Integrated Synthesis of Nucleotide and Nucleosides Directed by Amino Acids

Irene Suárez-Marina, Rebecca Turk-MacLeod, Yousef M. Abul-Haija, Piotr S. Gromski, Geoffrey J. T. Cooper and Leroy Cronin*

School of Chemistry, University of Glasgow, Glasgow G12 8QQ (UK)

Keywords: *amino acids, nucleotides, un-activated, one-pot synthesis, origin of life, molecular evolution*

Research on the origin of nucleic acids and proteins has been approached by either multi-step synthesis or simple one-pot reactions, but exploration of their prebiotic chemistry is normally done separately. However, if nucleotides and amino acids co-existed on early Earth, their mutual interactions and reactivity should be considered in exploring the emergence of complex chemical systems that can ultimately evolve. To explore this idea, we set out to investigate nucleotide/nucleoside formation by a simple dehydration reaction of the constituent building blocks (sugar, phosphate, and nucleobase) in the presence of amino acids (*i.e.* glycine, arginine, glutamic acid, threonine, methionine, phenylalanine and tryptophan). Herein, we report the first example of simultaneous formation of glycosidic bonds between ribose, purines, and pyrimidines under mild conditions without a catalyst or activated reagents, as well as nucleobase exchange. We observed not only the simultaneous formation of nucleotide and nucleoside isomers from several nucleobases, but also the selection of distribution of glycosylation products when glycine was present. This work shows how reaction networks of nucleotides and amino acids should be considered when exploring the emergence of catalytic networks in the context of molecular evolution.

Nucleotides and amino acids are vital building blocks in biology,[1] but although nature has mastered their synthesis and polymerisation,[2,3] their synthesis under mild, prebiotically plausible and simple

conditions without activation in the laboratory is challenging.[4] In this context, the uncatalysed synthesis of nucleosides and nucleotides from their precursors has been widely investigated. For instance Orgel and co-workers synthesised small amounts of adenosine and guanosine nucleosides in separate reactions *via* the dehydration reaction of ribose together with the corresponding purine base, adenine or guanine, in the presence of inorganic polyphosphate salts.[5] However, the synthesis of canonical pyrimidine cytosine, thymine, and uracil nucleotides is difficult, and they have not been synthesised directly from the base and ribose. Currently, the state of the art synthesis requires the stepwise addition of the components reacting under unique conditions, as well as a final photoanomerisation step.[6,7,8] Non-canonical nucleoside/nucleotide formation from nucleobase analogues and (5′-phosphorylated) ribose has been successfully achieved.[9,10] This is interesting as the synthesis presents a possibility for alternative pathways to genetic polymers during a "pre-RNA world", but it would be interesting to consider how amino acids and peptides might feature in such a world. This is because the formation of peptides has recently been found to be possible under mild conditions *via* simple hydration-dehydration cycles by heating solutions of amino acids to >90˚ C.[11,12] Recently, a mixture of amino acids and ribonucleotides in the presence of an activating agent (*i.e.* carbodiimide, ethylimidazole, or magnesium chloride) has been shown to lead to the formation of mixed polymers of nucleotides and amino acids[21,22] in addition to the formation of oligo-dipeptide backbones using thioester derivatives as mediators.[23]

Here, we have studied the co-reactivity of amino acids and nucleotide building blocks under simple dehydration conditions (90˚ C for 5 hours) (Fig. S1-S2). Through simple one-pot dehydration reaction of an aqueous mixture (pH 2.5) containing nucleotide building blocks (ribose, phosphate and nucleobase) without additional activated or catalytic agents, simultaneous formation of nucleotide and nucleoside isomer structures from both purines and pyrimidines were obtained. We also observed an exchange of nucleobases within and between nucleoside and nucleotide compounds, indicating a dynamic environment of early-forming nucleic acid monomers. Furthermore, we observed a clear isomeric selection on glycosylation products when we incorporated amino acids into our reactions. This indicates that amino acids may have had the capacity to direct the chemistry of prebiotic

nucleoside/nucleotide synthesis, further supporting the hypothesis of nucleic acid/amino acid coevolution. Our combined results suggest complex, possibly indirect pathways to a stable reservoir of nucleic acid monomers, which were almost certainly subject to dynamic isomeric exchange and interaction with neighbouring small molecules.

