Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Die präbiotische Entstehung von kanonischen und nicht-kanonischen RNA Bausteinen auf der frühen Erde

Sidney Marcel Becker
aus
Berlin, Deutschland

2018

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 11.08.2018

(Sidney Becker)

Dissertation eingereicht am	20.08.2018
1. Gutachter:	Prof. Dr. Thomas Carell
2. Gutachter:	Prof. Dr. Oliver Trapp
Mündliche Prüfung am	24.09.2018

Meinen Eltern gewidmet.

Im Rahmen meiner Promotion war ich an der Entstehung der folgenden Publikationen beteiligt:

[1] Sidney Becker[#], Ines Thoma[#], Amrei Deutsch, Tim Gehrke, Peter Mayer, Hendrik Zipse, Thomas Carell[*], A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway, *Science*, **2016**, 352, 833.

[2] Sidney Becker, Christina Schneider, Hidenori Okamura, Antony Crisp, Tynchtyk Amatov, Milan Dejmek, Thomas Carell, Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis, *Nat. Commun.*, **2018**, 9, 163.

[3] Christina Schneider, Sidney Becker, Dr. Hidenori Okamura, Antony Crisp, Dr. Tynchtyk Amatov, Michael Stadlmeier, Prof. Dr. Thomas Carell, Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth — Generation by Prebiotic Methylations and Carbamoylations. *Angew. Chem. Int. Ed.*, **2018**, 57, 5943.

Unveröffentlichte Manuskripte:

[1] Sidney Becker[#], Hidenori Okamura[#], Stefan Wiedemann, Martin Rossa, Antony Crisp, Katharina Iwan, Tynchtyk Amatov, Christina Schneider, and Thomas Carell, Simultaneous formation of all four Watson-Crick RNA building blocks under plausible prebiotic conditions (eingereicht).

Danksagung

Professor Dr. Thomas Carell danke ich für die sehr spannende und herausfordernde Themenstellung, das entgegengebrachte Vertrauen in meine Arbeit und für die damit einhergehende Übertragung an Verantwortung. Dein Enthusiasmus bei den vielen wissenschaftlichen Diskussionen und Anregungen haben mich stets motiviert und mitgerissen.

Professor Dr. Oliver Trapp danke ich für die Übernahme des Zweitgutachtens.

Bei den Mitgliedern der Prüfungskommission, Professor Dr. Herbert Mayr, Professor. Dr. Konstantin L. Karaghiosoff, Professor Dr. Anja Hoffmann-Röder und Dr. Henry Dube, bedanke ich mich für die Mitwirkung an der mündlichen Prüfung.

Frau Slava Gärtner möchte ich für die gewissenhafte Abwicklung aller bürokratischen Aufgaben danken.

Franziska Traube danke ich für das schnelle und gewissenhafte Korrekturlesen meiner Arbeit.

Dem Labor F4.004 mit seinen aktuellen und ehemaligen Mitgliedern Katharina Iwan, Sarah Schiffers, Dr. Iacovos Michaelidis, Dr. Milan Dejmek, Dr. Olesea Kosmatchev, Dr. Simon Geiger, Dr. Nobuhiro Tago danke ich für die herzliche Aufnahme, das stetige Mut machen und einfach für eine besondere, unvergessliche Zeit mit euch!

Es war mir eine Freude mit nahezu allen Mitgliedern des Arbeitskreises Carell auf irgendeine Art und Weise kooperiert haben zu dürfen. Ich danke euch allen für die sehr angenehme Arbeitsatmosphäre, dem offenen wissenschaftlichen Austausch, die stets große Hilfsbereitschaft, den vielen Kaffeeraumfeiern und vor allem für das Organisieren der vielen Aktionen außerhalb des Labors.

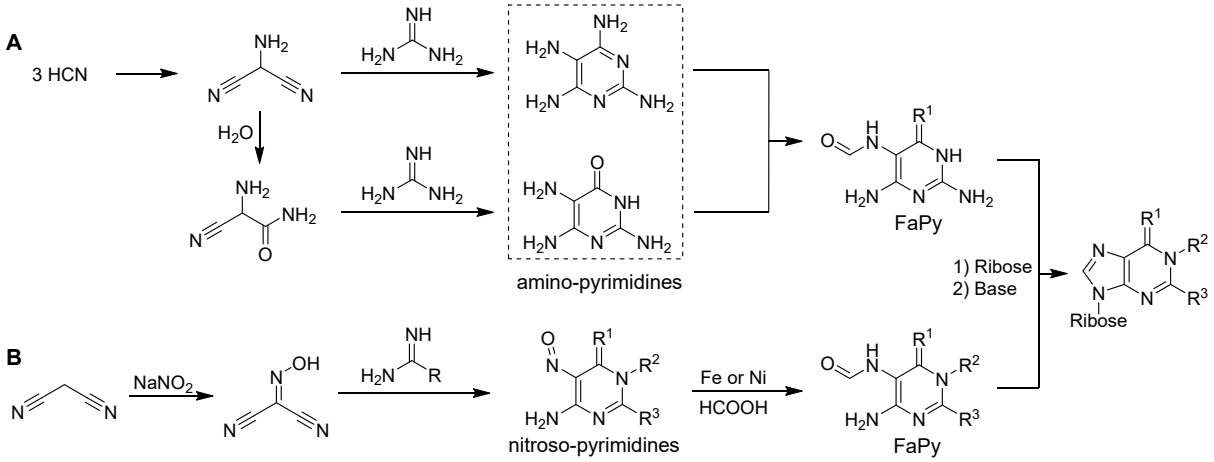
Ganz besonders bedanke ich mich bei meiner Familie für die große moralische und finanzielle Unterstützung. Im Speziellen danke ich meiner Mutter Sabine Alex, meinem Vater Frank Becker sowie meinen beiden Stiefeltern Carsten Alex und Barbara Becker. Ebenso möchte ich mich bei Sophia Slenczka für den stetigen Rückhalt in den letzten Jahren bedanken.

Summary

The origin of life is one of the greatest mysteries for mankind. Over the years, two main hypotheses have been developed for the origin of life on Earth, saying that life could have arisen from primitive metabolic cycles or from a genetic polymer. Based on the discovery of catalytic RNA molecules, the RNA world hypothesis was developed, which today found its way into common knowledge. It posits that life emerged from RNA polymers that were able to replicate themselves. Due to chemical evolution, complex chemical systems developed, which ultimately culminated in the emergence of life. The prerequisite for the RNA world hypothesis is the prebiotic synthesis of RNA on the early Earth. This requires the necessary purine and pyrimidine RNA building blocks, which are needed for the Watson-Crick base pairs A:U and G:C.

But how were these building blocks of RNA formed on early Earth? At the beginning of this doctoral thesis, a single prebiotic route to purine nucleosides has been proposed. This pathway, where the nucleobase is reacted with ribose, is considered to be inefficient due to lack of regioselectivity and low yields. The main reason for this is the low nucleophilicity of the nucleobases.

In this work, a new strategy for the prebiotic synthesis of RNA nucleosides was developed, in which not the nucleobase itself but a precursor is condensed with ribose. Suitable precursors are Formamidopyrimidines (FaPys), which can form the natural *N*-9-purine nucleosides with high regioselectivity and good yields (60%). The FaPys are synthesized by formylation of aminopyrimidines, which in turn can be formed from simple prebiotic molecules such as aminomalononitrile as a HCN trimer and guanidine (**Scheme 1**, pathway A). This provides a plausible scenario for the formation of purine nucleosides under early Earth conditions.

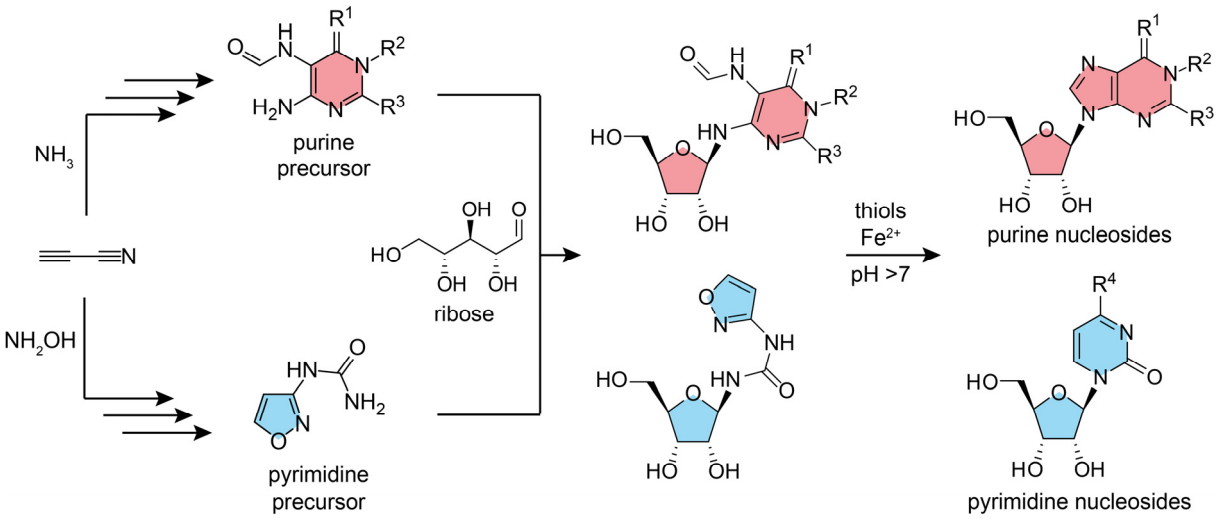


Scheme 1. Plausible prebiotic pathways leading to formamidopyrimidines (FaPys). When the FaPys are condensed with ribose, *N*-9 nucleosides are formed. Pathway A starts from HCN trimers via aminopyrimidines, while pathway B starts from malononitrile via nitroso-pyrimidines.

Unfortunately, the proposed pathway to FaPyA, as the adenosine precursor, was problematic. Therefore, a new synthetic strategy for the formation of these key building blocks was developed, starting from malononitrile instead of aminomalononitrile (**Scheme 1**, pathway B). Interestingly, a whole set of different purine precursors, which carry different substitution patterns, were formed in parallel. When reacted with ribose the canonical and some non-canonical purine nucleosides were obtained. In fact, RNA is not just build from the canonical nucleotides, but it contains more than 120 modified nucleotides. The RNA modifications obtained from the new prebiotic pathway can be found in contemporary RNA and are considered to be part of the genetic system of the last universal common ancestor (LUCA). Therefore, it can be assumed that these RNA modifications might be molecular fossils of an early Earth. Thus, this work offers a plausible solution to the question of whether RNA modifications have a chemical or biological origin. The entire synthesis pathway was designed in a way that all intermediates can be isolated or purified by fluctuations of physical parameters. These parameters include changes in pH, temperature or concentration, representing a plausible scenario for the formation of complex molecules on the early Earth. Furthermore, it was investigated whether more RNA modifications are accessible under prebiotic conditions. Therefore, a prebiotic approach for the methylation of canonical nucleosides was developed. These conditions can indeed produce a large number of RNA modifications. In addition, amino acid modified nucleosides by carbamoylation could also be detected. Our findings support the hypothesis that a large number of RNA modifications were available on the early Earth and might have taken part in the chemical evolution process.

However, one major problem in prebiotic chemistry is still unsolved. Although a plausible scenario for the synthesis of purine building blocks along the FaPy-pathway was provided, this route did not lead to the pyrimidine building blocks. For Watson-Crick base pairing, however, the availability of both purine and pyrimidine nucleosides is required. So far, no prebiotic pathway has been able to provide all RNA building blocks in a plausible geochemical environment in parallel. Therefore, a new pathway for pyrimidine nucleoside formation was developed, which is compatible with our FaPy-pathway. The new pathway allowed for the simultaneous synthesis of all RNA components under prebiotically plausible conditions (**Scheme 2**). In addition, the pathway is compatible with the formation of 5'-nucleotides in one-pot. Even 5'-diphosphates as activated RNA building blocks were detected.

In summary, our data suggests that a large variety of different RNA building blocks has probably existed on the early Earth. All of these RNA building blocks could have formed in the same geochemical environment as a prerequisite for RNA to emerge under prebiotic conditions.



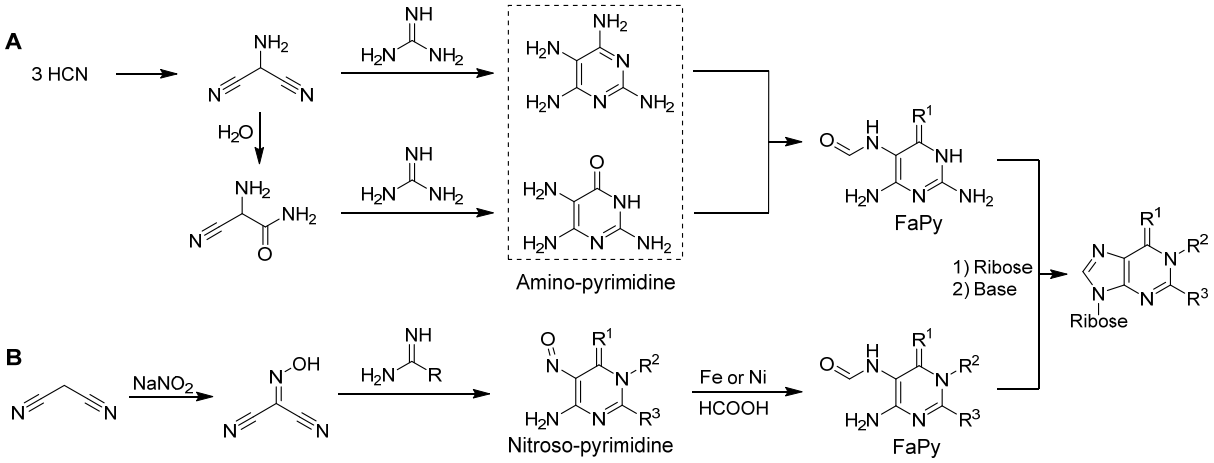
Scheme 2. Plausible prebiotic pathways towards purine (red) and pyrimidine (blue) nucleosides that were developed in the course of this thesis.

Zusammenfassung

Der Ursprung des Lebens ist eines der größten Rätsel für die Menschheit. Zurzeit gibt es zwei unterschiedliche Hypothesen zur Entstehung des Lebens auf der Erde. So könnte das Leben entweder aus primitiven Stoffwechselzyklen oder ausgehend von einem genetischen Polymer entstanden sein. Basierend auf der Entdeckung katalytisch aktiver RNA-Moleküle wurde die sogenannte RNA-Welt-Hypothese entwickelt, welche mittlerweile zum Allgemeinwissen gehört. Sie besagt, dass sich das Leben aus RNA-Polymeren entwickelt hat, welche in der Lage waren, sich selbst zu replizieren. Durch chemische Evolution entwickelten sich komplexe chemische Systeme, was letztendlich in der Entstehung des Lebens gipfelte. Voraussetzung für die Gültigkeit der RNA-Welt-Hypothese ist die präbiotische Entstehung von RNA auf der frühen Erde. Dafür bedarf es der notwendigen Purin- und Pyrimidin-RNA-Bausteinen, welche zusammen die Watson-Crick Basenpaare A:U und G:C bilden.

Aber wie konnten die Bausteine der RNA auf der frühen Erde entstehen? Zu Beginn dieser Doktorarbeit war bisher nur ein einziger präbiotischer Weg zu den Purinnukleosiden vorgeschlagen worden. Dieser Weg, bei dem die Nukleobase mit Ribose reagiert wird, gilt auf Grund fehlender Regioselektivität und geringen Ausbeuten als ineffizient. Der Grund dafür ist die geringe Nukleophilie der Nukleobasen.

