17835 (2014).

31. L. Da Silva, M.-C. Maurel, D. Deamer, *J. Mol. Evol.* **80**, 86 (2015).

32. J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).

33. S. Grimme, *J. Comput. Chem.* **27**, 1787 (2006).

34. N. Troullier, J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).

35. C. C. I. Corp., by Max Planck Institute Stuttgart, www.cpmd.org (2000-2017).

36. D. Branduardi, F. L. Gervasio, M. Parrinello, *J. Chem. Phys.* **126**, 054103 (2007).

37. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992).

38. M. Bonomi, *et al.*, *Comput. Phys. Commun.* **180**, 1961 (2009).

39. S. Ganguly, K. K. Kundu, *Can. J. Chem.* **72**, 1120 (1994).

40. A. P. M. Camargo, H. Baumgartel, C. Donner, *Phys. Chem. Chem. Phys.* **5**, 1657 (2003).

# Acknowledgements

**Funding:** This work was supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02, within the framework of the cluster of excellence MATériaux Interfaces Surfaces Environnement (MATISSE) led by Sorbonne Université. We acknowledge calculations performed on the *Ada* cluster at IDRIS, (Orsay, France) under GENCI allocations 2015-091387 and 2016-091387; and *Curie* Supercomputer at TGCC under GENCI allocations t201609s042 and A0010910143.

**Author contributions:** M.C.M., F.P. and A.M.S. designed research. A.P.V. performed computer simulations and analyzed the results. T.G. performed experiments. T.G and J.F.L analyzed and interpreted experimental data. A.P.V., F.P. and A.M.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