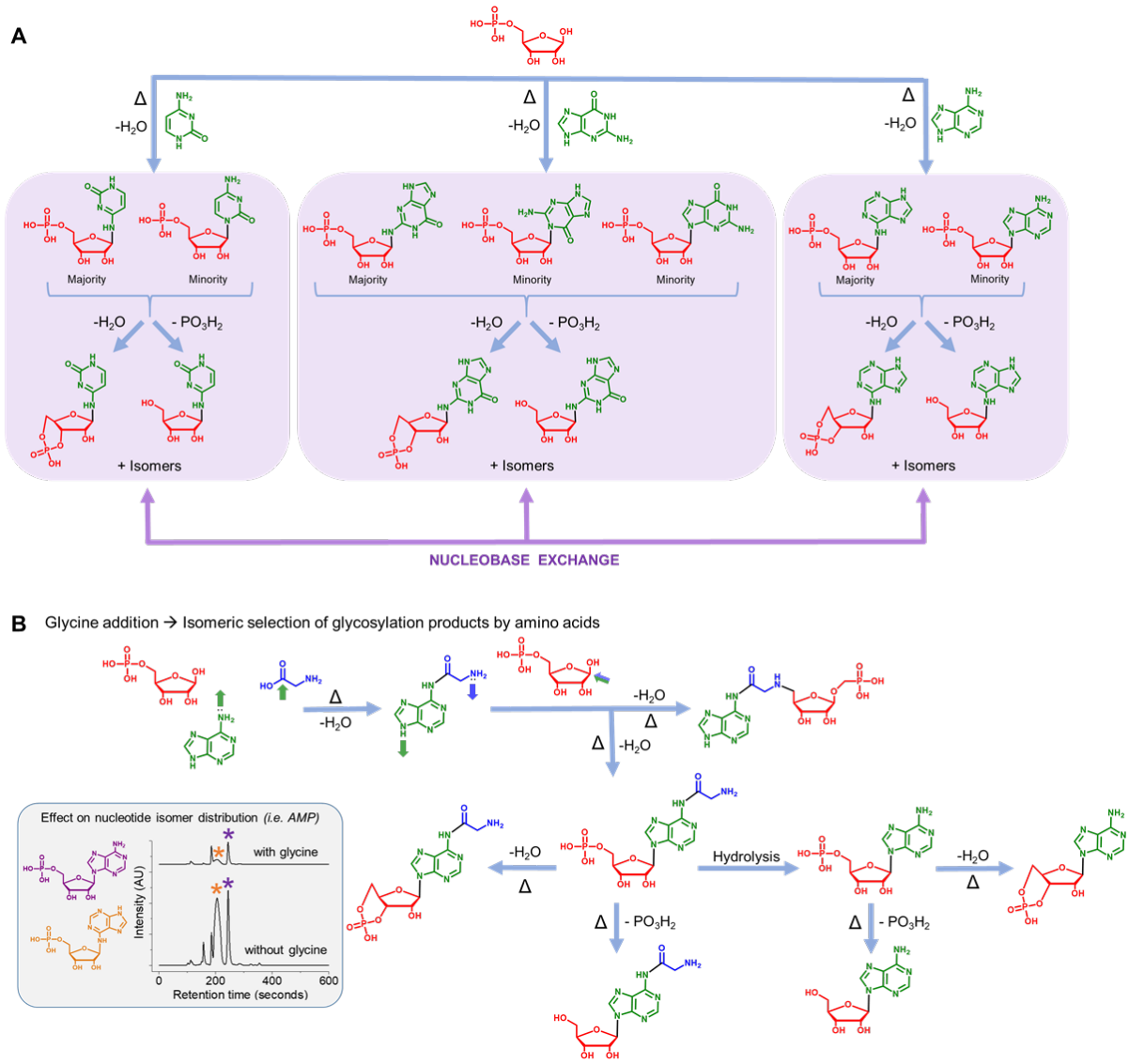

Figure 1. *Scheme showing (A) the network of reactions involving the simultaneous one-pot formation of glycosylation products (nucleotide, cyclic nucleotide and nucleosides) with three nucleobases (adenine, cytosine and guanine), and (B) the isomeric effect of glycine on the isomeric distribution of glycosylated structures. Arrows indicate preferred positions of reactivity. All reactions were carried out in equimolar conditions and heated at 90˚C for 5 hours (pH=2.5).*

The formation of nucleoside and nucleotide structures from simple precursors was achieved simultaneously with three nucleobases by dehydration reactions. This is in contrast to other work in this field, where conditions are optimised to favour specific products.[24,25,26] In a typical glycosylation

experiment, adenine and P-ribose were heated at 90˚ C for 5 hours at pH 2.5. We observed adenine monophosphate (AMP) nucleotide formation when adenine and P-ribose were combined. The extracted ion chromatograms (EICs) obtained from HPLC-MS analysis of the products revealed several peaks with the *m/z* of $[AMP+H]^+$ and $[2(AMP)+H]^+$ (Fig. S13-S15). The comparison to standards confirms that the AMP corresponds to the peak found at RT = 4.2 minutes (Fig. S5-S6); other major peaks are consistent with AMP isomers such as the N6-ribosylated isomer (Fig. S10). MS/MS analysis of our reaction products reveals fragmentation of AMP (and its isomers) to adenine (*m/z*=136.0617±0.01) (Fig. S20); this is consistent with fragmentation of the canonical AMP standard. Our proposed reaction mechanism consists of the formation of a glycosidic bond between a 1′-OH group of ribose and an amino group of adenine (see Fig. S12).

Time-course reactions evidence that water evaporation is the main driving force for glycosylation products formation, showing a significant increase in the formation of nucleoside and nucleotide structures in the time period between 2-4 hours (Fig. 2.A). This corresponds to the time range when the sample volume is drastically decreased, and reagents are extremely concentrated. After 5-6 hours of reaction, the sample reaches dryness and the rate of reaction (followed by intensity in HPLC-MS measurements) stabilises. In addition to AMP, glycosidic bond-containing products, including cyclic nucleotides (*i.e.* cAMP),[27] and nucleosides (*i.e.* adenosine), were detected using RP-HPLC-MS, MS/MS, and a test for 1,N6-etheno derivative formation,[28,29] and confirmed by comparison to standards (Fig. 2, Fig. S7-S11, Fig. S23-S27). The retention times of 2′,3′-cAMP and 3′,5′-cAMP canonical standards did not correspond with the retention times of the EIC peaks in the sample; however, the mass distributions ($[cAMP+H]^+$; $[2cAMP+H]^+$) and fragmentation patterns ($[adenine+H]^+$) were identical (Fig. S16-S17 and S21). These results show that while cyclic structures are formed, canonical cAMP is not the main product. Comparison to an adenosine analytical standard also shows that adenosine, together with a number of isomeric species, is formed in the dehydration reaction of P-ribose and adenine (Fig. S18-S19 and S22). While the canonical forms of AMP and

adenosine were confirmed in these experiments, they were not the main products in the dehydration reaction.