In dieser Arbeit wurde eine neue Strategie zur präbiotischen Synthese von Purinnukleosiden erarbeitet, bei der nicht die Nukleobase selbst, sondern eine Vorstufe mit Ribose kondensiert wird. Als geeignete Vorstufen wurden dabei Formamidopyrimidine (FaPys) verwendet, mit welchen die natürlichen *N*-9-Purinnukleoside mit hoher Regioselektivität und guten Ausbeuten (60%) dargestellt wurden. Die FaPys selbst können durch Formylierung von Amino-pyrimidinen entstehen, welche wiederum aus einfachen präbiotischen Molekülen wie Aminomalononitril als Blausäure (HCN)-Trimer und Guanidin entstehen können (**Schema 1**, Weg A). Damit konnte ein plausibles Szenario für die Entstehung von Purinnukleosiden auf der frühzeitlichen Erde geliefert werden.

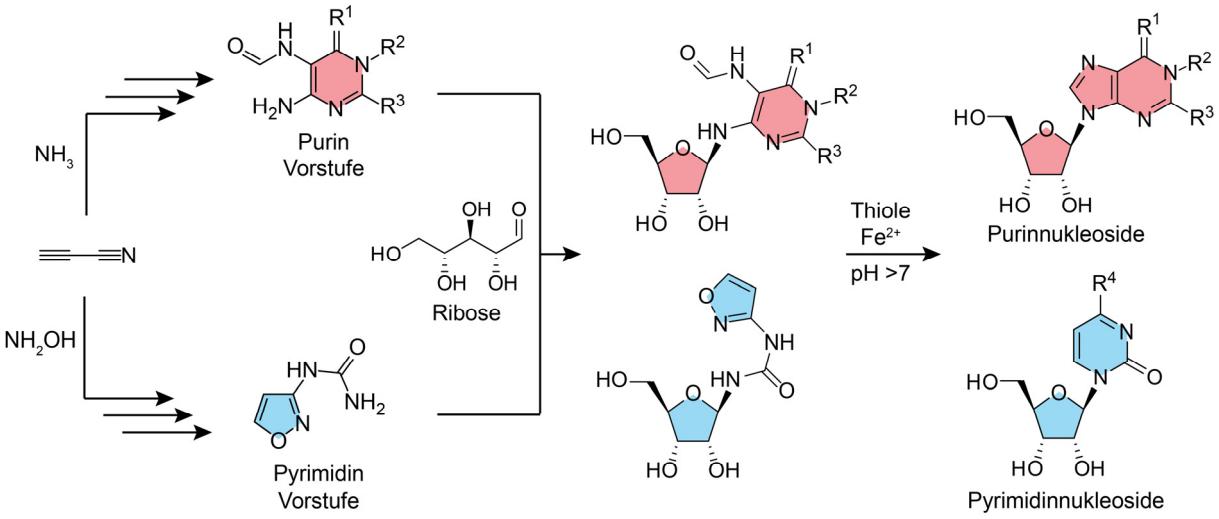


Schema 1. Plausible präbiotische Wege, die zu Formamidopyrimidinen (FaPys) führen. Wenn die FaPys mit Ribose kondensiert werden, bilden sich *N*-9 Nukleoside. Weg A beginnt mit HCN-Trimeren über Amino-pyrimidine, während Weg B mit Malonodinitril über Nitroso-pyrimidine verläuft.

Leider war der präbiotische Zugang zu FaPyA als Adenosin Vorstufe problematisch. Daher wurde ausgehend von Malonodinitril statt Aminomalononitril eine neue Strategie für die Entstehung dieses Schlüsselbausteins erarbeitet (**Schema 1**, Weg B). Interessanterweise konnte durch den neuen Syntheseweg ein ganzes Set an verschiedenen Purin-Vorstufen mit verschiedenen Substitutionsmustern parallel synthetisiert werden. Bei der Reaktion mit Ribose entstanden so die kanonischen, sowie einige nicht-kanonische Purinnukleoside. Tatsächlich besteht RNA nicht nur aus kanonischen Nukleosiden, sondern enthält darüber hinaus mehr als 120 modifizierte RNA-Nukleotide, die noch heute in allen Lebewesen zu finden sind. Die RNA-Basenmodifikationen, welche im Rahmen der neu entwickelten Syntheseroute entstehen, finden sich tatsächlich in der heutigen RNA wieder. Sie gelten sogar als Teil des genetischen Systems des letzten universellen gemeinsamen Vorfahren (LUCA). Dies impliziert, dass diese RNA-Modifikationen molekulare Fossilien einer frühen Erde sein könnten. Damit bietet diese Arbeit eine mögliche Lösung der Frage, ob RNA-Modifikationen chemischen oder biologischen Ursprungs sind. Unser Syntheseweg ist so ausgelegt, dass alle Zwischenstufen durch Änderungen von physikalischen Parametern isoliert oder aufgereinigt werden können. Diese Parameter sind pH-Wert, Temperatur oder Konzentration. Unser Reaktionsweg stellt damit ein plausibles Szenario für die Entstehung von komplexen Molekülen auf der frühen Erde dar. Zusätzlich wurde untersucht, ob weitere RNA-Modifikationen unter präbiotischen Bedingungen entstehen können. Dazu wurde ein präbiotischer Ansatz zur Methylierung kanonischer Nukleoside entwickelt. Die gefundenen Bedingungen können eine Vielzahl von RNA-Modifikationen hervorbringen. Darüber hinaus konnten auch Aminosäure-modifizierte Nukleoside durch eine Carbamoylierungsreaktion erzeugt werden. Dies unterstützt die Hypothese, dass viele RNA-Modifikationen auf der frühen Erde vorhanden waren und am chemischen Evolutionsprozess beteiligt gewesen sein könnten.

Ein grundlegendes Problem war jedoch noch immer ungelöst: Zwar wurde mit dem FaPy-Weg ein plausibles Szenario für die präbiotische Entstehung von Purinbausteinen geliefert, jedoch führte dieser Weg nicht zu den Pyrimidinbausteinen. Für die Watson-Crick Basenpaarungen ist aber die Verfügbarkeit sowohl von Purin- als auch Pyrimidinnukleosiden nötig. Bisher war kein Syntheseweg in der Lage, alle RNA-Bausteine in einer plausiblen geochemischen Umgebung darzustellen. Daher wurde in dieser Arbeit auch ein neuer Syntheseweg für die Pyrimidinnukleoside entwickelt, welcher chemisch mit unserem FaPy-Weg kompatibel ist. Damit ist es nun möglich alle RNA-Bausteine parallel unter präbiotisch plausiblen Bedingungen darzustellen (**Schema 2**). Zusätzlich konnte gezeigt werden, dass diese Bedingungen nicht nur mit der Synthese von 5‘-Nukleotiden kompatibel sind, sondern auch 5‘-Diphosphate als aktivierte RNA-Bausteine liefern können.

Zusammenfassend zeigen diese Ergebnisse, dass vermutlich eine Vielzahl an RNA-Bausteinen auf der frühen Erde existierte. All diese RNA-Bausteine könnten in derselben geochemischen Umgebung entstanden sein, um so die Voraussetzung für die präbiotische Entstehung von RNA zu schaffen.



Schema 1. Plausible präbiotische Wege zu Purin- und Pyrimidinnukleosiden, die im Rahmen dieser Arbeit entwickelt wurden.

Inhaltsverzeichnis

1 Präbiotische Chemie

1.1 Die Entstehung des Lebens: Gegebenheiten auf der frühen Erde

Leben, wie es heute auf der Erde existiert, ist extrem vielfältig. Es reicht von einfachen Spezies wie den Einzellern, über Pflanzen bis hin zu sehr intelligenten Lebensformen wie dem Menschen. Die menschliche Spezies besitzt ein so hoch entwickeltes Bewusstsein, dass sie in der Lage ist, ihre eigene Entstehung zu hinterfragen. Dieses Bewusstsein führte zur Evolutionstheorie, welche durch paläontologische Funde bestätigt werden konnte. Aus der Evolutionstheorie wiederrum lässt sich die monophyletische Abstammungstheorie herleiten. Diese besagt, dass alles Leben auf einen gemeinsamen Vorfahren zurückzuführen ist: LUCA („**L**ast **U**niversal **C**ommon **A**ncestor“, auf Deutsch „letzter allgemeiner gemeinsamer Vorfahr“). LUCA war vermutlich ein thermophiler Organismus der Tiefsee, wie durch phylogenetische Analysen angenommen wird.[1] Gleichzeitig zeigt dieser phylogenetische Ansatz, dass LUCA bereits ein sehr komplexes Lebewesen mit vielen evolutionären Innovationen war. Aus diesem Grund darf LUCA nicht mit der ersten lebenden Zelle gleichgesetzt werden.[2] Das erste Lebewesen muss nicht zwangsläufig am gleichen Ort wie LUCA entstanden sein, sondern könnte sich im Laufe der Evolution an die Tiefseebedingungen angepasst haben. Die erste lebende Zelle, aus der LUCA durch Evolution entstanden war, muss höchst wahrscheinlich eine weitaus geringere Komplexität aufgewiesen haben, um so überhaupt eine Chance für eine spontane Entstehung aus unbelebter Materie gehabt zu haben. Dieser Übergang von einem teilweise lebensfeindlichen, leblosen Planeten bis hin zu einem Planeten, welcher heute Millionen von verschiedenen Lebensformen beherbergt, ist eines der fundamentalen Puzzles der Wissenschaft. Während die Physik kontinuierlich weitere Erkenntnisse für die Entstehung des Universums findet und die Biologie immer detaillierter darlegen kann, was Leben eigentlich ist, bleibt es an der Chemie zu zeigen, wie aus einfachsten Molekülen die erste lebende Zelle entstehen konnte. Es liegt nahe, dass diese Frage immer wieder als eine der fundamentalen Aufgaben der Chemie bezeichnet wird.[3-4]

Die Disziplin, welche sich mit der Frage nach der Entstehung des Lebens beschäftigt, wird als präbiotische Chemie oder chemische Evolution bezeichnet. Sie beschreibt die Entstehung von Lebewesen aus organischen Molekülen vor über 4 Milliarden Jahren,[5] welche der zellulären Evolution vorausging. Das Zitat „The origin of life cannot be ‘discovered’, it has to be ‘re-invented’“ von Albert Eschenmoser,[6] umschreibt treffend die Grundproblematik der präbiotischen Chemie. Es existieren heute nur noch wenige Anhaltspunkte über die geologischen Gegebenheiten, die verfügbaren Moleküle, sowie deren Reaktionswege auf der jungen Erde. Die Vielzahl an Variablen macht es unmöglich nachzuvollziehen, wie genau die einzelnen Prozesse, welche zur Entstehung des Lebens nötig waren, abgelaufen sein könnten. Die präbiotische Chemie muss daher sinnvolle Annahmen treffen, um so neue Reaktionswege zu finden, welche auf der frühen Erde plausibel erscheinen. Ausgehend von einfachsten Molekülen wie z.B. Ammoniak (NH_3), Blausäure (HCN) oder Methan (CH_4), müssen diese chemischen Reaktionen zu immer komplexeren Systemen wie z.B. Lipiden oder Nukleinsäuren führen, welche

essentielle molekulare Bestandteile des Lebens sind. Erst das Zusammenspiel dieser verschiedenen Molekülklassen ermöglicht dann den Übergang von lebloser Materie zu einer lebenden Spezies.

1.1.1 Die Ursprünge der präbiotischen Chemie

Die konzeptionellen Ursprünge der präbiotischen Chemie gehen bis zum Anfang des 20ten Jahrhunderts zurück. Unabhängig voneinander stellten Aleksander Oparin (1924)[7-8] und John Haldane (1929)[9] die Hypothese auf, dass eine reduzierende Uratmosphäre durch eine angemessene Energiequelle wie Lichtblitze oder UV-Strahlung in der Lage wäre eine große Vielfalt an Molekülen zu produzieren. Die entstandenen Moleküle würden sich im Meer ansammeln und dort mit Hilfe von Sonnenenergie weiter zu komplexeren Strukturen bis hin zum ersten Lebewesen reagieren. Als Ausdruck für das Meer, mit seiner Vielzahl an Molekülen, verwendete Oparin den Begriff „premordial Soup“, zu Deutsch präbiotische Suppe, welcher noch heute als Umschreibung für die Haldane-Oparin Sichtweise verwendet wird. Obwohl die Konzepte bereits vorhanden waren, entwickelte sich die präbiotische Chemie erst mit dem wegweisenden Miller-Urey Entladungsexperiment (1953) zu einer eigenen chemischen Disziplin.[10] Doch auch schon ältere Experimente wie die Harnstoff- oder Oxalsäuresynthese von Friedrich Wöhler,[11-12] die Aminosäuresynthese von Adolph Strecker,[13-14] die Formosereaktion von Alexander Butlerow[15] oder die Entladungsexperimente von Walther Löb[16-17] sind aus heutiger Sicht als Meilensteine der präbiotischen Chemie zu betrachten. Für seine Entladungsexperimente nutzte Löb feuchtes CO_2, welches zu Zuckervorläufern wie Formaldehyde und Spuren von Glycolaldehyd reagierte.[16] In einem ähnlichen Versuch konnte Löb die Aminosäure Glycin als Baustein von Peptiden nachweisen, indem er feuchtes Formamid einer elektrischen Entladung aussetzte.[17] Im Vergleich zu Löbs Entladungsexperiment konnten von Miller und Urey neben Glycin eine Vielzahl weiterer Aminosäuren (z.B. Alanin, Glutaminsäure, Asparaginsäure) und andere Substanzen (z.B. Ameisensäure, Harnstoff, *N*-Methylharnstoff, Essigsäure) nachgewiesen werden.[10, 18-19] Miller zeigte später, dass durch die Entladung nicht direkt Aminosäuren, sondern lediglich HCN und Aldehyde entstanden, welche dann zu Amino- und Hydroxynitrilen weiterreagierten und schließlich zu den eigentlichen Säuren in der wässrigen Phase hydrolysiert wurden.[20]

1.1.2 Die Uratmosphäre

Die von Miller und Urey angenommene reduzierende Uratmosphäre setzte sich aus Wasser (H_2O), Methan (CH_4), Ammoniak (NH_3) sowie Wasserstoff (H_2) zusammen. Diese bis dahin geltende Überzeugung einer reduzierenden Atmosphäre geriet in den darauffolgenden Jahren immer weiter in die Kritik.[21] Neue wissenschaftliche Erkenntnisse in den 70er Jahren[22-25] führten schließlich zu der Annahme einer CO_2-reichen und damit neutralen Uratmosphäre, welche maximal Spuren von H_2, CH_4, NH_3 oder O_2 enthielt.[26-27] Dies führte jedoch zu einem Dilemma, da die Synthese von organischen Molekülen, unter anderem HCN, unter solchen Bedingungen deutlich weniger effizient abläuft.[28-29] HCN, als zentrales Molekül, ist nicht nur für die Entstehung von Aminosäuren und Nukleobasen von

Bedeutung, sondern wurde auch zur Synthese von Lipid- oder Zucker-Vorläufermolekülen verwendet.[30-34] Der Grund für die sehr ineffiziente Synthese von organischen Molekülen in einer neutralen Atmosphäre ist die Stickstofffixierung, da die vorhandenen Stickstoffatome als N_2 gebunden und damit chemisch praktisch inert sind. Um Stickstoff für präbiotische Synthesen nutzbar zu machen, muss dieses als „fixierte“ (reduzierte) Spezies, z.B. als NH_3, zu Verfügung stehen. In einer neutralen Atmosphäre, hauptsächlich bestehend aus CO_2 und N_2, reagiert der Stickstoff durch elektrische Entladung jedoch nicht zu NH_3 sondern zu Stickstoffmonoxid (NO).[35-37] Durch effiziente Umsetzung von NO zu NH_3 bzw. HCN unter möglichen präbiotischen Bedingungen konnte dieses anfangs problematische Resultat später gelöst werden.[38-40] Andere Annahmen gehen davon aus, dass organische Substanzen wie HCN in einer N_2-Atmosphäre als Resultat von Meteoriteneinschlägen während der Zeit des „Großen Bombardement“ entstanden sein könnten.[41-43]