**Compelling interests:** The authors declare no competing financial interests.

**Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

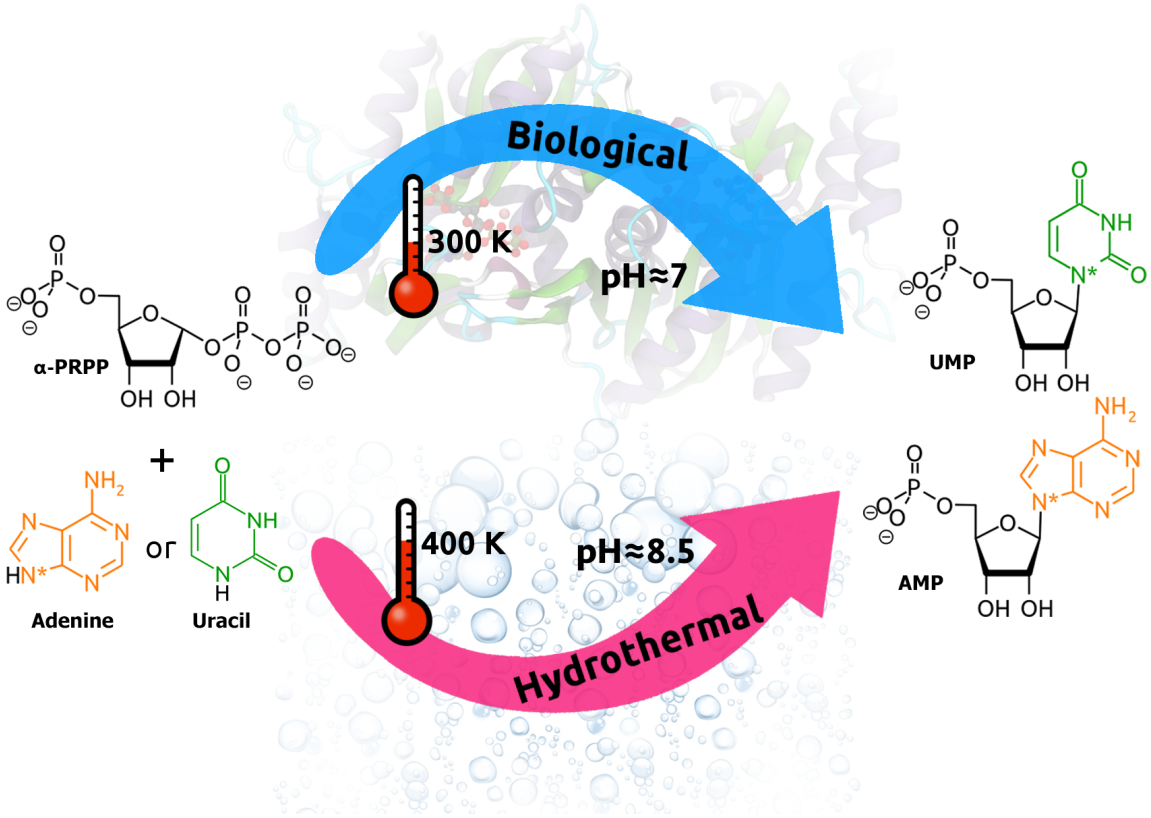


**Fig. 1.** Nucleotide formation from PRPP and purine/pyrimidine bases: current biological pathway versus putative prebiotic hydrothermal conditions. Current living forms exploit the enzymatic catalysis of phosphoribosyltransferases.