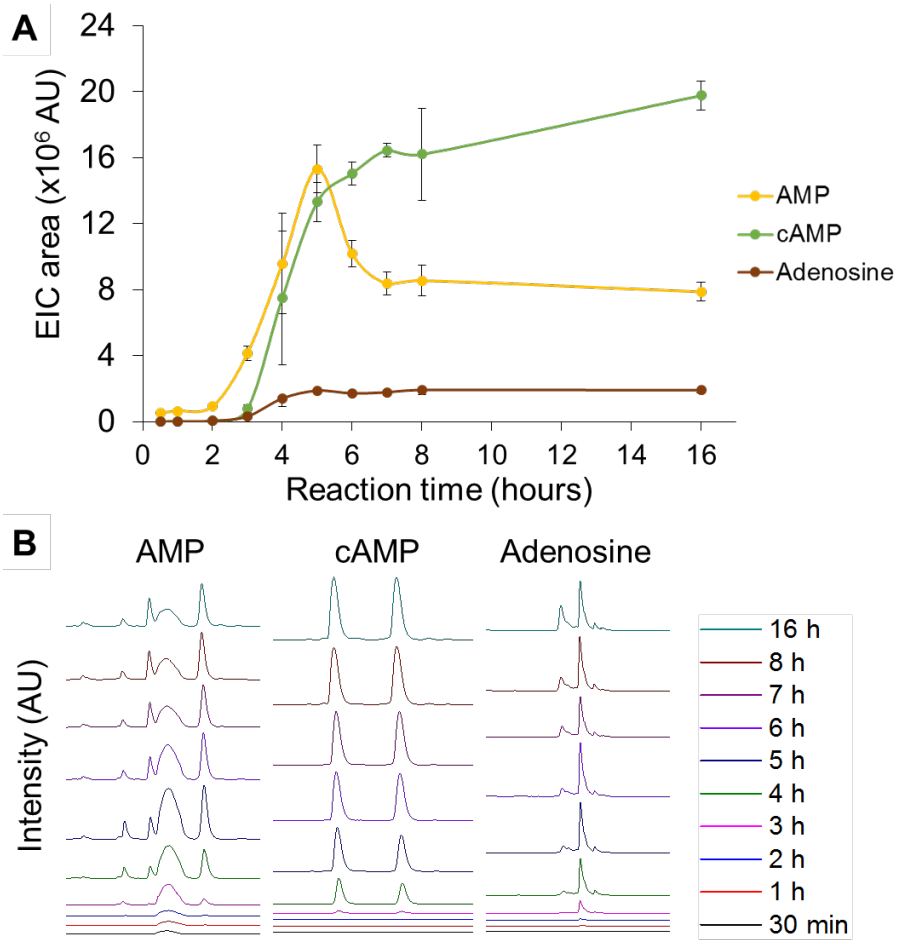

Figure 2. Adenine nucleotide products. *Time course of a reaction of P-ribose with adenine at 90˚ C; products were analysed by RP-HPLC-MS. A) Representation of the total peak areas in the Extracted Ion Chromatogram (EIC) for the masses of adenine glycosylation products. B) EICs for adenine glycosylation products over time.*

When phosphate was supplied separately as pyrophosphate and reacted with adenine and ribose, a compound with the mass of adenosine was detected and a low amount of AMP and cAMP were detected, and adenosine still formed when no phosphate source was present (Fig. S33-S41). We propose the formation of a glycosidic bond between the hydroxyl group of ribose and an amino group of adenine, which is triggered by the loss of a water molecule due to evaporation. The relative reactivity of the primary and secondary amine groups in adenine is well studied[30] and without activation or the presence of a protecting group, glycosidic linkage at the primary amine is normally preferred. Therefore, the canonical isomer of adenosine / AMP is not expected to be a major product, as it would require reaction

exclusively at the secondary amine. However, the reactivity is sufficiently high at the secondary amine site for the canonical isomers to be formed, though not as the major product. In other nucleobases, such as guanine, there are even more accessible amine groups, and the potential for isomeric products is greater. Reactivity of the ribose meanwhile is likely to be predominantly through the anomeric position, leading to fewer possible isomers, though some other minor products may be observed.