1.1.3 Mögliche präbiotische Ausgangsverbindungen

Als „Großes Bombardement“ wird die Zeit in der Entwicklung des Sonnensystems vor ca. 4 Milliarden Jahren bezeichnet, in der es zu zahlreichen Einschlägen durch große Asteroiden auf der jungen Erde kam.[44-46] Einige präbiotische Theorien gehen davon aus, dass ein großer Teil der vorhanden organischen Substanzen über Asteroiden und Kometen auf die Erde gelangt sein könnte.[42, 47-48] Obwohl die Himmelskörper oftmals eine Vielzahl an einfachen organischen Molekülen beherbergen, ist es fraglich, ob die Menge, die einen Einschlag übersteht, wirklich signifikant gewesen sein kann. Unabhängig davon waren die Meteoriten in der Lage, komplexe Moleküle über unbekannte Reaktionspfade zu generieren. Unter anderem konnten auf verschiedenen Meteoriten biologisch relevante Nukleobasen oder Aminosäuren nachgewiesen werden.[49-54] Fraglich bleibt, ob diese erst durch die Energie der Einschläge oder bereits im interstellaren Raum entstanden waren.[42] Da die Meteoriten bei Ihren Einschlägen großen Temperaturen ausgesetzt sind, ist es schwer, Rückschlüsse auf die ursprünglich an der Oberfläche vorhandenen reaktiven Moleküle zu ziehen. Doch schon Anfang der 70er Jahre war es möglich, erste organische Substanzen in interstellaren Staubpartikeln über Infrarotspektroskopie nachzuweisen.[55] Im Laufe der nächsten Jahrzehnte konnten immer zahlreichere Verbindungen im interstellaren Raum in Eis oder Staubpartikeln identifiziert werden.[56-58] Die Liste an gefundenen Substanzen wurde durch aufwendige Weltraumprojekte wie z.B. der Stardust Mission auf Komet „81P/Wild 2“ oder der Rosetta Mission auf Komet „67P/Churyumov-Gerasimenko“ immer wieder erweitert. [59-61] Die Suche nach Molekülen auf Kometen oder im interstellaren Raum ist aus präbiotischer Sicht hoch interessant, da sie Anhaltspunkte darüber liefern kann, welche möglichen Substanzen auf der frühen Erde existiert haben könnten. Im Weltraum herrschen ein Hochvakuum mit sehr geringer Teilchendichte sowie eine Temperatur nahe dem absoluten Nullpunkt. Unter diesen Bedingungen kann davon ausgegangen werden, dass Moleküle nur sehr langsam reagieren, weshalb die „Zeit“ chemisch gesehen fast stillsteht. Die gefundenen Moleküle können daher in der Präbiotik als mögliche Ausgangsverbindungen für chemische Synthesen angenommen werden.

Bei den möglichen präbiotischen Ausgangsverbindungen handelt es sich überwiegend um sehr einfache, zum Teil hoch reaktive Substanzen wie Nitrile, Isocyanate oder Aldehyde bzw. Ketone. Vor allem gasförmige Moleküle reagierten vermutlich mehrheitlich in der Atmosphäre durch UV-Strahlung oder Entladungen (siehe Abschnitt 1.1.2). Die entstandenen Verbindungen konnten sich anschließend, z.B. durch Regen, im Meer oder anderen Gewässern ansammeln. Die zu ca. 70% von Ozeanen bedeckte Erdoberfläche lässt vermuten, dass die Mehrheit der präbiotischen Reaktionen in Wasser abgelaufen sein muss. Neben Wasser als präbiotisches Lösungsmittel muss auch Formamid in Betracht gezogen werden. Formamid als Hydrolyse Produkt von HCN kann sich auf Grund seines deutlich höheren Siedepunkts im Vergleich zum Wasser anreichern.[62] Durch Evaporation von Wasser kann neben Formamid auch ein komplett Lösungmittel-freier (trockener) Zustand entstehen, in welchem Feststoffe reagieren konnten.

1.1.4 Die Temperatur auf der frühen Erde

Die Frage, welche Temperaturen auf der frühen Erde geherrscht haben, unter denen die präbiotischen Reaktionen abgelaufen sind, kann nicht ganz klar beantwortet werden. Während atmosphärische oder geologische Chemiker von höheren Temperaturen ausgehen, nehmen präbiotische Chemiker oftmals kühlere Temperaturen an. Der Fakt, dass organische Substanzen bei höheren Temperaturen deutlich geringere Halbwertszeiten besitzen, wird immer wieder als Argument für kältere Temperaturen verwendet.[63-67] In der Tat sind die Halbwertszeiten von Cytosin (ca. 3 Wochen) bzw. Adenine (ca. 1 Jahr) durch Hydrolyse bei 100 °C erstaunlich gering im Vergleich zur geologischen Zeitskala.[65, 68-69] Ribose als essentieller Bestandteil von RNA ist im Vergleich zu den Nukleobasen noch labiler mit Halbwertszeiten von wenigen Minuten bei 100 °C.[70] Über Komplexierung mit Silicaten oder Boraten kann Ribose jedoch stabilisiert und damit deren Halbwertszeit deutlich erhöht werden.[71-72] Da die Verfügbarkeit von Substanzen immer vom Gleichgewicht aus Synthese und Zerfall abhängt, ist eine geringere Zerfallskinetik zum Beispiel durch Stabilisierung der Reaktionsprodukte ein entscheidender Faktor für die Anreicherung eines Moleküls unter präbiotischen Bedingungen. Das Problem der Anreicherung ist vor allem für den zentralen Baustein HCN gegeben. Juan Oro konnte 1960 zeigen, dass NH_4CN zu Adenin reagieren kann.[31] Diese Reaktion über Polymerisation von HCN läuft jedoch nur bei hohen Konzentrationen (> 1M) ab. HCN als volatiles Molekül hätte sich in Wasser demnach sehr stark anreichern müssen.[73] Wie Sanchez et al. zeigen konnte, ist eine Anreicherung von HCN über Bildung einer eutektischen Lösung durch Gefrieren prinzipiell möglich.[74] Ein solches Szenario ist im Einklang mit der Annahme, dass die junge Sonne eine 25-30% geringere Strahlungsleistung besaß.[75-77] Als Resultat wäre jegliches Wasser an der Erdoberfläche gefroren gewesen.[78] Da gefrorene NH_4CN Lösungen ebenfalls zu den biologisch relevanten Nukleobasen reagieren können, sind tiefe Temperaturen auf der frühen Erde ein mögliches Szenario für die Entstehung des Lebens.[79]
Im Gegensatz dazu sprechen geologische Funde für eine warme Erde, auf der Wasser im flüssigen Zustand trotz der geringeren Strahlungsleistung der Sonne existiert haben muss.[80-84] Dieser Umstand wird auch als Paradoxon der schwachen jungen Sonne bezeichnet.[85] Atmosphärische Chemiker gehen davon aus, dass der Treibhauseffekt dabei eine entscheidende Rolle gespielt haben muss. Durch hohe

Konzentrationen an Treibhausgasen (vor allem CO_2) konnte sich die Erde trotz der geringeren Sonneneinstrahlung erwärmen.[25, 27, 86-87] Weitere Faktoren, welche zu einer frühen warmen Erde beigetragen haben könnten, sind die Gezeiten durch den geringeren Mondabstand zur Erde sowie radioaktiver Zerfall oder Einschläge durch große Meteoriten.[88-92]Dabei hätten die Ozeane nach klimatischen Modellen eine Temperatur im Bereich von 40-85 °C erreicht,[83, 93-95] wobei modernere Modelle von Ozeantemperaturen von <40 °C ausgehen.[96-97] Die Entdeckung von hyperthermophilen Organismen bestätigte, dass Leben bei hohen Temperaturen bis deutlich >100°C grundsätzlich möglich ist.[98-100] Typische Besiedlungsräume dieser Organismen sind hydrothermale Quellen in der Tiefsee oder am Land. Auf Grund der auffälligen Parallelen zwischen der in hydrothermalen Quellen vorhandenen Redoxchemie und den grundlegenden Energie erzeugenden metabolischen Reaktionen in autotrophen Organismen gehen einige Evolutionsbiologen davon aus, dass die heißen Tiefseequellen ein möglicher Ort für die Entstehung des Lebens sein könnten.[101-102]

1.1.5 Präbiotische Chemie: allgemeine Rahmenbedingungen

Zusammenfassend kann gesagt werden, dass es schwer ist, die genauen klimatischen und geologischen Bedingungen auf der frühen Erde zu rekonstruieren. Nichtsdestotrotz gibt es einige allgemeine Übereinstimmungen über die Bedingungen einer „präbiotischen“ Reaktion:[103]

1) Sauerstoff war maximal in Spuren vorhanden.
2) Reaktionen müssen in Wasser, Formamid oder in Abwesenheit eines Lösungsmittels erfolgen.
3) Aus Plausibilitätsgründen müssen die Startmaterialien der Synthese in ausreichender Menge und am Ort der Reaktion vorhanden gewesen sein.
4) Zur Aktivierung bzw. Katalyse stehen Wärme, Licht, Mineralien, Säuren oder Basen zur Verfügung.

Diese sehr weit gefasste Definition der „präbiotischen Bedingungen“ kann und sollte auf Grund der geologischen und klimatischen Unsicherheiten nicht genauer definiert werden. Unabhängig davon, ob die Erde im Durchschnitt eher warm oder kalt war, kommen alle plausiblen Temperaturen von weit unter 0 °C bis zu mehreren 100 °C in Frage. Wie auch heute auf unserer Erde waren die physikalischen Bedingungen auf der frühen Erde vermutlich nicht einheitlich. So können die Temperaturen an Land in der Nähe von vulkanischer Aktivität vermutlich mehrere 100 °C, in der Tiefsee aber konstant bei 4°C gelegen haben. Doch auch die Tiefsee kann z.B. in Bereichen von hydrothermalen Quellen Temperaturen von deutlich über 4°C erreicht haben. Die Anzahl an plausiblen geologischen Umgebungen, welche in Betracht gezogen werden können, ist fast grenzenlos. Dies eröffnet die Möglichkeit in verschiedenen Umgebungen verschiedene Reaktionen ablaufen zu lassen. Dabei können die gleichen Edukte je nach Bedingung zu unterschiedlichen Produkten reagieren. Ein anschauliches Beispiel dafür ist die Reaktion von NH_3 und HCN, welche unter hohen Konzentrationen und Wärme zu polymeren Strukturen reagieren (u.a. Nukleobasen),[31] unter verdünnten Bedingungen und UV-Strahlung jedoch zu Harnstoff und Guanidin umgesetzt werden.[104] Ein Blick auf die Bausteine des

Lebens zeigt ebenfalls, dass z.B. Lipide, Zucker oder Nukleobasen unterschiedliche Chemie bzw. Bedingungen für ihre Entstehung benötigen. Diese Chemie muss nicht immer kompatibel mit denen anderer Molekülklassen sein, weshalb die chemische Evolution vermutlich nicht nur an einem bestimmten Ort abgelaufen ist. Aus chemischer Sicht ergibt es mehr Sinn, dass unterschiedliche Molekülklassen in verschiedenen geologischen Umgebungen entstanden und erst später zusammentrafen. Durch dieses Zusammentreffen wurden dann neue Reaktionswege eröffnet, welche weitere Molekülklassen hervorbringen konnten. Ein Beispiel hierfür ist z.B. die Reaktion von Nukleobasen bzw. deren Vorläufermolekülen mit Zuckern für die präbiotische Synthese von Nukleosiden.[105-107]

Ganz gleich, über welche Reaktionswege und Bedingungen die entscheidenden präbiotischen Reaktionen abgelaufen sind, müssen diese früher oder später zu selbst replizierenden Systemen geführt haben. Während an dieser Stelle noch Übereinstimmung herrscht, spaltet die Frage, welches das erste selbstreplizierende System war, die präbiotische Gemeinschaft in zwei Lager. Die beiden in Frage kommenden Systeme könnten entweder aus metabolischen Kreisläufen durch autokatalytische Reaktionen bestehen oder aus selbstreplizierenden Polymeren, welche in der Lage waren, Informationen zu speichern. Beide Theorien sollen im nachfolgenden Abschnitt detaillierter beschrieben werden.

2 Selbstreplizierende Systeme

2.1 Die Entstehung des Lebens: proto-Metabolismus oder genetisches Material?

Alle Lebensformen, wie wir sie heute auf der Erde kennen, nutzen ein System aus Biopolymeren (Nukleinsäuren und Proteine), um genetische Informationen zu speichern bzw. biochemische Reaktionen zu katalysieren. Diese beiden Funktionen können als Grundvoraussetzung des Lebens betrachtet werden. Entscheidend ist hierbei, dass die Biopolymere voneinander abhängig sind. Die genetische Information ist in der DNA gespeichert, welche in RNA transkribiert wird, um anschließend mit Hilfe der Ribosomen durch Translation in Proteine übersetzt zu werden. Diesen Informationsfluss von der DNA über die RNA bis hin zu den Proteinen wird als das `Zentrale Dogma der Molekularbiologie´ bezeichnet. Interessanterweise ist die Biosynthese der DNA wiederum über Enzyme katalysiert, welche durch die DNA selbst codiert werden. Diese biologische Rückkopplung ermöglicht die Replikation (Zellteilung) durch Duplikation des Erbgutes. Der molekulare Mechanismus, mit welchem die Nukleinsäuren ihre genetische Information speichern und duplizieren, konnte basierend auf der von Watson und Crick vorgeschlagenen Doppelhelix Struktur aufgeklärt werden.[108]
Es ist unwahrscheinlich, dass erste primitive Lebensprozesse bereits auf einem so komplexen System aus DNA, RNA und Proteinen basierten. Diese so genannte DNA/RNA/Protein Welt, wie sie heute existiert, ist das Resultat einer Milliarde von Jahren andauernden Evolution ausgehend von einem möglicherweise bedeutend einfacheren selbstreplizierenden System. Im Fokus der präbiotischen Chemie steht demnach die Suche und Entstehung von potentiellen frühzeitlichen autokatalytischen Systemen. Eine essentielle Voraussetzung für solch selbstreplizierende Systeme muss die Speicherung und zuverlässige Übertragung von Informationen sein, um so Darwin'sche Evolution zu ermöglichen. Nach Ansicht der Wissenschaft kommen dabei zwei verschiedene Systeme in Frage:

1) Autokatalytische „proto-metabolische“ Kreisläufe wie z.B. präbiotische Varianten des reduktiven Citrat-Zyklus
2) Autokatalytische Replikationskreisläufe von Templat-basierten Polymeren (z.B. RNA)