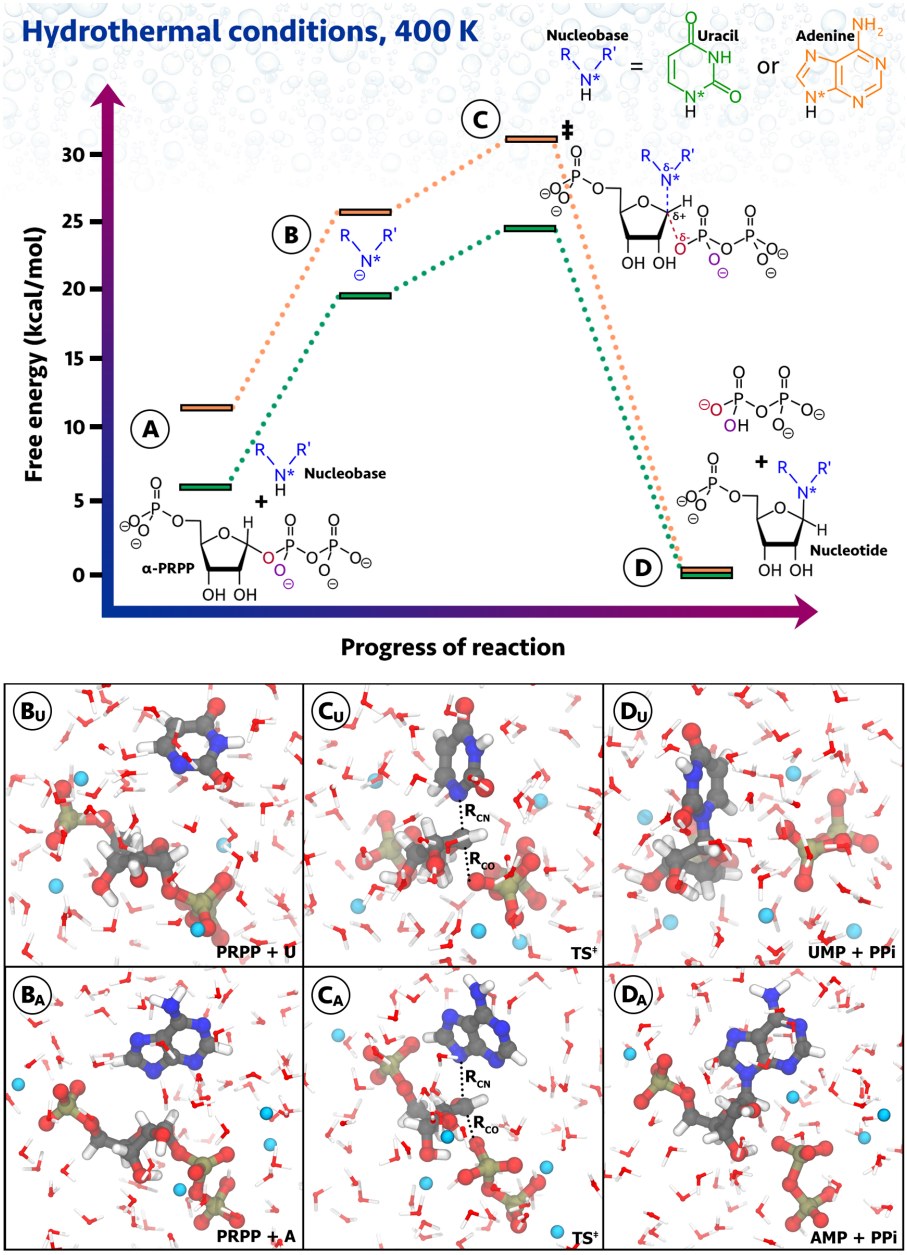


**Fig. 2.** Relative free-energy values for reactants (**A**), intermediates (**B**), transition states (**C**), and products for the nucleotide synthesis (**D**). Free-energy levels for uracil (pyrimidine) and adenine (purine) pathways are depicted in green and orange respectively. "**N***" represents the nitrogen from the nucleobase participating in the glycosidic bond. Relative free-energy values for reactants (**A**), are estimated based on the pKa values of the nucleobases (*39, 40*). Structural representation of reactants ($\mathbf{B}_\mathrm{U}$ and $\mathbf{B}_\mathrm{A}$), transition states ($\mathbf{C}_\mathrm{U}$ and $\mathbf{C}_\mathrm{A}$), and products ($\mathbf{D}_\mathrm{U}$ and $\mathbf{D}_\mathrm{A}$) during the nucleotide formation from PRPP and uracil or adenine, respectively. The transition states display the typical geometry of $S_N2$ mechanisms.

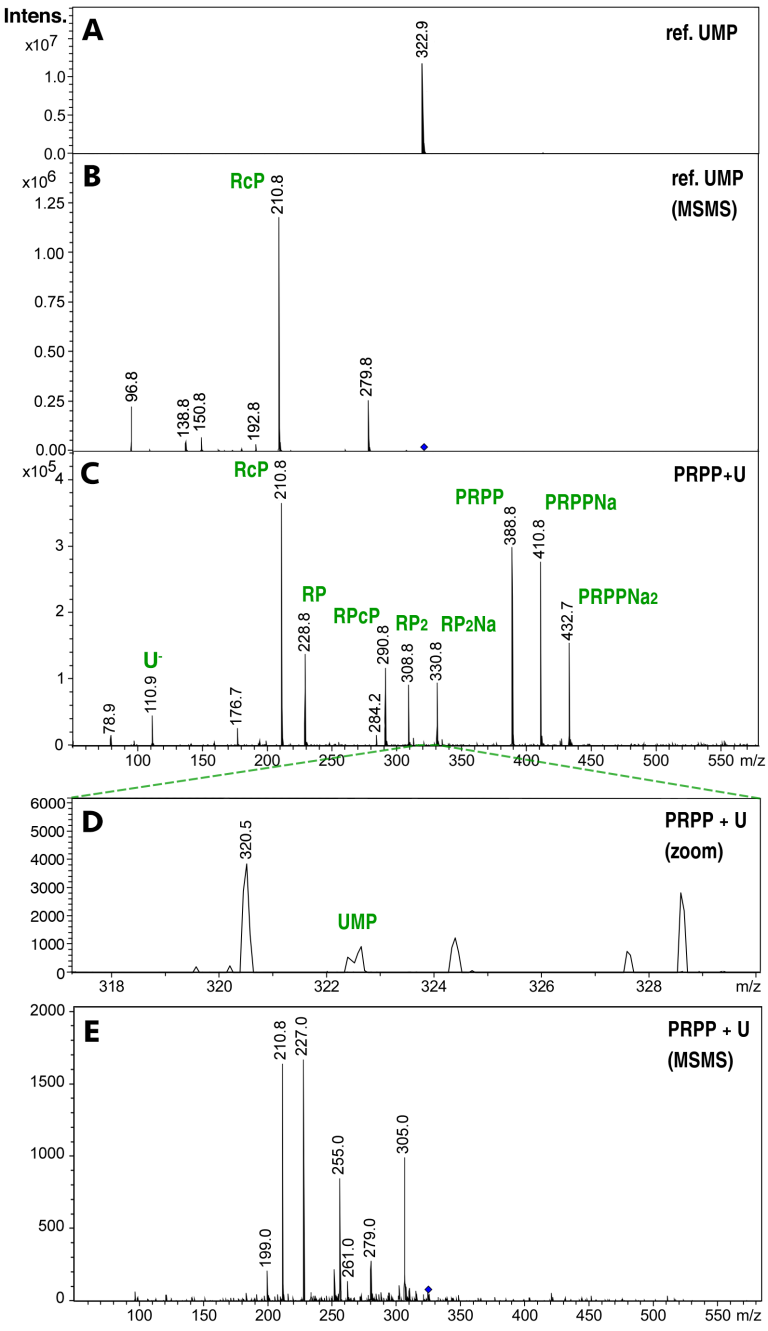


**Fig. 3.** Mass spectrometry (ESI-IT-MS – negative mode) of (**A**) UMP reference, with the peak at 322.9 amu corresponding to $(UMP–1H)^-$, (**B**) fragmentation of peak 322.9 amu in UMP reference, (**C**) complete spectrum for the sample PRPP + uracil, (**D**) zoomed spectra around UMP region (317-330 amu), and (**E**) fragmentation of peak at 322.9 amu from sample PRPP + uracil. The detailed labeling of peaks is explained in Supporting Information.