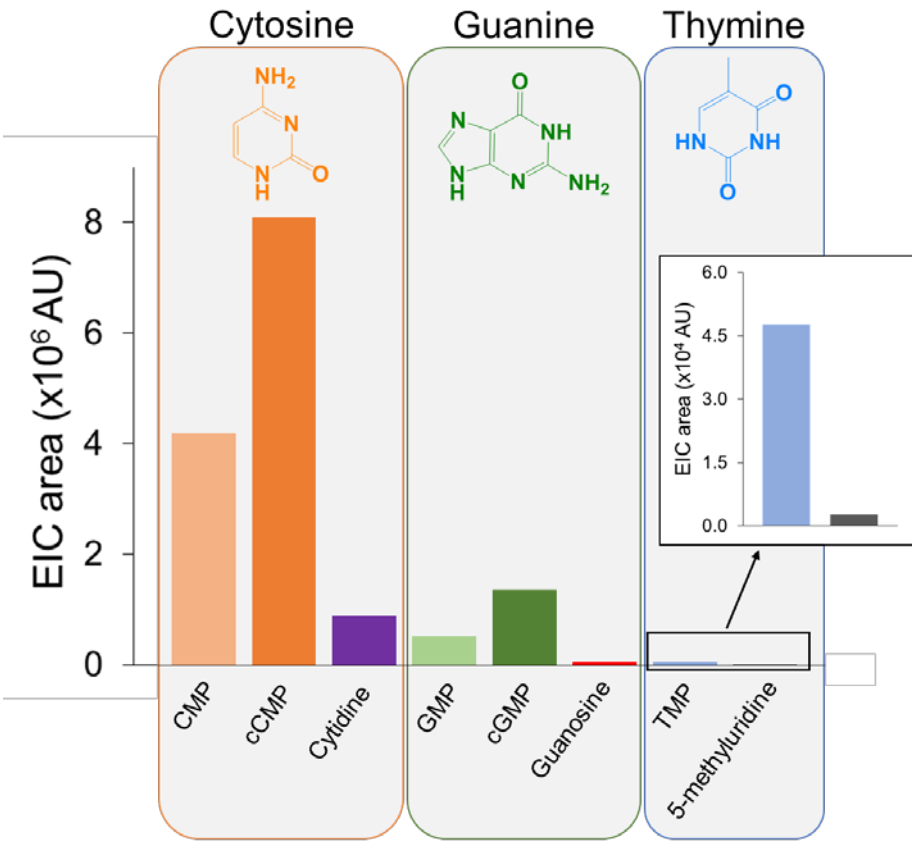

Figure 3. Other products from P-ribose and nucleobase. *All aqueous reaction mixtures consisting of 25 mM P-ribose + 25 mM nucleobase were heated at 90˚ C for 5 hours, and products were analysed by RP-HPLC-MS. Representation of the total peak areas of the EIC for the masses of cytosine, guanine and thymine glycosylation products.*

The reactivity of other canonical nucleobases (cytosine, guanine and thymine) with P-ribose was also investigated. Masses corresponding to nucleoside and nucleotide structures were detected following the dehydration reaction of guanine and cytosine with P-ribose (Fig. 3 and S50-S64). Guanine glycosylation structures were formed to a relatively low extent, likely due to the limited solubility of guanine at low pH. The product quantities measured for TMP and 5-methyluridine (thymine nucleoside) were even

lower than their equivalents from guanine and cytosine (Fig. 3 and S65-S67). These results can be explained by the presence of a primary amino group in guanine and cytosine, which is lacking in thymine. While all three nucleobases have secondary amine groups, these are less reactive in glycosidic bond formation. Hence, the formation of nucleoside and nucleotide structures through a secondary amine reaction, as is required to form canonical glycosylation products, is disfavoured in nucleobases where primary amines are available. This has an interesting implication for the adoption of nucleic acid chemistry in the origin of life, as it suggests that the canonical nucleotides may have been initially unsuitable until further biochemical machinery had emerged to enhance selectivity towards the correct isomers. Therefore, as expected, canonical nucleotide and nucleoside products of cytosine and guanine were formed in our experiments, however they did not correspond to the main peaks observed (Fig. S42-S49). By combining these two observations (different retention times but same mass distribution as canonical standards in the EICs) we can conclude that the nucleotide and nucleoside species formed from the dehydration reaction of guanine / cytosine with P-ribose were mainly isomeric species of the canonical nucleotides and nucleosides (some possible structures are shown in Fig. S50-S51).

Typically, the formation of nucleotide structures has been performed under specific conditions depending on the nucleobase used, targeting a specific reaction product. In an alternative approach, we decided to include multiple nucleobases simultaneously in the reaction with P-ribose. Our aim was to determine whether product formation with multiple nucleobases under the same reaction conditions would yield a mixture of products, or be dominated by one. This reaction was carried out by including two or three (adenine, guanine and cytosine) nucleobases simultaneously in the reaction vessel, together with P-ribose. A mixture of glycosylation products was obtained, comprising nucleotides (CMP, AMP and GMP) as well as the respective cyclic nucleotide (cCMP, cAMP and cGMP) and nucleoside products (cytidine, adenosine and guanosine) (Fig. S68-S96). Guanine glycosylation products were formed in a lower yield than those of adenine and cytosine, as expected due to the low solubility of guanine under acidic conditions. Furthermore, nucleobase exchange was observed when Na^{+}AMP was heated for 5 hours at 90˚ C in acidic aqueous media with cytosine or guanine. Nucleobase exchange resulted in the formation of nucleotide (CMP or GMP), cyclic nucleotide (cCMP or cGMP) and

nucleoside (cytidine or guanosine) structures (see Fig. S97-S103, and S146 for semi-quantitative yields). The formation of cytosine and guanine glycosylation products demonstrated that cleavage of the AMP glycosidic bond occurred under our reaction conditions. When the EICs of AMP, cAMP and adenosine were analysed, several peaks were observed in each chromatogram, supporting the theory that glycosidic bonds undergo dynamic hydrolysis / formation during the dehydration reaction. The reactions of cytosine and guanine nucleotides with adenine were also investigated. In the case of the dehydration reaction of CMP and adenine, no products corresponding to nucleobase exchange could be observed (Fig. S104-S105 and S109.A). However, adenine glycosylation products were clearly detected in the reaction of GMP with adenine (Fig. S106-S108 and 109.B). This suggests that the hydrolysis of glycosidic bonds in GMP molecules is more facile than for CMP molecules, under the same reaction conditions.