2.1.1 Der proto-Metabolismus

Das Leben auf der Erde basiert auf Stoffwechselprozessen, welche aus komplexen chemischen Reaktionen für den Aufbau (Anabolismus) und Abbau (Katabolismus) von Biomolekülen bestehen. Auch wenn diese Prozesse heute von Enzymen gesteuert werden, muss während der chemischen Evolution der Aufbau von komplexen Molekülen in Abwesenheit von Enzymen stattgefunden haben. Hierbei

gehen die Befürworter des proto-Metabolismus davon aus, dass einige grundlegende biologische Stoffwechselprozesse schon vor der Existenz von Enzymen, Nukleinsäuren oder Zellen in einer geochemischen Umgebung spontan entstehen konnten.[109] Diese Stoffwechselprozesse könnten die Entstehung und den grundlegenden Aufbau der heutigen biochemischen Prozesse erklären. In der Tat ähnelt die Redoxchemie in hydrothermalen Quellen den grundlegenden Energie erzeugenden metabolischen Reaktionen in autotrophen Organismen.[101] Hierbei sind vor allem reduktive CO_2-fixierungs Kreisläufe für den Aufbau von molekularer Komplexität von Bedeutung.[110-111] Gegenstand der Forschung sind vor allem der Acetyl-CoenzymA-Weg (Acetyl-CoA-Weg, auch Wood-Ljungdahl-Weg genannt)[102, 112] und der reduktive Citrat Zyklus[113-114] sowie theoretische Kombinationen aus beiden.[115-116] Beim Acetyl-CoA-Weg wird CO_2 und H_2 zu Acetat umgewandelt. Da bei dieser Umsetzung Energie frei wird, könnte dieser metabolische Kreislauf unabhängig von Energiequellen wie UV-Strahlung oder Wärme existiert haben. Die präbiotische reduktive Synthese von Acetat aus CO_2 konnte in geringen Ausbeuten mit Eisen (Fe^0) oder elektrochemisch mit Hilfe von Greigite [$(Fe^{II}(Fe^{III})_2S_4$] erfolgreich gezeigt werden.[117-118] Wächtershäuser konnte schon früher parallelen zwischen der Chemie von Eisen-Schwefel-Verbindungen und dem Acetyl-CoA-Weg aufzeigen, wobei er jedoch nicht von CO_2 sondern von Kohlenstoffmonoxid (CO) ausging. Mit einem Gemisch aus Eisensulfid (FeS) und Nickelsulfid (NiS) konnte aus CO und Methanthiol (MeSH) mit guten Ausbeuten Acetat synthetisiert werden (~0.5%). Die Reaktion läuft über eine aktivierte Thioessigsäure als Intermediat, ähnlich dem Essigsäureester im Acetyl-CoA-Weg, welches anschließend zum Acetat hydrolysiert wird.[119]

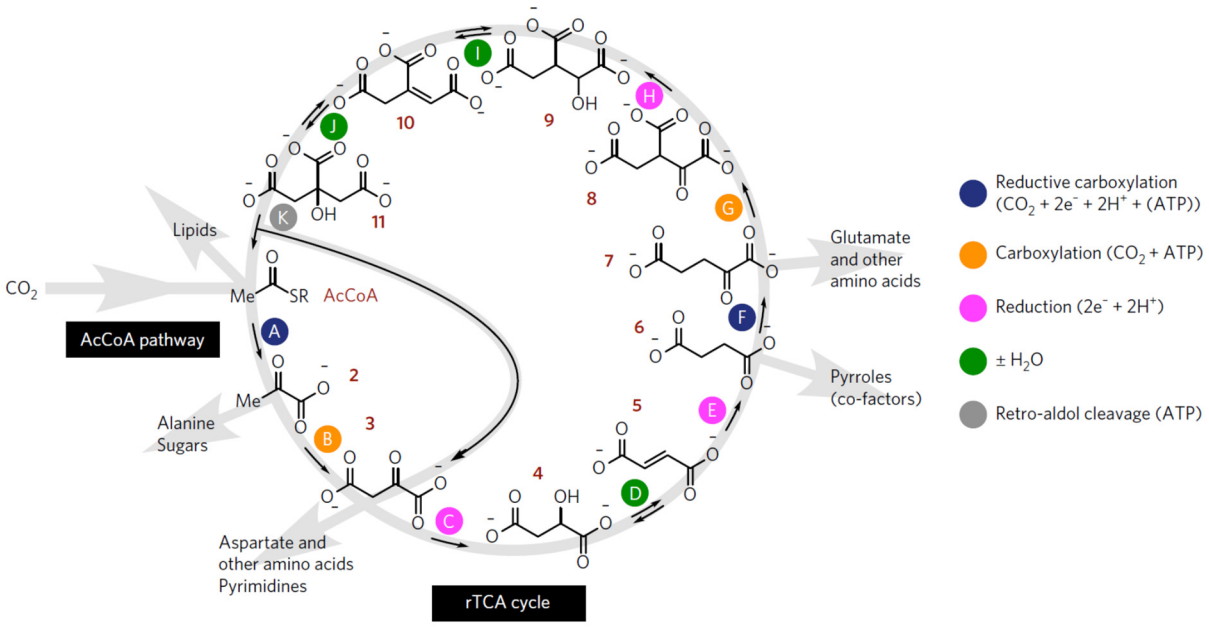


Abbildung 1. Hypothetisches proto-anabolisches Netzwerk, welches den Acetyl-CoA Weg (CO_2 zu AcCoA) als auch den reduktiven Citratzyklus enthält. Aus Muchowska et al., *Nature Ecol. Evol.* **2017**, 1, 1716. Nachdruck mit Genehmigung von SpringerNature.

Zusammen mit dem reduktiven Citratzyklus könnten die grundlegenden Stoffwechselprozesse von autotrophen Organismen wie LUCA genutzt worden sein.[1] Ein mögliches proto-anabolisches Netzwerk aus Kombination des reduktiven Citratzyklus und Acetyl-CoA-Wegs ist in **Abb. 1** gezeigt. Dieses Netzwerk ist autokatalytisch, da jedes Citrat in Oxaloacetat (**3**, **Abb. 1**) und Acetyl-CoA gespalten wird. Über CO_2-Fixierung wird Acetyl-CoA ebenfalls in Oxaloacetat **3** umgewandelt. Somit gibt jedes

eingespeiste Citratmolekül nach Durchlaufen des Kreislaufs zum Schluss zwei Citratmoleküle. Interessanterweise beinhaltet dieser proto-anabolische Zyklus alle fünf universellen metabolischen Vorstufen für die Biosynthese von Aminosäuren, Zuckern, Nukleobasen, Lipiden und von Pyrrolen. Alle Transformationen lassen sich auf nur fünf mechanistisch unterschiedlichen Reaktionsklassen zurückführen, von denen einige durch Metalle katalysiert werden können (**Abb. 1**).[114]

Ein proto-anabolisches Netzwerk ist zwar in der Lage, den grundlegenden Aufbau der heutigen Biochemie zu erklären, muss jedoch eine Vielzahl an Voraussetzungen erfüllen.[120] Laut dem zweiten Hauptsatz der Thermodynamik strebt das Universum nach steigender Entropie, was im Gegensatz zu hochgradig organisierten Stoffwechselprozessen steht. Nichtsdestotrotz kann die Entropie lokal sinken, wenn sie außerhalb stärker ansteigt. Daraus folgt, dass metabolische Kreisläufe eine Abgrenzung von der Umwelt (z.B. Protomembran) benötigen. Um in solch einem lokalen Kompartiment eine hohe Organisation an Stoffwechsel Prozessen aufrecht zu erhalten, ist eine Energiequelle (z.B. Redoxsysteme) essentiell. Reduktive CO_2-Fixierungskreisläufe müssen somit an ein oxidierendes System gekoppelt sein. Dies könnte z.B. durch die Oxidation von Eisensulfid (FeS) zu Pyrit (FeS_2) wie von Wächtershäuser vorgeschlagen geschehen.[110, 112] Um sich nicht selbst auszulöschen, muss ein metabolisches Netzwerk außerdem in der Lage sein zu wachsen, indem neues Material schneller akquiriert als verbraucht wird. Erst durch Reproduktion des Systems nach Erreichen einer bestimmten Größe (z.B. physische Teilung des Kompartiments) ist die Möglichkeit für Darwin'sche Evolution überhaupt denkbar.

Diese Vielzahl an Voraussetzungen für die mögliche Entstehung des Lebens aus einem proto-Metabolismus führt immer wieder zu Kritik an diesem Konzept.[121] Wie könnten Kompartimente ausgesehen haben, um eine Abgrenzung von der Umwelt zu ermöglichen, bevor Lipide überhaupt entstanden waren? Immer wieder werden Poren in hydrothermalen Quellen als mögliche Abgrenzung diskutiert. In solchen Poren könnte das System jedoch nicht unbegrenzt wachsen und replizieren, da die Poren räumlich fest begrenzt sind. Es erscheint fast unmöglich, ein gesamtes metabolisches System von einer Pore in die Nächste wandern zu lassen ohne dabei z.B. vom Ozean verdünnt zu werden. Des Weiteren setzt Wachstum voraus, dass die in metabolischen Kreisläufen katalysierten Reaktionen extrem effizient ablaufen müssen. Einige der Reaktionen sind jedoch reversibel (Reaktion D, I, J in **Abb. 1**) oder aber führen zu einer Vielzahl an Nebenprodukten.[122] Außerdem sind viele Moleküle sich strukturell sehr ähnlich, was sehr spezifische Katalysatoren benötigt, um das eine Molekül umzusetzen das Andere aber nicht. Ein Beispiel hierfür liefert der reduktive Citratzyklus für die gewünschte Reduktion von Oxaloacetat (**3**, **Abb. 1**) zu Malat (**4**, **Abb. 1**). Ein unspezifischer Katalysator würde vermutlich das strukturell sehr ähnliche α-Ketoglutarat (**7**, **Abb. 1**) ebenfalls reduzieren und damit aus dem katalytischen Kreislauf entfernen. Selbst wenn alle Voraussetzungen für ein metabolisches Netzwerk, das heißt ein Kompartiment, eine Aufrechterhaltung der Ordnung, Wachstum, sowie Reproduzierbarkeit des Systems, gegeben sind, bleibt die Frage, ob metabolische Netzwerke überhaupt die Fähigkeit zur evolutiven Weiterentwicklung besitzen können.[121] Dafür ist die Akquirierung von neuen autokatalytischen Reaktionen bzw. neuen autokatalytischen Kreisläufen definiert. Dies ist notwendig, um aus dem angenommenen Ist-Zustand eine höhere Komplexität des metabolischen

Netzwerks zu erreichen. Wie theoretische Studien anhand von Computermodellen zeigen konnten, ist diese Fähigkeit für metabolische Netzwerke nicht gegeben.[123] Schon 1971 schlussfolgerte Eigen, dass die Idee autokatalytisch funktionierender Proteine zur Evolution metabolischer Netzwerke nicht funktionieren kann, da Peptidsequenzen Informationen nicht speichern können. Daraus folgt, dass die Fähigkeit zur Evolution nicht gegeben ist, da ein neues funktionelles Protein (z.B. durch Produktionsfehler) nicht reproduziert werden kann und damit über die Zeit wieder verloren geht.[124]

Zusammenfassend kann gesagt werden, dass die Entstehung von proto-metabolischen Kreisläufen wie z.B. Varianten des Acetyl-CoA-Wegs oder des reduktiven Citratzyklus in einer geochemischen Umgebung nicht auszuschließen ist. Dennoch ist die Entstehung solcher Kreisläufe alleine nicht ausreichend, um die Entstehung des Lebens zu erklären. Selbst wenn alle Voraussetzungen (wie oben beschrieben) für solche Kreisläufe erfüllt wären, besitzen diese nicht die Fähigkeit zur evolutiven Weiterentwicklung. Bis heute fehlt jeglicher experimenteller Beweis für ein autokatalytisches Netzwerk oder Kreislauf, welches in der Lage wäre, eine höhere Komplexität des eigenen Systems zu erreichen. Im Gegensatz dazu sind Templat-basierte Replikatoren (z.B. Nukleinsäuren) in der Lage, Mutationen zu reproduzieren und damit prinzipiell in der Lage, Darwin'sche Evolution einzugehen. Nachfolgend werden genetische Polymere als Möglichkeit für die Entstehung des Lebens diskutiert.

2.1.2 Die Templat-basierte Replikation

Jedes Lebewesen speichert seine Erbinformation in einem genetischen Biopolymer, der DNA. Die dafür verwendeten kanonischen Nukleobasen Adenin (A), Guanin (G), Cytosin (C) und Thymin (T) sind verantwortlich für die Sequenzinformation. Im Vergleich zu einem proto-Metabolismus, bei dem die Informationsübertragung durch positive Rückkopplung anhand von autokatalytischen Reaktionen erfolgt, nutzen Nukleinsäuren ein Templat-basiertes System. Dies ermöglicht das Ablesen und die Replikation der in der Basensequenz gespeicherten Information. Hierfür wird eine spezifische Interaktion für die molekulare Erkennung der einzelnen Nukleobasen benötigt. Dies geschieht über Wasserstoffbrückenbindungen, welche zu den spezifischen Basenpaarungen G/C sowie A/T führen. Zusammen mit den π-Interaktionen der übereinanderliegenden Basenpaare bildet die DNA eine stabile sequenzunabhängige Doppelhelix Struktur aus. Diese Struktur ermöglicht die kompakte Speicherung von Erbinformationen auf engstem Raum. Die Sequenzinformation der DNA wird in RNA transkribiert und anschließend mit Hilfe der Ribosomen durch Translation in Proteine übersetzt. Interessanterweise ist die Biosynthese der DNA wiederum über Enzyme katalysiert, welche durch die DNA selbst codiert werden. Somit werden die beiden Grundfunktionen des Lebens auf zwei verschiedene Biopolymere, den Proteinen (Katalyse) und der DNA (Informationsspeicherung), aufgeteilt. Auf Grund der gegenseitigen Abhängigkeit dieser beiden Biopolymere ergibt sich für den Ursprung des Lebens ein klassisches Henne-Ei-Problem. Was entstand als erstes: Proteine oder DNA?