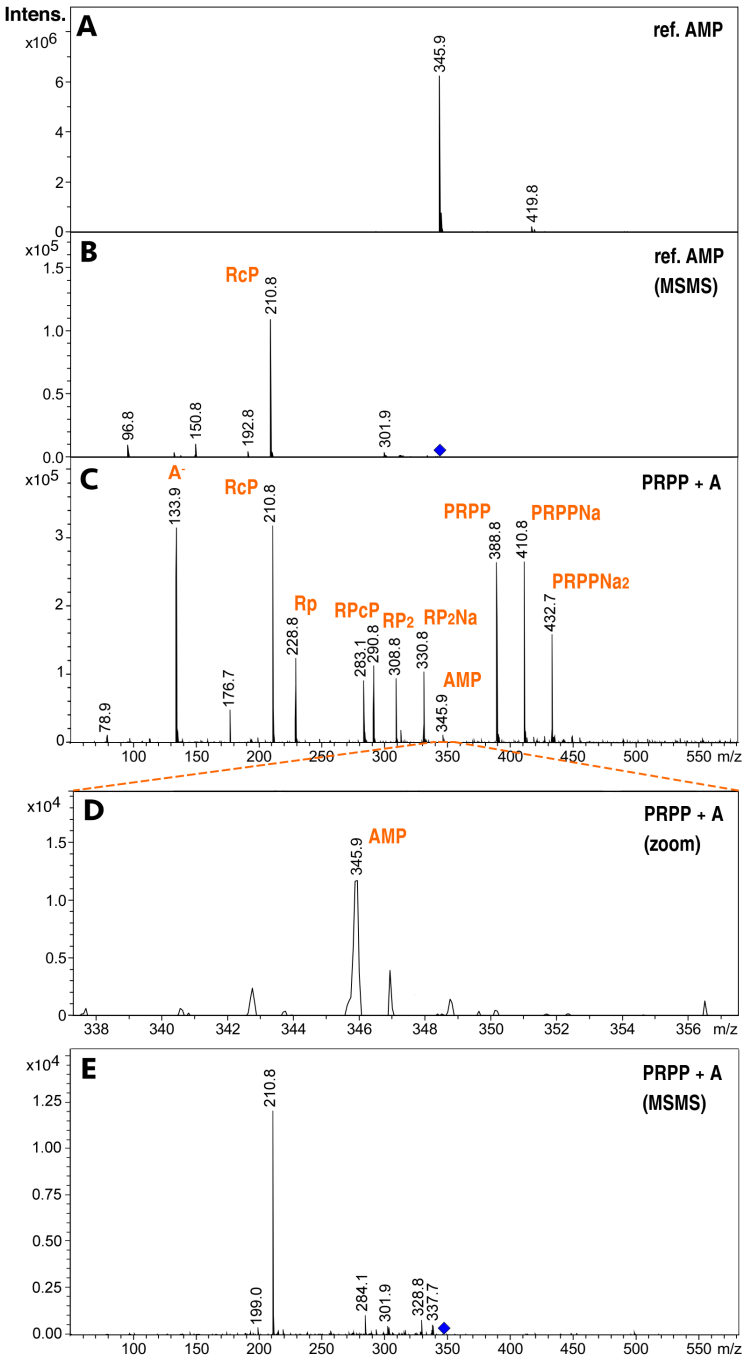


**Fig. 4.** Mass spectrometry (ESI-IT-MS – negative mode) of (**A**) AMP reference, with the peak at 345.9 amu corresponding to $(AMP–1H)^-$, (**B**) fragmentation of peak 345.9 amu in AMP reference, (**C**) complete spectrum for the sample PRPP + adenine, (**D**) zoomed spectra around AMP region (337-358 amu), and (**E**) fragmentation of peak at 345.9 amu from sample PRPP + adenine. The detailed labeling of peaks is explained in Supporting Information.