As previously mentioned, amino acids, nucleotides and their building blocks could have been present on early Earth at the same time. Therefore, products of a co-polymerisation reaction, or even products resulting from some catalytic effect of one type of polymer over the other, could have occurred under a prebiotic environment. To study the co-reactivity of nucleotide building blocks and amino acids in a one-pot dehydration, glycine, the simplest amino acid, was included in the dehydration reactions of P-ribose and respective nucleobases. The incorporation of glycine had a clear effect on the formation of glycosylation products, causing the overall yield of products with the mass of AMP, cAMP and adenosine (Fig. 4.A and S28-S32) to decrease. This indicates that glycine plays a role in either consuming nucleotide building blocks (P-ribose and/or adenine) through a side reaction, or that it becomes attached to the product structure, changing its mass. EIC analysis for these reactions reveals peaks corresponding to the mass of glycine adducts (*i.e.* AMP-Gly, cAMP-Gly, adenosine-Gly and adenine-Gly; see Fig. S114) (Fig. S115-S122), though these side products are not formed in sufficient quantity to account for all the changes observed. Glycine adducts were also confirmed by using deuterated glycine as starting material together with P-ribose and adenine, which caused changes in the isotopic distribution of the adduct masses (Fig. S123-S124). A maximum semi-quantitative yield of 59% was obtained for the formation of glycosylation products (AMP, cyclic AMP and adenosine) in

the reaction of P-ribose + Adenine; however, only 46% was obtained when glycine was also present in the reaction medium (Fig. S144).

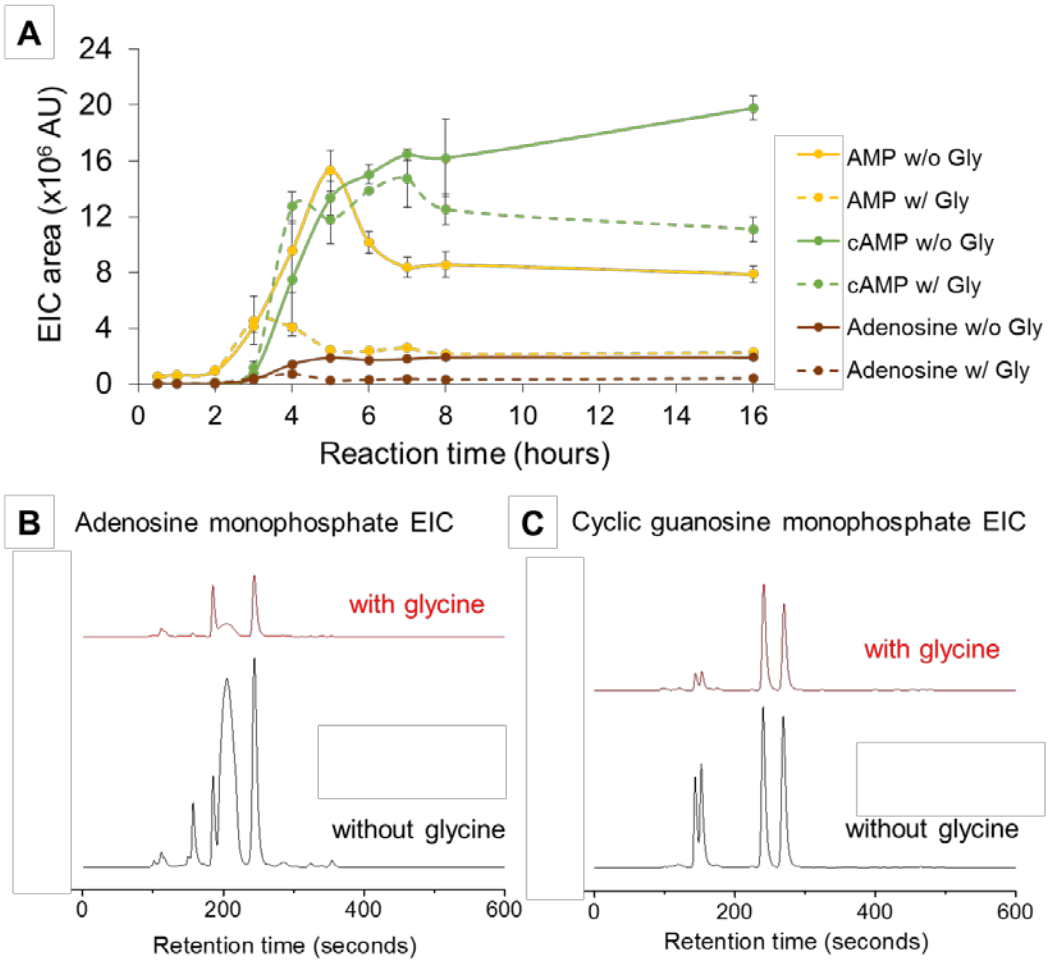

Figure 4. Glycosylation products in the presence and absence of glycine. A) *Products of the reaction of 25 mM P-ribose and 25 mM adenine are shown with a solid line; products of the reaction of 25 mM glycine, 25 mM P-ribose and 25 mM adenine are shown with a dashed line. All reactions were performed by heating the starting materials at 90˚ C for the indicated time in an acid aqueous media, after which samples were analysed by RP-HPLC-MS.* ***B)*** *EIC for adenosine monophosphate (m/z=348.0683±0.01) comparing the reaction of 25 mM adenine + 25 mM P-ribose in the presence (red) and absence (black) of 25 mM glycine.* ***C)*** *EIC for cyclic guanosine monophosphate (m/z=346.0547±0.01) comparing the reaction of 25 mM guanine + 25 mM P-ribose in the presence (red) and absence (black) of 25 mM glycine.*