Die Antwort auf die Frage könnte die RNA liefern. Aus chemischer Sicht ist die RNA der DNA sehr ähnlich. Sie kann daher ebenfalls Informationen in ihrer Basensequenz speichern. Zusätzlich zur

Informationsspeicherung ist RNA außerdem im Stande komplexe tertiäre Strukturen auszubilden, welche es ermöglichen, chemische Reaktionen zu katalysieren. Noch bevor die erste katalytische RNA nachgewiesen werden konnte, stellten Orgel und Woese fest, dass autonome RNA „Organismen“ im Prinzip existieren könnten, aber nur wenn RNA in der Lage wäre, einige Funktionen von Proteinen zu übernehmen. Sie spekulierten, dass Coenzyme molekulare Fossilien, aus einer Zeit, in der RNA Moleküle ohne die Hilfe von Proteinen arbeiteten, wären.[125-127] Mit dem Nachweis der katalytischen Aktivität von RNA in den 1980er Jahren durch Sidney Altman[128] und Thomas Cech[129] rückte die RNA endgültig in den Fokus als zentrales Molekül während der chemischen Evolution. Selbst die für Proteinbiosynthese verantwortlichen Ribosomen konnten als Ribozyme identifiziert werden.[130] Dies unterstützt die Hypothese, dass unsere heutige DNA/RNA/Protein Welt ein evolutionäres Produkt einer ersten RNA-Welt sein könnte, in der DNA und Proteine erst später auftraten.[131]

Die RNA-Welt-Hypothese geht davon aus, dass die DNA-Basen zum Großteil von der RNA übernommen wurden. Der einzige Unterschied zwischen den kanonischen Basen der RNA und der DNA ist die Verwendung von T statt U in der DNA. Evolutionär ergibt dieser Unterschied Sinn, da die DNA für die Langzeitspeicherung der genetischen Information optimiert ist. Somit kann die DNA durch Verwendung von T statt U die Identität der G/C Basenpaarung nach einer ungewollten Deaminierung von C erhalten. Interessanterweise wird der DNA-Baustein Thymidin aus dem RNA-Baustein Uridin biochemisch synthetisiert. Dies impliziert, dass DNA tatsächlich erst später aus der RNA entstanden sein könnte. Ein weiterer evolutionär sinnvoller Unterschied von RNA zu DNA ist die Abwesenheit der 2`-OH Gruppe im Zucker (daher der Name Desoxyribonukleinsäure). Dieser Unterschied ist ebenfalls essentiell für die zuverlässige Langzeitspeicherung von Erbinformationen. Die fehlende 2`-OH Gruppe in der DNA führt zu stabileren Phosphodiesterbindungen und damit zu einer erheblich höheren Halbwertszeit der DNA im Vergleich zur RNA.[132]

Trotz aller biologischen Indizien, welche darauf hinweisen, dass RNA tatsächlich vor der DNA und den Proteinen entstanden sein könnte, erklärt ein reiner RNA-Organismus noch nicht die Entstehung des Lebens per se. Da RNA beide fundamentalen Prozesse des Lebens (Katalyse und Informationsspeicherung) in einem Molekül vereint, ergibt sich aber die Möglichkeit für die Entstehung eines selbstreplizierenden RNA-Moleküls, welches exponentielles Wachstum unabhängig von Proteinen ermöglichen könnte. Im Prinzip wäre ein einziges selbstreplizierendes RNA-Molekül ausreichend gewesen, um Leben über Darwin’sche Evolution entstehen zu lassen, solange die RNA-Bausteine in genügender Menge vorhanden waren. Die Fähigkeit zur Evolution durch Mutation konnte für RNA in Abwesenheit von Zellen durch Spiegelmann in einem wegweisenden Experiment bewiesen werden.[133] Zwar konnte bisher noch keine einzelne selbst replizierende RNA experimentell nachgewiesen werden, dennoch zeigen neuere Arbeiten im Bereich der *in vitro* Evolution einige Erfolge auf.[134-137] Auch wenn die RNA-Welt-Hypothese die Entstehung des Lebens vereinfacht, indem sie die Zahl an beteiligten Polymeren reduziert, gibt es wenig Grund zur Annahme, dass nicht auch Lipidmembranen, Polysaccharide oder nicht-codierte Polypeptide gleichzeitig mit Nukleinsäuren entstanden sein könnten. Basierend auf diesen Überlegungen ist die Entstehung von RNA auf einer frühen Erde von zentraler Bedeutung, um den Ursprung des Lebens zu verstehen. Unterstützer des

proto-Metabolismus argumentieren hingegen, dass RNA strukturell zu komplex sei für eine spontane Entstehung auf der frühen Erde.[120]

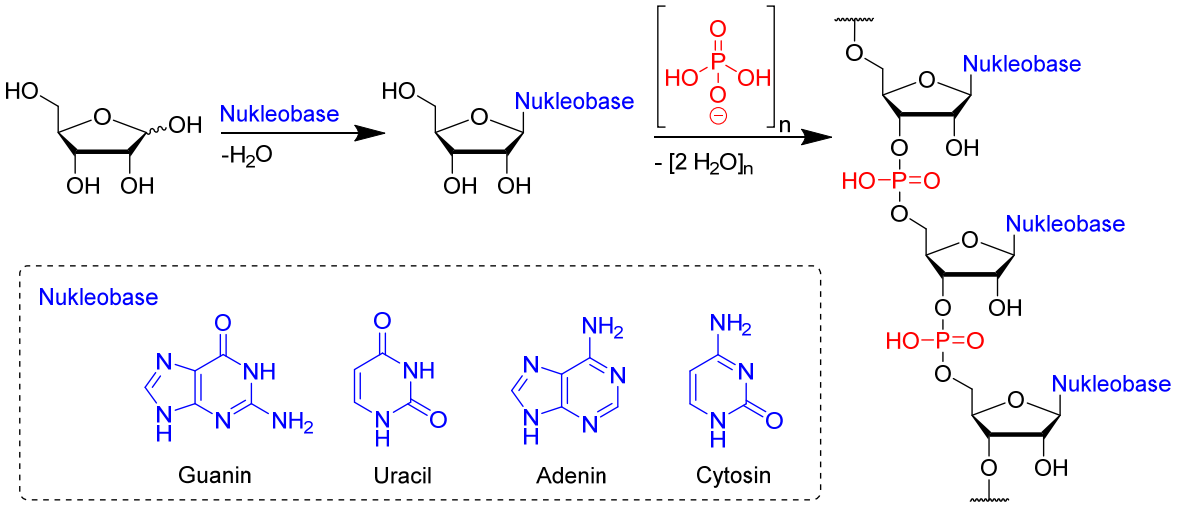


Abbildung 2. Aufbau der RNA bestehend aus Nukleotidmonomeren. Die Nukleobasen (blau) sowie Phosphat (rot) sind hervorgehoben. Verknüpfung von Ribose (schwarz) mit der Nukleobase entspricht einer Kondensationsreaktion. Verknüpfung der Nukleosidmonomere zu einem RNA-Polymer entspricht abermals einer Kondensationsreaktion mit Phosphat.

Ein genauer Blick auf die Struktur der RNA zeigt, dass diese aus den Nukleotidmonomeren bestehen, welche über Phosphodiesterbindungen verknüpft sind. Die Nukleotide bestehen aus drei Substrukturen: den Nukleobasen, der Ribose und einem Phosphat (**Abb. 2**). Während die heterozyklischen Nukleobasen vermutlich polymere Strukturen aus HCN[31] oder eng verwandten Molekülen (Formamid, Amidin, Ammoniumformiat, Harnstoff oder Malonodinitril)[138-140] sind, scheint Ribose ein polymeres Produkt von Formaldehyde zu sein.[15] Hierbei entsteht Ribose zusammen mit einer Vielzahl an anderen Zuckern mit einer Ausbeute von <1%. Phosphat ist auf der Erdoberfläche zwar vorhanden, war vermutlich aber nur in kleinen Mengen in Lösung verfügbar, da es unlösliche Mineralien vor allem mit divalenten Metallkationen (z.B. Ca^{2+}, Mg^{2+} oder Fe^{2+}) bildet. Die drei Nukleotidkomponenten sind über chemische Bindungen miteinander verknüpft. Diese Bindungen sind im Prinzip das Resultat von Kondensationsreaktionen (**Abb. 2**), welche aus thermodynamischer Sicht nur in Abwesenheit von Wasser effizient ablaufen können. Vor allem die glykosidische C-N Bindung zwischen den Nukleobasen (Purine bzw. Pyrimidine) und dem Zucker stellt in der präbiotischen Chemie eine große Herausforderung dar. Die direkte Verknüpfung von kanonischen Purinen und Ribose funktioniert nur in sehr geringen Ausbeuten, wohingegen die gleiche Reaktion mit den kanonischen Pyrimidinen überhaupt nicht funktioniert. Es scheint also, dass die präbiotische RNA-Synthese tatsächlich mit einer Vielzahl an Problemen verknüpft ist. Wie kann Ribose z.B. in großen Mengen synthetisiert werden und woher kommt ihre Chiralität? Wie kann Phosphat für präbiotische Reaktionen verfügbar gemacht werden? Wie und in welcher Umgebung sind die Bindungen entstanden, die zu den RNA-Monomeren führten? Wie konnten die Monomere zu langen Polymeren verknüpft werden? Gab es möglicherweise ein einfacheres genetisches Polymer, aus dem die RNA hervorgegangen ist? Auf viele Fragen gibt es noch keine oder nur teilweise Antworten. Trotzdem hat sich die präbiotische Chemie seit dem Urey-Miller Experiment vor ca. 60 Jahren rasant entwickelt und weist regelmäßig neue Erfolge auf. Nachfolgend soll detaillierter auf die Frage eingegangen werden, wie eine RNA-Welt aus präbiotischer Sicht entstanden sein könnte.

3 Die RNA-Welt

3.1 Die Entstehung des Lebens: Präbiotische Synthese von RNA

Vor ca. 4.5 Milliarden Jahren entstand unser Planet, die Erde. Sie war ein Planet, der kaum lebensfeindlicher hätte sein können, mit Oberflächentemperaturen von ca. 1000°C. Doch nach nur 100-200 Millionen Jahren war die Erde soweit abgekühlt, dass die Oberfläche zur Erdkruste mit hoher vulkanischer Aktivität erstarrt war und Wasser im flüssigen Zustand existieren konnte.[80] Interessanterweise wird das erste Leben durch fossile Funde bereits auf ein Alter von ca. 3.5-4 Milliarden Jahren geschätzt.[5, 141-143] Auf der geologischen Zeitskala ist das eine verblüffend geringe Zeitspanne von lediglich einigen 100 Millionen Jahren von der Entstehung der Erde hin zum ersten Leben. In diesem Zeitraum müssen über unbekannte Reaktionspfade die Moleküle des Lebens wie Nukleoside, Lipide oder Aminosäuren entstanden sein. Vor allem die Entstehung von RNA als zentrales Molekül für die chemische Evolution ist von enormer Bedeutung. Unter welchen geologischen Gegebenheiten und über welche präbiotisch plausiblen Reaktionspfade konnte RNA aus kleinsten anorganischen Molekülen entstehen, um so die chemische Evolution in Gang zu setzen?

3.1.1 Von anorganischen zu organischen Verbindungen

Ausgehend von einer Ur-Atmosphäre bestehend vor allem aus H_2O, N_2 und CO_2 zusammen mit geringen Mengen an Thiolen, H_2, einfachen Aminen und CH_4 entstanden die ersten komplexeren organischen Moleküle. Diese Reaktionen wurden unter anderem durch Sonnenwinde, UV-Strahlung, elektrische Entladung oder vulkanische Aktivität vorangetrieben. Durch Reaktion von H_2O und CO_2 konnten einfachste Zuckervorläufer wie Formaldehyd oder Glycolaldehyd entstehen.[16] Auch HCN, als Produkt von N_2 und CH_4, ist in der Lage, Zuckervorläufermoleküle über Photoredoxchemie zu liefern.[34, 144] Wässriges HCN kann ebenfalls zu verschiedenen Nitrilen, Amidinen oder Harnstoff mit Hilfe von NH_3 reagieren.[104, 145] NH_3 entstand vermutlich durch Reduktion von Nitrit bzw. Nitrat, welche sich in Wasser aus NO bildete.[38-40] Das benötigte NO entstand direkt aus einer Uratmosphäre durch Reaktion von N_2 und CO_2.[35-36] Damit konnte aus einfachsten anorganischen Molekülen bereits ein ganzes Repertoire an präbiotischen Vorstufen für die Synthese von Zuckern und Heterozyklen entstehen. HCN und dessen Hydrolyseprodukt Formamid sind hierbei die wohl am häufigsten verwendeten Substanzen, da sie eine Vielzahl an präbiotisch relevanten Molekülen liefern können.[62]

3.1.2 Die Nukleobasen

Juan Oró konnte als erstes zeigen, dass Adenin über HCN-Polymerisation entstehen kann.[31] Durch seine Struktur kann Adenin als HCN-Pentamer betrachtet werden, welches aus einer konzentrierten wässrigen NH_4CN Lösung mit Ausbeuten von ca. 0,05-0,23% entsteht. Neben Adenin wurden andere Purin Vorstufen wie 4-Aminoimidazol-5-carboxamid, 4-Aminoimidazole-5-carboxamidine sowie Formamidin und Formamid gefunden. Auf Grund dieser Produkte entwickelte Oró den Reaktionsweg, welcher zum Adenin und anderen Purinen führte (**Abb. 3**).[146-147] Später wurde auch Guanin, wenn auch in deutlich geringeren Ausbeuten, nachgewiesen. Guanin könnte direkt oder aber über Hydrolyse von Diaminopurin (DA) entstanden sein.[148] In den Reaktionen konnten später auch weitere Purine wie Xanthin oder Inosin nachgewiesen werden.[79]

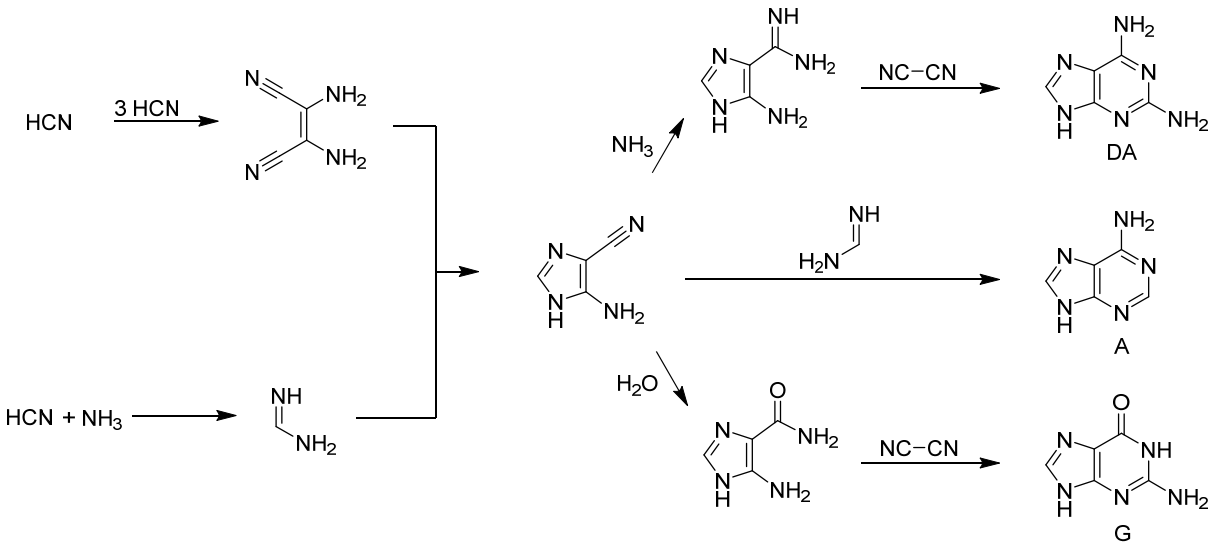


Abbildung 3. Purinbasensynthese aus einer NH_4CN Lösung.

Interessanterweise wurden die kanonischen Pyrimidinbasen in den ersten NH_4CN-Reaktionen nicht gefunden. Leidglich Orotsäure, 5-Hydroxyuracil und 4,5-Dihydroxypyrimidin wurden als Pyrimidine identifiziert.[149] Erst in einer 27 Jahre alten, bei -78°C aufbewahrten Probe konnte Uracil (0,0017%), aber nicht Cytosin als kanonische Nukleobase neben vielen anderen Heterozyklen direkt aus einer NH_4CN Lösung nachgewiesen werden.[79] Cytosin entsteht jedoch mit bis zu 5% Ausbeute, wenn Kaliumcyanat (KOCN) zusammen mit Cyanoacetylen reagiert, welches nach HCN das zweithäufigste Produkt einer elektrischen Entladung von N_2 und CH_4 ist.[150] Cyanoacetylen ist allgegenwärtig im Universum zu finden und kann damit als mögliches präbiotisches Ausgangsmaterial angenommen werden.[151-152]

Insgesamt scheint die präbiotische Synthese der Nukleobasen aus HCN sehr ineffizient abzulaufen. Zum einen sind die Ausbeuten sehr gering, da HCN hauptsächlich zu polymerem Material reagiert, zum anderen entstehen die Heterozyklen nur bei hohen Konzentrationen (1-15 M), da unter verdünnten Bedingungen Hydrolysereaktionen überwiegen.[145] Auf Grund der Volatilität von HCN scheint es fast unmöglich, solch hohe Konzentrationen in den Ozeanen oder in anderen Gewässern zu erreichen. HCN könnte jedoch über zwei verschiedene Mechanismen angereichert worden sein. Zum einen bildet HCN im Komplex mit Fe^{2+} stabile Hexacyanoferrate,[153] zum anderen kann sich bei -23.4 °C eine eutektische

Lösung bilden bei der HCN bis zu einer Konzentration von 74.5 mol% angereichert hätte werden können.[74] Trotz dieser Möglichkeiten zur Anreicherung bleibt die Frage, ob HCN überhaupt in signifikant hoher Menge auf einer frühen Erde vorhanden war, um schneller angereichert als hydrolysiert zu werden. Die Bildung von HCN läuft nur effizient ab, wenn große Mengen an CH_4 vorhanden sind. In einer neutralen Atmosphäre (N_2, CO_2 und H_2O), wie auf der frühen Erde angenommen, entsteht HCN jedoch nur in Spuren.[154-155] Damit sind hohe Konzentrationen, wie für die Synthese von Nukleobasen benötigt, wahrscheinlich niemals erreicht worden.