Glycine also affected the distribution of isomeric species, and clear differences were observed comparing the base peak chromatogram (BPC) from the reaction of P-ribose and adenine, in the presence and absence of glycine (Fig. 4.B and S110-S113). These data indicated the presence of

different chemical species and the change in mass distribution between the reactions with and without glycine. Individual EICs were then analysed for each adenine glycosylation product, resulting in clear differences being observed in the peak relative intensities when glycine was added. These results clearly show that glycine has a selective effect on which isomeric species are preferentially formed. Glycine is known to react readily with other amines under dehydrating conditions,[12] and it is likely to react with the primary amines of nucleobases. Side hybrid products including glycine (Gly-AMP, Gly-cAMP, Gly-Adenosine, Gly-Adenine) were detected in approx. 1% (Fig. S145), however, this small percentage has an important effect on the isomeric distribution of the adenine glycosylation products (see Fig. S147-S148). Change on the isotopic distribution of the mases of hybrids products was detected (Fig. S123-S124) when deuterated glycine was included in the dehydration reaction, confirming glycine inclusion on hybrid structures.

A similar effect on isomer distribution was also observed when P-ribose was reacted with cytosine/guanine in the presence of glycine (Fig. 4.C, S125-S140). The maximum intensities of the different isomer species decreased (GMP, CMP, cGMP, cCMP, guanosine and cytidine), while the distribution of the relative intensities was also affected. The differences between experiments were highlighted using cluster analysis of EIC data to divide samples, in the presence and absence of glycine, into constituent groups/clusters with common characteristics (Fig. 5). The objective of cluster analysis in this study is to group data (*i.e.* nucleotide and nucleoside structure formation) into constituent assemblies with shared characteristics (*e.g.* glycine addition *versus* no glycine). This should demonstrate high internal homogeneity within clusters/groups and high external heterogeneity between clusters/groups. Figure 5 displays a dendrogram with "Wards" linkage.[31] To identify clusters in the dendrogram, we have coloured the spectra according to the presence of glycine (glycine - red, no glycine - black, blank - blue). As can be observed, glycine-present samples cluster together; one cluster corresponds to P-ribose + adenine + glycine (three samples), which is separated from samples without glycine. A second cluster corresponds to P-ribose + guanine + glycine and P-ribose + cytosine + glycine. Samples containing adenine separate into a larger cluster that is distinguished from the other samples, indicating a strong influence of adenine in the reaction.

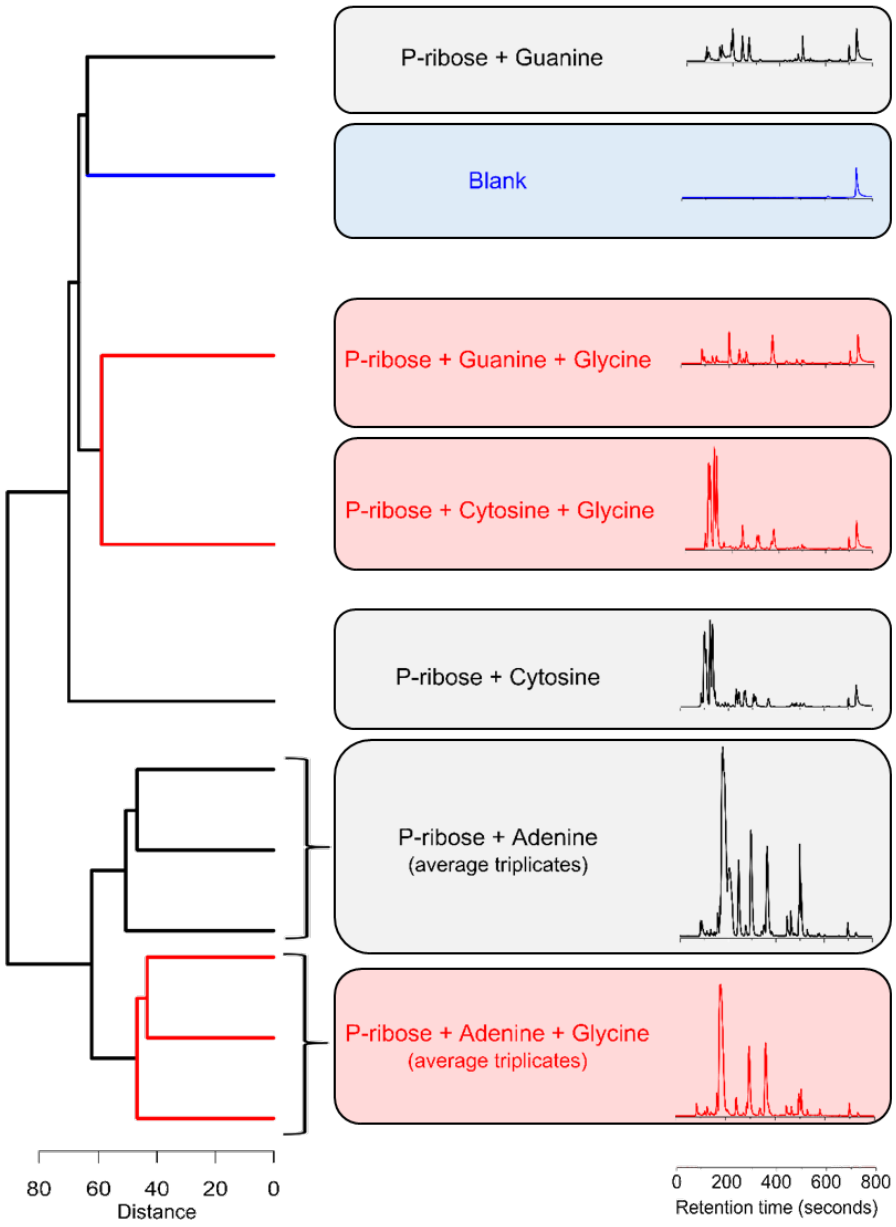

Figure 5. Dendrogram and base peak chromatograms (BPC). *Cluster analysis was used to bundle the samples into constituent groups with common characteristics. Here the dendrogram displays high internal homogeneity within clusters for three replicates each of reactions performed in the presence and absence of glycine. At the same time, the method displays high external heterogeneity between clusters where adenine samples compose a larger cluster that is more distant than other nucleotides.*

Other amino acids were also included in the dehydration reaction of adenine with P-ribose to test if they would also have some effect on the isomeric distribution of the glycosylation products (Fig. S141-S143). The six amino acids selected for this study (arginine, glutamic acid, threonine, methionine,

phenylalanine and tryptophan) have different side chains, with different chemical natures and functional groups. When the results were compared with the data obtained from the reaction of only P-ribose and adenine, changes in the isomeric distribution of AMP were observed in all the reactions, except in the case of tryptophan. Analysing cAMP EICs, smaller changes in the relative intensity of the isomeric peaks were noticed. However, a clear difference was observed in the EIC of adenosine only for the reactions including phenylalanine and threonine.

Conclusions

In conclusion, our results show that glycosidic bonds can be simultaneously formed with three different nucleobases in a one-pot dehydration reaction under acidic conditions. Nucleotide products were observed upon reacting D-ribose-5′-phosphate, adenine, guanine and cytosine as the only starting materials, without the need for any mineral, catalyst or activating agents. Glycosidic bond formation between P-ribose and the nucleobase is found to preferentially occur at the primary amine sites of the nucleobase (where available), leading to non-canonical nucleobases forming as major products. However, reactivity at secondary amine sites also occurs, leading to a distribution of isomeric products. Furthermore, while the reaction of ribose is most likely to occur through the anomeric position, glycosidic bond formation at other positions is also possible, further increasing the number of possible isomeric products. Addition of amino acids was found to significantly alter the relative intensity and distribution of product isomers, and we suggest that this is partially through its competition for reactive amine sites on the nucleobase, though amino acid adducts do not account for all the changes observed. Since the conformation of the nucleobases in a nucleic acid is critical for hydrogen bonding and base pairing, future work will investigate how this selectivity may be tuned using different amino acids or reactive species, in order to promote specific nucleobase isomers.

References

1. Ruiz-Mirazo, K., Briones, C. & De La Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.*, **114**, 285–366 (2014).

2. Rentergent, J., Driscoll, M. D. & Hay, S. Time course analysis of enzyme-catalyzed DNA polymerization. *Biochemistry*, **55**, 5622-5634 (2016).
3. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The cost of protein production. *Cell Reports*, **14**, 22-31 (2016).
4. Hud, N. V., Cafferty, B. J., Krishnamurthy, R. & Williams, D.L. The Origin of RNA and ‘‘My Grandfather’s Axe’’. *Chemistry & Biology*, **20**, 466–474 (2013).
5. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. VII Solid-state synthesis of purine nucleosides. *J. Mol. Evol.*, **1**, 249–257 (1972).
6. Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. V Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.*, **47**, 531–543 (1970).
7. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, 239–242 (2009).
8. Xu, J., Tsanakopoulou, M., Magnani, C. J., Szabla, R., Šponer, J. E., Šponer, J., Góra, R. W. & Sutherland, J. D. A prebiotically plausible synthesis of pyrimidine β-ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.*, **9**, 303–309 (2017).
9. Bean, H. D., Sheng, Y., Collins, J. P., Anet, F. A. L., Leszczynski, J. & Hud, N. V. Formation of a β-Pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. *J. Am. Chem. Soc.*, **129**, 9556-9557 (2007).
10. Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nat. Commun.*, **7**, 11328 (2016).
11. Fox, S. W. Thermal polymerization of amino-acids and production of formed microparticles on lava. *Nature*, **201**, 336–337 (1964).
12. Rodriguez-Garcia, M., Surman, A. J., Cooper, G. J. T., Suarez-Marina, I., Hosni, Z., Lee, M. P. & Cronin, L. Formation of oligopeptides in high yield under simple programmable conditions. *Nat. Commun.* **6**, 8385 (2015).
13. Leman, L., Orgel, L. & Ghadiri, M. R. Carbonyl sulfide–mediated prebiotic formation of peptides. *Science*, **306**, 283–286 (2004).