Auf Grund dieser Gegebenheit könnte ein Großteil des vorhandenen HCN zu Formamid und Ameisensäure hydrolysiert oder durch Sonnen- bzw. UV-Strahlung zu Guanidin, Harnstoff und Cyanamid umgesetzt worden sein.[104] Für alle erwähnten Produkte konnte gezeigt werden, dass diese ebenfalls über unterschiedliche Reaktionen in der Lage sind, neben Purinen und Pyrimidinen auch andere Heterozyklen, z.B. Triazine, zu bilden.[139, 156-159] Über die präbiotische Relevanz vor allem von nicht-kanonischen Basen soll später detailliert eingegangen werden (siehe Abschnitt 3.2). Eine Reaktion, die besonders hervorgehoben werden sollte, ist die erste effiziente Synthese von Cytosin aus Harnstoff und Cyanoacetylen mit 30-50% Ausbeute. Über Hydrolyse ist Uracil ebenfalls direkt verfügbar.[160] Bis heute kann Cytosin und andere Pyrimidine nur in Kombination mit Cyanoacetylen bzw. dessen Hydrolyseprodukt Cyanoacetaldehyd präbiotisch effizient synthetisiert werden.[150, 157, 161] In diesen Reaktionen lassen sich Purine maximal in Spuren finden. Es scheint also, dass Purine eher über HCN-Polymerisation mit einem Imidazolderivat als Zwischenstufe entstehen (**Abb. 3**), während Pyrimidine über die direkte Verknüpfung von C3 Einheiten mit Harnstoff, Guanidin oder Cyanat entstehen. Je nach Ausgangsmaterial können sich also entweder Purine oder Pyrimidine bilden, aber selten beide zusammen in guten Ausbeuten.

Im Gegensatz dazu ist Formamid als Hydrolyseprodukt von HCN in der Lage sowohl die kanonischen Purine als auch Pyrimidine in geringen Ausbeuten darzustellen. Je nach Katalysator können unterschiedlichste Produktverteilungen erreicht werden.[158, 162-165] Hierbei ist der genaue Mechanismus für die Entstehung der Heterozyklen noch nicht so detailliert erforscht wie für die Entstehung von Heterozyklen aus HCN.[166]

Ein alternativer Syntheseweg zu den Purinen, welcher nicht über ein Imidazol Intermediat verläuft, wurde von Traube beschrieben. Dabei verwendet Traube Formamidopyrimidine (FaPys), welche unter basischen oder thermischen Bedingungen zu Purinen reagieren.[167] Die FaPys können aus Aminopyrimidinen über Formylierung mit Ameisensäure oder Formamid hergestellt werden. Wie von der Arbeitsgruppe Carell mechanistisch gezeigt werden konnte, ist diese Formylierung extrem regiospezifisch für das N-5 formylierte Produkt.[107] Die benötigten Aminopyrimidine können z.B. aus dem HCN-Trimer Aminomalonodinitiril sowie dessen Hydrolyseprodukt Aminocyanoacetamid mit Hilfe von Guanidin dargestellt werden. Alternativ können Aminopyrimidine über Reduktion von Nitroso-Pyrimidinen und anschließender Formylierung entstehen. Nitroso-Pyrimidine lassen sich aus Malonodinitril und verschiedenen Amidinen darstellen (**Abb. 4**).[140]

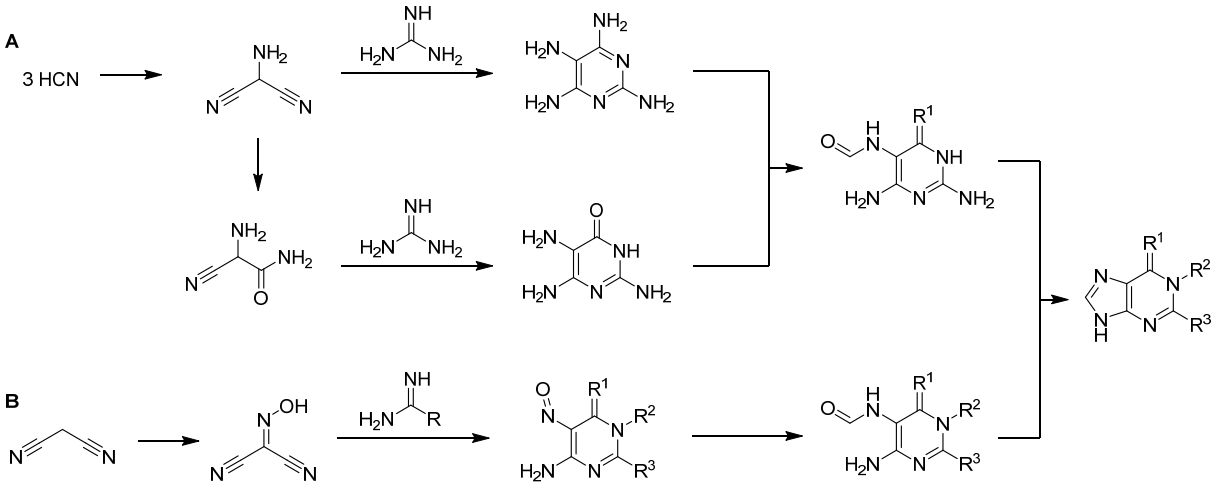


Abbildung 4. Purinsynthese aus Formamidopyrimdinen (FaPys). Weg A ausgehend von HCN, Weg B ausgehend von Malonodinitril.

3.1.3 Die Zucker

Ein weiterer wichtiger molekularer Baustein des Lebens sind die Zucker bzw. Zuckervorläufer, welche unter präbiotischen Bedingungen entstanden sein müssen. Pentosen sind zum Beispiel Untereinheiten von DNA und RNA. Die älteste und bekannteste Zuckersynthese unter präbiotischen Bedingungen ist die Formosereaktion nach Butlerow.[15] Die beiden Ausgangsstoffe Formaldehyd und Glycolaldehyd, können aus einer feuchten, CO_2-reichen Atmosphäre über elektrische Entladung entstehen.[16] Der entstandene Formaldehyd, welcher Spuren von Glykolaldehyd enthält, wird bei der Formosereaktion in Anwesenheit von $Ca(OH)_2$ erhitzt. Im ersten Schritt gibt die Aldoladdition von Formaldehyd und Glycolaldehyd den C3-Baustein Glycerinaldehyd. Anschließend finden eine Vielzahl von weiteren Aldol-, Cannizzaro- und Lobry de Bruyn-Reaktionen statt, die zu einem komplexen Produktgemisch führen.[168] Hierbei entsteht Ribose als Baustein der RNA in <1% mit einer Vielzahl an weiteren Zuckern, was die ursprüngliche Formosereaktion als Zuckerquelle in der präbiotischen Synthese von RNA problematisch erscheinen lässt. Die Formosereaktion wurde deshalb vielfach modifiziert vor allem durch Zugabe verschiedener Katalysatoren, welche die Synthese von Ribose begünstigen sollten.[72, 169-171] Hierbei konnte durch die Gruppe um Benner gezeigt werden, dass besonders Borate in der Lage sind, Pentosen zu stabilisieren, was deren Entstehung begünstigt.[71, 172] Eine weitere interessante Modifizierung der Formosereaktion ist die Verwendung von Glycolaldehydphosphat als C2-Baustein sowie Formaldehyd als C1-Baustein, was in einer Ausbeute von 16% an Ribopyranose-2,4-diphosphat führt.[173] Kritisch ist hierbei jedoch die selektive Ausbildung des Pyranoserings statt des natürlichen Furanoserings. Das verwendete Glycolaldehydphosphat kann mit Hilfe von Amidotriphosphat oder Diamidophosphat aus Glycolaldehyd dargestellt werden.[174-175] Ein weiterer Ansatz zur Optimierung der Formosereaktion ist die Synthese von Zuckern aus Formaldehyd in Micellen, was zu sehr guten Ausbeuten an Pentosen von ca. 65% führt.[176] Leider liefert auch dieser elegante Ansatz Ribose nicht selektiv. Im Gegensatz dazu scheint Hydroxylapatit $Ca_5(PO_4)_3(OH)$ in der Lage zu sein die stereoselektive Synthese von Ribose zu katalysieren. Dabei zeigen die Autoren, dass ein 10:1 Gemisch aus Formaldehyd und Glykolaldehyd nach 42 Std. fast quantitativ zu Dihydroxyaceton umgesetzt wird.

Das Dihydroxyaceton kann dann mit Glykolaldehyd (1:1) zu den zwei möglichen Ketosen (Ribulose bzw. Xylulose) reagieren. Die Analyse der Reaktion konnte zeigen, dass Ribose (20%) aber keine Arabinose entsteht. Damit muss das Hydroxylapatit in der Lage sein, Ribulose (Ketose) zu Ribose (Aldose) stereoselektiv zu isomerisieren. Die Isomerisierung scheint extrem selektiv abzulaufen, da weder Xylose noch Lyxose, als Isomer der Xylulose (11%), in der Reaktion gefunden wurden (Abb. 5).[177]

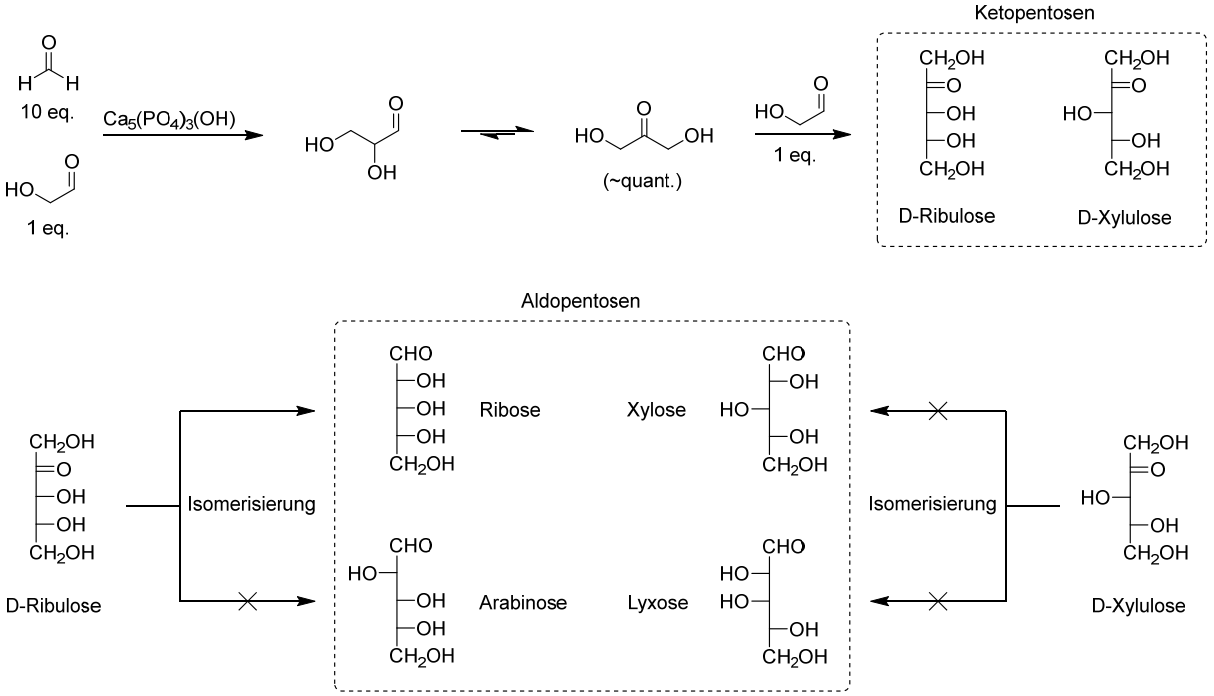


Abbildung 5. Zuckersynthese mit Hydroxylapatit als Katalysator nach Usami und Okamoto.[177] Formaldehyd und Glykolaldehyd (10:1) reagieren zu Glycerinaldehyd, welches zu Dihydroxyaceton isomerisiert. Das Dihydroxyaceton kann mit Glykolaldehyd (1:1) zu D/L-Ketopentosen reagieren (nur D-Enantiomere sind gezeigt). Dabei scheint die homo-Aldolkondensation des Glykolaldehyds zu Tetrosen, durch das Hydroxylapatit unterdrückt zu sein. Nur die entstandene Ribulose wird stereoselektiv zu Ribose (20%) aber nicht Arabinose isomerisiert. Die Isomerisierung der entstandenen Xylulose (11%) wird von Hydroxylapatit in dieser Reaktion nicht katalysiert. Ebenfalls findet sich Glykolsäure (40%), Dihydroxyaceton (20%) sowie nicht isomerisierte Ribulose (8%) in der Reaktion.