14. Forsythe, J. G., Yu, S-S., Mamajanov, I., Grover, M. A., Krishnamurthy, R., Fernandez, F. M. & Hud, N. V. Ester-mediated amide bond formation driven by wet–dry cycles: A possible path to polypeptides on the prebiotic Earth. *Angew. Chem. Int. Ed.* **54**, 9871–9875 (2015).
15. Forsythe, J. F., Petrov, A. S., Millar, W. C., Yu, S-S., Krishnamurthy, R., Grover, M. A., Hu, N. V. & Fernández, F. M. Surveying the sequence diversity of model prebiotic peptides by mass spectrometry. *PNAS*, **114**, E7652–E7659 (2017).
16. Duncan, K. L. & Ulijn, R. V. Short peptides in minimalistic biocatalysts design. *Biocatalysis*, **1**, 67-81 (2015).
17. Nakashima, T. & Fox, S. W. Synthesis of peptides from amino acids and ATP with lysine-rich proteinoid. *J. Mol. Evol.*, **15**, 161-168 (1980).
18. Gorlero, M., Wieczorek, R., Adamala, K., Giorgi, A., Schininà, M. E., Stano, P. & Luisi, P. L. Ser-His catalyses the formation of peptides and PNAs. *FEBS Letters*, **583**, 153–156 (2009).
19. Tyagi, S. & Ponnamperuma, C. Nonrandomness in prebiotic peptide synthesis. *J. Mol. Evol.*, **30**, 391–399 (1990).
20. Jauker, M., Griesser, H. & Richert, C. Spontaneous formation of RNA strands, peptidyl RNA, and cofactors. *Angew. Chem. Int. Ed.* **54**, 14564 –14569 (2015).
21. Griesser, H., Tremmel, P., Kervio, E., Pfeffer, C., Steiner, U. E. & Richert, C. Ribonucleotides and RNA promote peptide chain growth. *Angew. Chem. Int. Ed.*, **55**, 1219 –1223 (2016).
22. Griesser, H., Bechthold, M., Tremmel, P., Kervio, E. & Richert, C. Amino acid-specific, ribonucleotide-promoted peptide formation in the absence of enzymes. *Angew. Chem. Int. Ed.*, **128**, 1224 –1228 (2016).
23. Ura, Y., Beierle, J. M., Leman, L. J., Orgel, L. E. & Ghadiri, M. R. Self-assembling sequence-adaptive peptide nucleic acids. *Science*, **325**, 73–77 (2009).
24. Stairs, S., Nikmal, A., Bučar, D-K., Zheng, S-L., Szostak, J. W. & Powner, M. W. Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.*, **8**, 15270 (2017).

25. Kim, H-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *PNAS*, **114,** 11315-11320 (2017).
26. Becker, S., Thoma, I., Deutsch, A., Gehrke, T., Mayer, P., Zipse, H. & Carell, T. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science*, **352,** 833-836 (2016).
27. Cassone, G., Sponer, J., Saija, F., Di Mauro, E., Saitta, A. M. & Sponer, J. E. Stability of 2',3' and 3',5' cyclic nucleotides in formamide and in water: a theoretical insight into the factors controlling the accumulation of nucleic acid building blocks in a prebiotic pool. *Phys. Chem. Chem. Phys.*, **19**, 1817-1825 (2017).
28. Kochetkov, N. K., Shibaev, V. N., Kost, A. A. & Zelinsky, N. D. New reaction of adenine and cytosine derivatives, potentially useful for nucleic acids modification. *Tetrahedron Letters*, **22**, 1993–1996 (1971).
29. Bhatt, D. P., Chen, X., Geiger, J. D. & Rosenberger, T. A. A sensitive HPLC-based method to quantify adenine nucleotides in primary astrocyte cell cultures. *J. Chromatogr. B*, **889-890**, 110-115 (2012).
30. Brotzel, F., Ying Cheung Chu, Y-C. & Mayr, H. Nucleophilicities of primary and secondary amines in water. *J. Org. Chem.*, **72**, 3679-3688 (2007).
31. Jain, A. K., Murty, N. M. & Flynn, P. J. Data Clustering: A Review. *ACM Computing Survey*, **31**, 264-323 (1999).

Supplementary Information is linked to the online version of the paper on www.nature.com/naturechemistry.

Acknowledgements The authors gratefully acknowledge financial support from the EPSRC (Grant Nos EP/H024107/1, EP/I033459/1, EP/J00135X/1, EP/J015156/1, EP/K021966/1, EP/K023004/1, EP/K038885/1, EP/L015668/1, EP/L023652/1), the ERC (project 670467 SMART-POM), and the John Templeton Foundation Grant ID 60625, We would like to thank Dr. Maria D. Castro for her help with

HPLC-MS analysis, Dr. Marc Rodriguez-Garcia and Dr. Andrew J. Surman for initial experimental discussions.

Author Contributions. LC conceived the idea, designed the project and coordinated the efforts of the research team with IS-M, RT-M and YMA. IS-M developed and performed the synthetic and analytical protocols. IS-M and RT-M conducted data analysis, and PSG carried out statistical analysis. RT-M, YMA and GC contributed to experimental discussions and paper writing. LC, IS-M, and RT-M co-wrote the paper with input from all the authors.

Author Information. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/naturechemistry. Correspondence and requests for materials should be addressed to Leroy Cronin (Lee.Cronin@glasgow.ac.uk).