Obwohl deutliche Optimierungen der Formosereaktion für die Synthese von Ribose gezeigt werden konnten, steht diese Art der Zuckersynthese in Teilen der präbiotischen Gemeinschaft immer noch als ineffizient in der Kritik. Bei näheren Untersuchungen der Formosereaktion konnte gezeigt werden, dass vor allem C2- und C3-Bausteine zu Pentosen in hohen Ausbeuten (ca. 47%) reagieren können.[178] Das Problem hierbei war, dass der bis dahin einzige präbiotische Zugang zu diesen Zuckervorläufermolekülen über die Formosereaktion selbst verlief. Erst vor kurzem konnte die Gruppe um Sutherland zeigen, dass größere Mengen an Glykol- bzw. Glycerinaldehyd über Photoreduktion ausgehend von $HCN_{(aq)}$ entstehen können.[34, 144] Jedoch wird auch diese Syntheseroute über die direkte Verwendung von C2- und C3-Bausteinen als kritisch gesehen. Der größte Schwachpunkt ist hierbei die hohe Reaktivität von Glycol- bzw. Glycerinaldehyd, welche eine Anreicherung der beiden Moleküle fast unmöglich macht. Die Gruppe um Powner konnte jedoch zeigen, dass sich Glykolaldehyd und Glycerinaldehyd über Aminalbildung nicht nur stabilisieren, sondern auch anreichern lassen.[179] Damit könnten Pentosen mit Hilfe dieser C2- und C3-Bausteine effizient hergestellt werden. Jedoch bleibt hierbei immer noch die Frage der Selektivität offen, da alle vier Aldopentosen mit etwa gleicher Ausbeute entstehen.[178] Das Problem der Selektivität wurde teilweise durch Verwendung von Zink-

Prolin-Komplexen gelöst. Hierbei entstehen aus Glykolaldehyd und Glycerinaldehyd die Aldopentosen Ribose (19%) und Lyxose (21%) sowie Arabinose (14%) und Xylose (9%) (**Abb. 6**).[180]

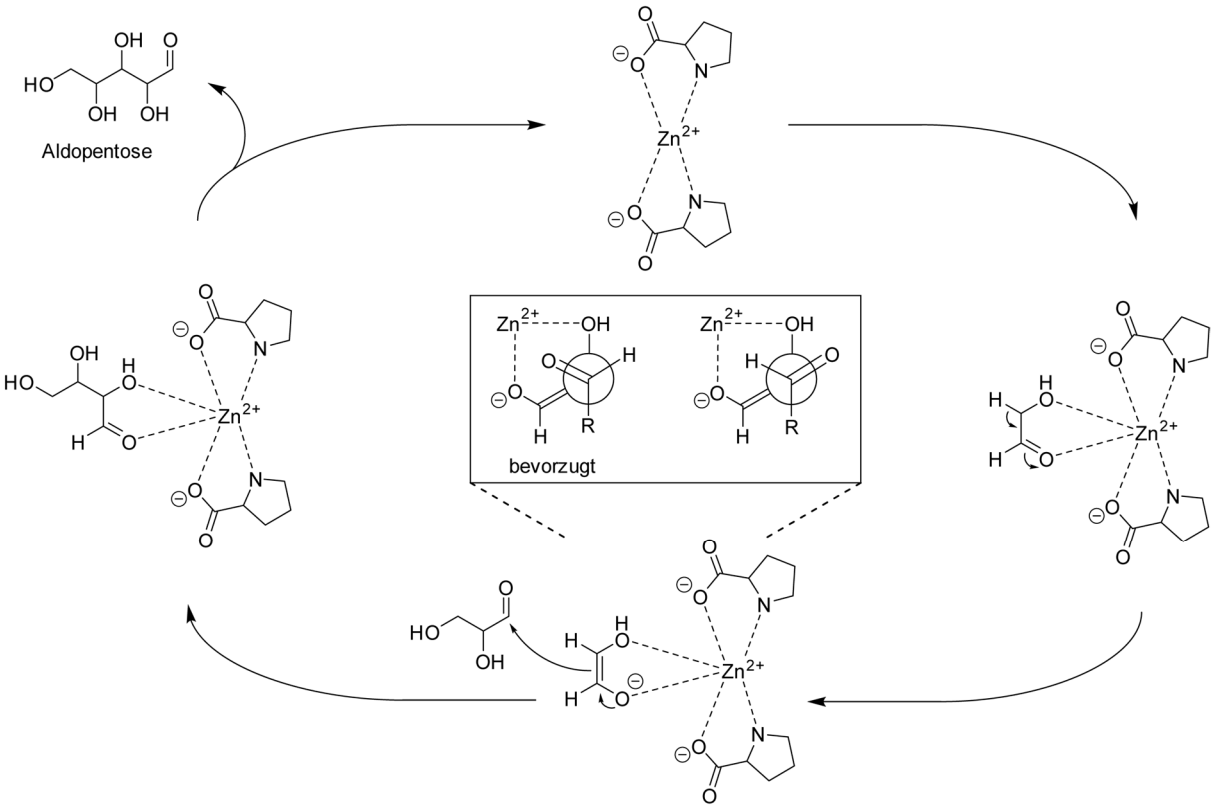


Abbildung 6. Synthese von Pentosen mit Zink-Prolin-Komplexen aus Glykolaldehyd und Glycerinaldehyd.[180] Durch homo-Aldolkondensation von Glykolaldehyd entstehen Erythrose (7%) und Threose (13%). Hexosen (14%) wurden ebenfalls im Reaktionsgemisch gefunden.

Weiterhin bleibt die Synthese von Ribose unter präbiotischen Bedingungen problematisch. Die kürzlich beschriebene Zuckersynthese mit Hilfe von Hydroxylapatit lässt aber hoffen, dass Ribose schon bald stereoselektiv in größeren Mengen aus Dihydroxyaceton und Glykolaldehyd hergestellt werden kann. Auf Grund der verschiedenen Problematiken wurden kürzlich ganz neue Ansätze zur Synthese von Zuckern entwickelt. So konnte die Gruppe um Meierhenrich zeigen, dass Ribose mit Hilfe von UV-Strahlung aus interstellaren Eisanaloga (H_2O, NH_3 und MeOH) entstehen kann.[181] Doch auch dieser Ansatz löst ein grundsätzliches Problem nicht: die Chiralität der Zucker. Bisher ist völlig unklar, wie D-Ribose, als Baustein der RNA selektiv entstehen oder angereichert werden kann. Ein mögliches Szenario für einen Enantiomerenüberschuss von Biomolekülen könnte polarisiertes Licht sein. Dies konnte zumindest für Aminosäuren schon gezeigt werden.[182] Auch chirale Mineralien könnten bei der enantioselektiven Synthese von präbiotischen Molekülen eine Rolle gespielt haben.[183] Alternativ könnte ein Enantiomerenüberschuss nicht über die Synthese selbst, sondern lokal über Anreicherung eines Enantiomers aus einem racemischen Gemisch entstanden sein.[184-185]

3.1.4 Die RNA-Bausteine

Zucker sowie Nukleobasen konnten, wie oben diskutiert, unter präbiotischen Bedingungen entstehen. Die Frage bleibt, wie sich aus diesen beiden Komponenten die RNA-Bausteine (Nukleoside bzw. Nukleotide) bilden konnten, aus denen die RNA gebildet wird. Der einfachste Weg wäre die direkte Verknüpfung von Ribose mit der entsprechenden Nukleobase. Dieser Syntheseweg wurde ausführlich von Orgel für die Purine und Pyrimidine untersucht. In diesen Untersuchungen konnte gezeigt werden, dass die Purine Hypoxanthin, Guanin und Adenin mit Ribose säurekatalysiert zu den entsprechenden Nukleosiden reagieren können (**Abb. 7**). Auf Grund fehlender Regioselektivität und geringer Nukleophilie können die natürlichen N-9 β-Isomere allerdings nur in geringen Ausbeuten erhalten werden (8% für Inosin, 4% für Adenosin und 9% für Guanosin).[105] Orgel versuchte die gleiche Reaktion auch mit den Pyrimidinbasen (Uracil, Thymin und Cytosin), konnte aber keine Nukleoside detektieren.[106]

Der Hauptgrund für die ineffiziente Glykosidierung ist die geringe Nukleophilie der Nukleobasen. Dieses Problem kann über unterschiedliche Strategien umgangen werden. Ein möglicher Ansatz ist die Verwendung von aktivierten Zuckern für die Glykosidierung, ähnlich dem heutigen Salvage-Pathway.[186] Die freie Enthalpie (Gibbs Energie) ist bei der Kondensation der Nukleobase mit aktivierter Ribose im Vergleich zu nicht aktivierter Ribose meist negativ.[187-188] In verschiedenen Arbeiten wurde dieser energetische Vorteil genutzt, um zuerst eine aktivierte Ribose zu synthetisieren und diese mit einer Nukleobase umzusetzen. Wenn Ribose und Phosphat auf einer Siliziumdioxid (SiO_2) Oberfläche adsorbiert und erwärmt werden, entsteht 5-Phosphoribosyl-1-pyrophosphat. Wurde zu diesem Ansatz Adenin gegeben, konnte über Matrix–Assistierte Laser–Desorption–Ionisierung (MALDI) ein Adenosinmonophosphat-Signal, ohne die Angabe von exakten Ausbeuten, gefunden werden.[189] Die Gruppe von Zare konnte ebenfalls zeigen, dass aktivierte Zucker (Ribose-1-phosphat) in Mikrotröpfchen aus Ribose und Phosphorsäure entstehen. Wurden die kanonischen Nukleobasen in Gegenwart von Mg^{2+} zu dieser Reaktionsmischung gegeben, konnten Uridin (2,5%), Adenosin (2,5%), Cytidin (0,7%) und Inosin (1,7%) nachgewiesen werden. Guanosin wurde vermutlich auf Grund der geringen Löslichkeit von Guanin in Wasser nicht gefunden.[188, 190] Die Gruppe um Benner konnte zeigen, dass die direkte Kondensation der Nukleobase mit Hilfe von Ribose-1,2-cyclic Phosphat[175] stereoselektiv mit hohen Ausbeuten erfolgen kann. Dabei konnten immer die natürlichen β-ribofuranosyl-Isomere von Adenosin (15%), Inosin (14%) sowie DA-Nukleosid (12%) nachgewiesen werden. Guanosin konnte durch diese Methode jedoch nicht synthetisiert werden (**Abb. 7**). Ebenfalls konnten auch keine kanonischen Pyrimidine nachgewiesen werden. Lediglich Zebularin (30%),[191] wie schon vorher mit freien Zuckern gezeigt,[192] war reaktiv genug, um die C-N glykosidische Bindung auszubilden. Überraschenderweise konnte über die gleiche Methode auch ein Nicotinamidnukleosid (36%) synthetisiert werden, obwohl hierbei ein positiv geladener Pyridinium-Heterozyklus durch die glykosidische Bindung entsteht.[193]

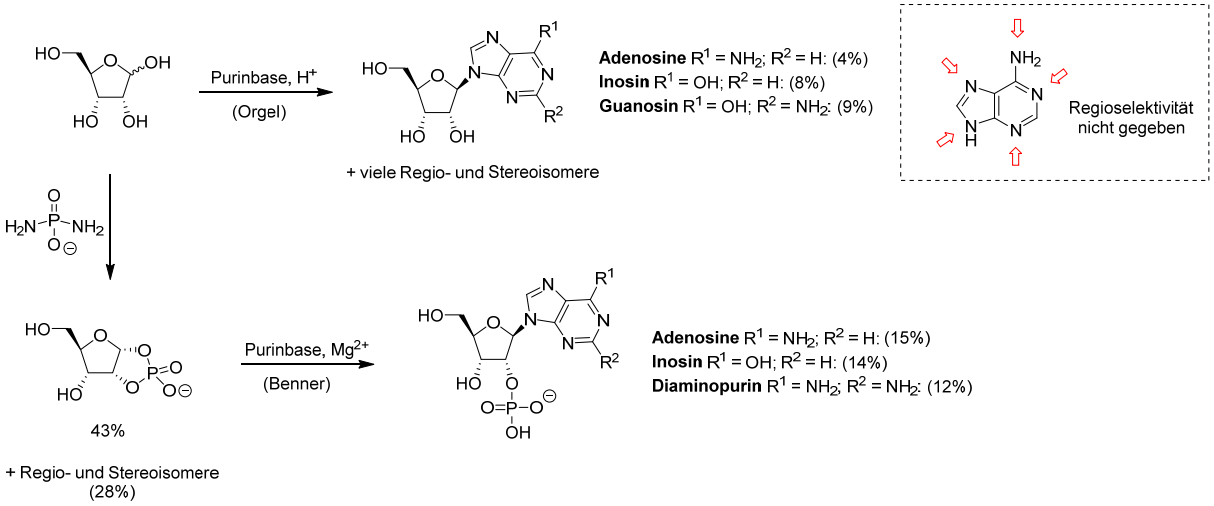


Abbildung 7. Direkte glykosidische Verknüpfung von Purinnukleobasen nach Orgel und Benner. Die Phosphorylierung der Ribose nach Krishnamurthy gibt ebenfalls Ribose-2,3-cyclo-phosphat (28%) als Furanosyl- und Pyranosyl-Isomer.[175] Die Glykosidierung von Hypoxanthin mit aktivierter Ribose gibt ein zweites, nicht identifiziertes Isomer. Guanosin ist unter den Reaktionsbedingungen von Benner nicht verfügbar.

Andere nicht kanonische Pyrimidine sind ebenfalls reaktiv genug, um direkt mit freier Ribose zu reagieren. 2,4,6-Triaminopyrimidin (TAP) bildet Nukleoside mit 60-90% Ausbeute, wenn es in mehreren Nass-Trocken-Zyklen mit Ribose reagiert wird. Dabei bildet sich das β-Ribofuranosylnukleosid (20%, TARC) als eines der Hauptprodukte. Interessanterweise besitzt TARC jedoch keine C-N, sondern eine C-C glykosidische Bindung zwischen dem C5 des Pyrimidinrings und dem C1 der Ribose. In Anwesenheit von Cyanursäure bildet TARC nicht kovalente polymere Strukturen aus (**Abb. 8a**).[194] Analog zu TAP reagiert Barbitursäure (BA) mit Ribose-5-phosphat unter Ausbildung einer C-C glykosidischen Bindung. Auf Grund höherer Reaktivität des Heterozyklus läuft die Reaktion direkt in Lösung mit hohen Ausbeuten (>80%) ab. Melamin kann ebenfalls mit Ribose-5-phosphat in Lösung zu Nukleosiden reagieren (33-55%). Hierbei entsteht die glykosidische Bindung über ein exozyklisches Amin des Melamins. Auf Grund von komplementären Wasserstoff-Brückenbindungen können sich die beiden Nukleoside (BA und Melamin) spontan zu nicht kovalenten Polymeren selbst organisiert assemblieren (**Abb. 8b**).[195]

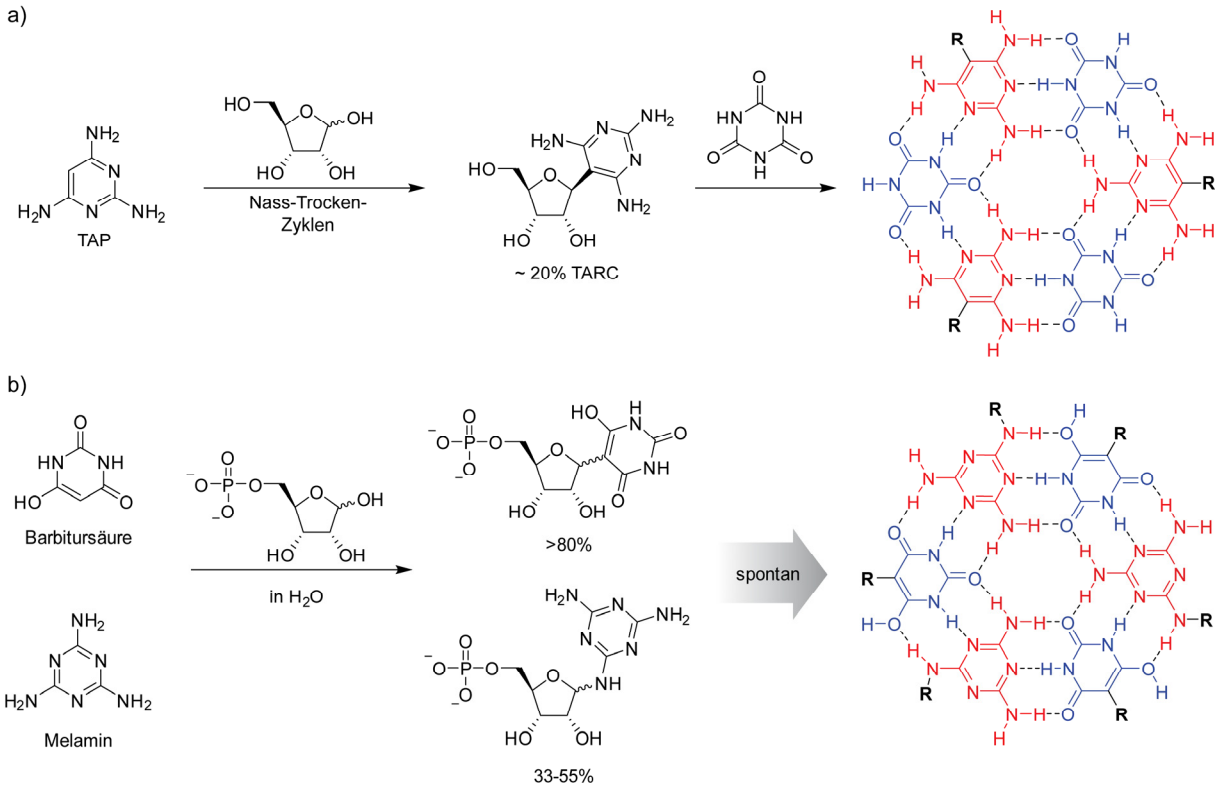


Abbildung 8. Entstehung von nicht-kanonischen Nukleosiden. a) Reaktion von 2,4,6-Triaminopyrimidin (TAP) mit Ribose zu 2,4,6-Triaminopyrimidin Nukleosid (TARC) sowie Selbstorganisation in Anwesenheit von Cyanursäure. b) Entstehung von Melamin und Barbitursäure Nukleotiden sowie spontane Selbstorganisation der beiden Einheiten.

Obwohl einige nicht-kanonische Pyrimidine mit Ribose teilweise sogar in Lösung reagieren (**Abb. 8b**), bleibt die Frage, wie die kanonischen Pyrimidin Nukleoside entstehen konnten. Da die kanonischen Pyrimidine nicht direkt mit Ribose kondensiert werden können, entwickelte die Gruppe um Sutherland eine andere Strategie, bei der der Zucker sowie die Nukleobase schrittweise aufgebaut werden (**Abb. 9**).[196] Diese Strategie bietet zwei Vorteile: Zum einen ist der Syntheseweg nicht von freier Ribose abhängig, zum anderen wird die glykosidische C-N Bindung vor dem Aufbau des Heterozyklus geknüpft. Besonders die Notwendigkeit von Ribose als Ausgangsverbindung in präbiotischen Nukleosidsynthesen wird von Sutherland immer wieder auf Grund dessen geringer Stabilität sowie Verfügbarkeit kritisiert.[34, 196-197]

Für die Pyrimidin Synthese gehen die Autoren von den beiden präbiotisch plausiblen Molekülen Cyanamid und Glykolaldehyd aus, welche in einem Phosphatpuffer (1M) zu 2-Amino-oxazol reagieren. Anschließend kann dieses mit Glycerinaldehyd zu Pentose-aminooxazolin (50%) reagieren (**Abb. 9a**). Entscheidend ist die zeitversetzte Zugabe der jeweiligen C2 und C3-Bausteine, da sonst die gewünschten Produkte nicht entstehen. Die Gruppe um Powner konnte dieses Problem der zeitversetzen Verfügbarkeit durch Anreicherung der jeweiligen C2- und C3-Aldehyde mit Hilfe von Aminalbildung lösen (**Abb. 9b**).[179] Trotzdem bleibt das Problem der Stereospezifität für die Entstehung von Pentose-aminooxazolin bestehen. Als Produkt entstehen alle vier möglichen Pentosen (**Abb. 9a**): Ribose (25%), Arabinose (15%), Xylose (6%) und Lyxose (4%) zusammen mit Eliminierungsprodukten (20%).[196] Dieser Syntheseschritt stellt damit den Flaschenhals des Syntheseweges dar, da nur das Arabinose-Derivat das gewünschte natürliche β-Isomer später bilden kann. Das Arabinose-aminooxazolin wird mit Cyanoacetylen (7 Äq.) zum Anhydro-arabinonukleosid umgesetzt[198] und

anschließend mit Phosphat durch Substitution der 2'-Position in das Cytidin-2',3'-cyclo-phosphat umgewandelt (**Abb. 9a**). Auf Grund seiner Elektrophilie sind 7 Äq. Cyanoacetylen notwendig, um den Pyrimidinring zu bilden. Das benötigte Cyanoacetylen entsteht durch eine Kreuzkupplung von Acetylen und HCN mit Hilfe von Cu(II) in Wasser.[197] Interessanterweise stellt Sutherland zwar die Stabilität und Verfügbarkeit der Ribose in Frage, vernachlässigt jedoch, dass Cyanaoacetylen in Wasser eine deutlich geringere Halbwertszeit im Vergleich zur Ribose besitzt. Die Verwendung von Cyanoacetylen im vorletzten Syntheseschritt könnte somit potentiell zu Verfügbarkeitsproblemen führen. Erstrebenswert wäre demnach eine Nukleosidsynthese, welche Cyanoacetylen als Startmaterial verwendet, welches dann in stabilere Moleküle mit höheren Halbwertszeiten überführt werden kann. Nichtsdestotrotz konnte Sutherland über seinen Syntheseweg als erstes kanonische Pyrimidin-, jedoch keine Purinnukleoside präbiotisch verfügbar machen.

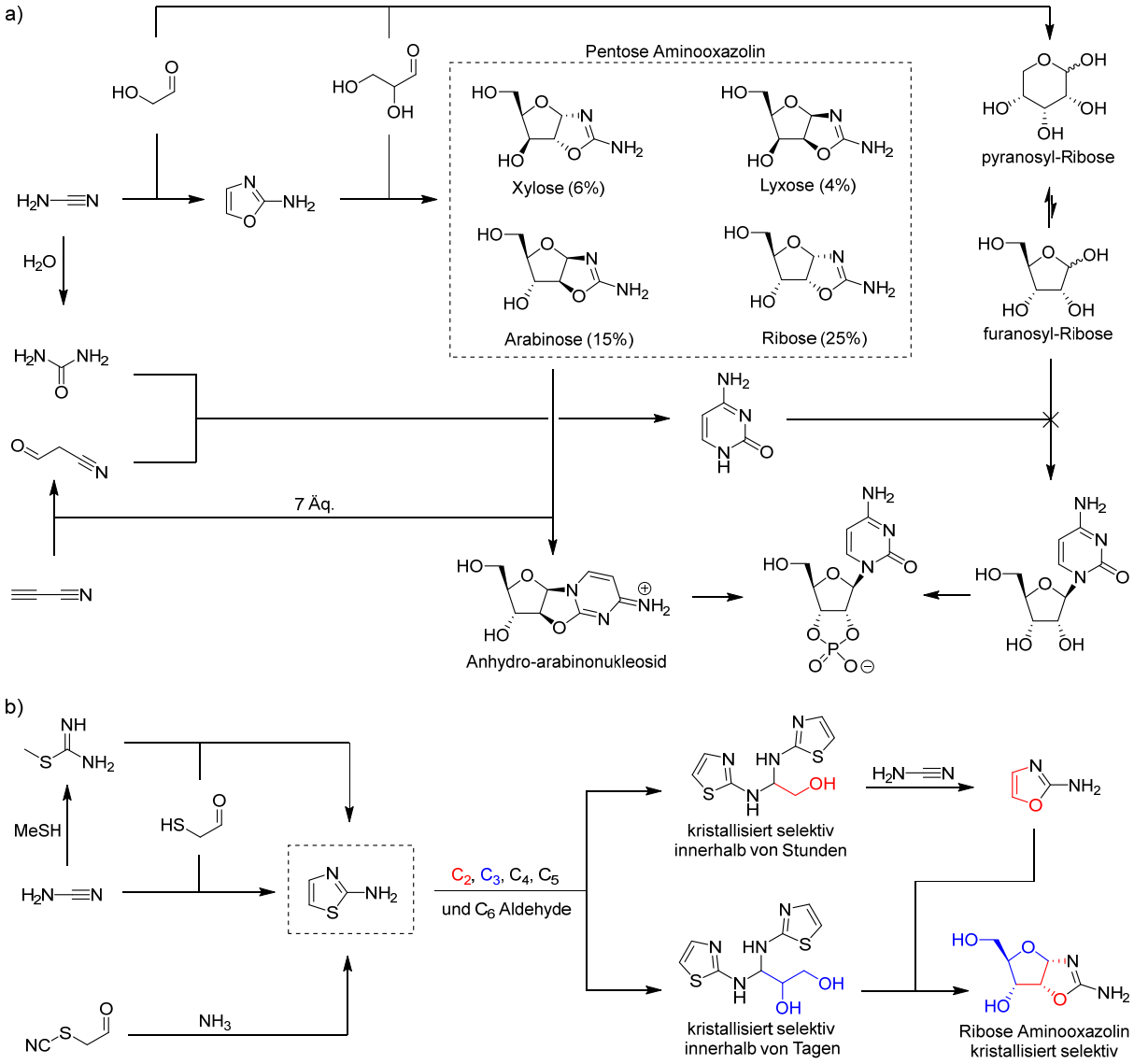


Abbildung 9. Synthese der Pyrimidin Nukleoside nach *Sutherland et al.* a) Syntheseweg für die Bildung der Pyrimidin Nukleotiden.[196] b) Zeitversetzte Verfügbarkeit von Glykolaldehyd und Glycerinaldehyd für die Synthese von Pentose Aminooxazolin.[179]

Durch Modifizierung seines Syntheseweges war Sutherland später in der Lage, auch das Ribose-aminooxazolin über ein α-2-Thiocytidin und anschließender Photo-Anomerisierung für die Synthese

natürlicher Pyrimidin Nukleoside nutzbar zu machen.[199] Die Anomerisierung des 2-Thioderivates (76%) funktioniert gut im Vergleich zu nicht modifizierten α-Cytidin, welches nur zu 5% anomerisiert (**Abb. 10**).[200] Wird Cytidin jedoch acetyliert, kann das α-Cytidin zu 22% in das β-Cytidin mit Hilfe von UV-Strahlung überführt werden.[201] Erst kürzlich entwickelte die Gruppe um Powner eine Erweiterung des Pyrimidinsynthesewegs, sodass jetzt zusätzlich auch die 8-oxo-Purin Derivate von Inosin und Adenosin dargestellt werden können.[202] Frühere Versuche desselben Autors brachten bisher keine Purinnukleoside hervor.[203]

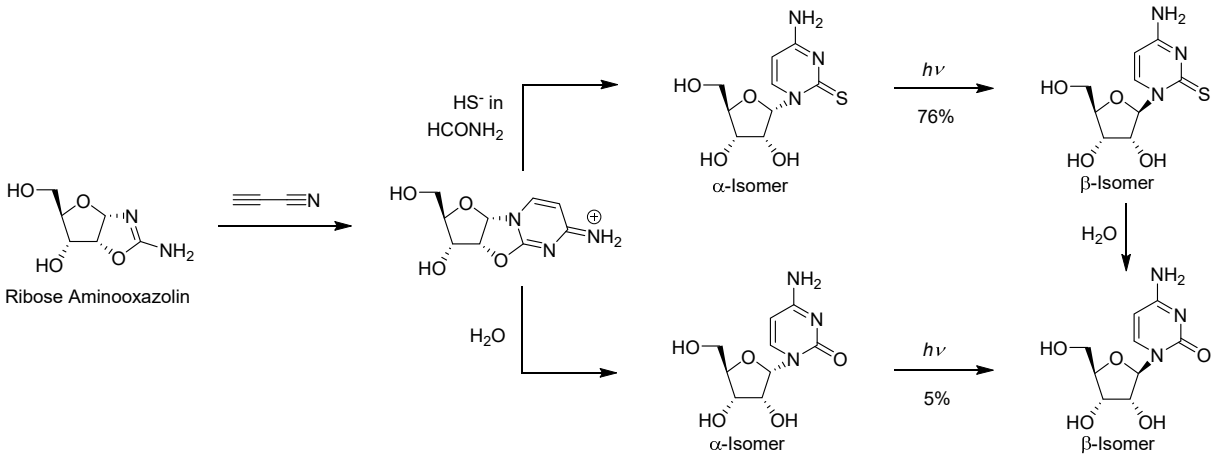


Abbildung 10. Photo-Anomerisierung für die Synthese von β-Cytidin aus α-Cytidin Derivaten.[199]

Beide kanonischen Purinnukleoside (Adenosin und Guanosin) konnten bisher nur von Orgel in sehr geringen Ausbeuten über Verknüpfung von Ribose und der entsprechenden Nukleobase präbiotisch zugänglich gemacht werden.[105-106] Andere Strategien, welche über eine aktivierte Ribose verlaufen, liefern zwar Adenosin, aber kein Guanosin.[190-191] Auch die Strategie für den schrittweisen Aufbau der Nukleoside bringt weder Guanosine noch dessen 8-oxo-Derivat hervor.[196, 202] Dies stellt die präbiotische Verfügbarkeit der kanonischen Purinnukleoside in Frage. Die Arbeitsgruppe Carell entwickelte daher eine neue Strategie, um die N-9-Purin Nukleoside regioselektiv darstellen zu können. Hierbei wurde nicht die Nukleobase, sondern eine Vorstufe mit Ribose reagiert, welche eine höhere Nukleophilie im Vergleich zum N-9 Atom der Purine aufweist. Diese so genannten FaPys können unter thermischen oder basischen Bedingungen nach erfolgreicher Glykosidierung in die Nukleobase überführt werden. Auf Grund von Symmetrieeigenschaften der FaPys entsteht dabei immer das natürliche N-9 Nukleosid mit bis zu 60% Ausbeute. Dieser Ansatz gibt damit präbiotischen Zugang sowohl zu Adenosin als auch zu Guanosin, welche über Phosphorylierung in die kanonischen Nukleotide überführt werden können.[204-206] Später konnten auch nicht-kanonische Nukleoside wie m^2A, DA, ms^2A, m^1G, m^2G oder $m^2{}_2G$ parallel mit den kanonischen Nukleosiden präbiotisch über das gleiche Prinzip synthetisiert werden. Interessanterweise sind all diese Modifikation in der heutigen RNA zu finden. Eine detaillierte Darstellung der präbiotischen Purinsynthesen ist in Abschnitt 5.1 und 5.2 zu finden.

Die Kondensationsreaktion zur Verknüpfung von Zucker und Nukleobase ist thermodynamisch in Wasser nicht begünstigt. Die Arbeitsgruppen um Saladino und DiMauro studierten deshalb die Bildung von Adenosin in Formamid in Abwesenheit von Wasser. Adenosin (5,6%) konnte dabei aus Ribose und

Adenin mit Hilfe von Protonenbestrahlung synthetisiert werden. Wurde Meteoritenpulver der Reaktion beigemengt, konnte die Ausbeute für Adenosin (20%) deutlich erhöht werden.[207] Schon vorher konnten die Autoren zeigen, dass Meteoritenpulver die Synthese aller vier kanonischen Nukleoside aus reinem Formamid in Spuren katalysieren kann.[138]

3.2 RNA-Modifikationen und ihre potentiellen Rollen auf der frühen Erde

Wie oben dargestellt, führen unterschiedliche präbiotische Reaktionen zu einer Vielzahl an kanonischen und nicht-kanonischen RNA-Bausteinen. Dies impliziert, dass auf der frühen Erde nicht nur die kanonischen Nukleoside über zielgerichtete Synthese vorhanden waren, sondern eine hohe Diversität an RNA-Bausteinen verfügbar war. Ein frühes genetisches Polymer könnte also sehr heterogen gewesen sein, vor allem bezüglich der Zucker oder in Bezug auf die Nukleobasen. Über Selektion muss daraus dann die RNA hervorgegangen sein (**Abb. 11**).[208-211] Heutige RNA besitzt nicht nur die kanonischen Basen (A, G, C und U), sondern mehr als 120 modifizierte Nukleoside, welche diverse fundamentale biologische Prozesse steuern.[212] Die meisten davon sind im Translationsapparat, vor allem in der Transfer-RNA (tRNA) und ribosomalen RNA (rRNA), zu finden. Dieser Translationsapparat ist im Prinzip hochkonserviert in allen Lebewesen vorhanden und muss sich deshalb schon sehr früh während der Entstehung des Lebens entwickelt haben.[213-215] Nicht-kanonische RNA-Bausteine könnten somit früh eine entscheidende Rolle gespielt und möglicherweise die Entstehung des Ribosoms und damit die Translation überhaupt möglich gemacht haben. Daher soll nachfolgend die Rolle von RNA-Modifikationen auf der frühen Erde diskutiert werden.

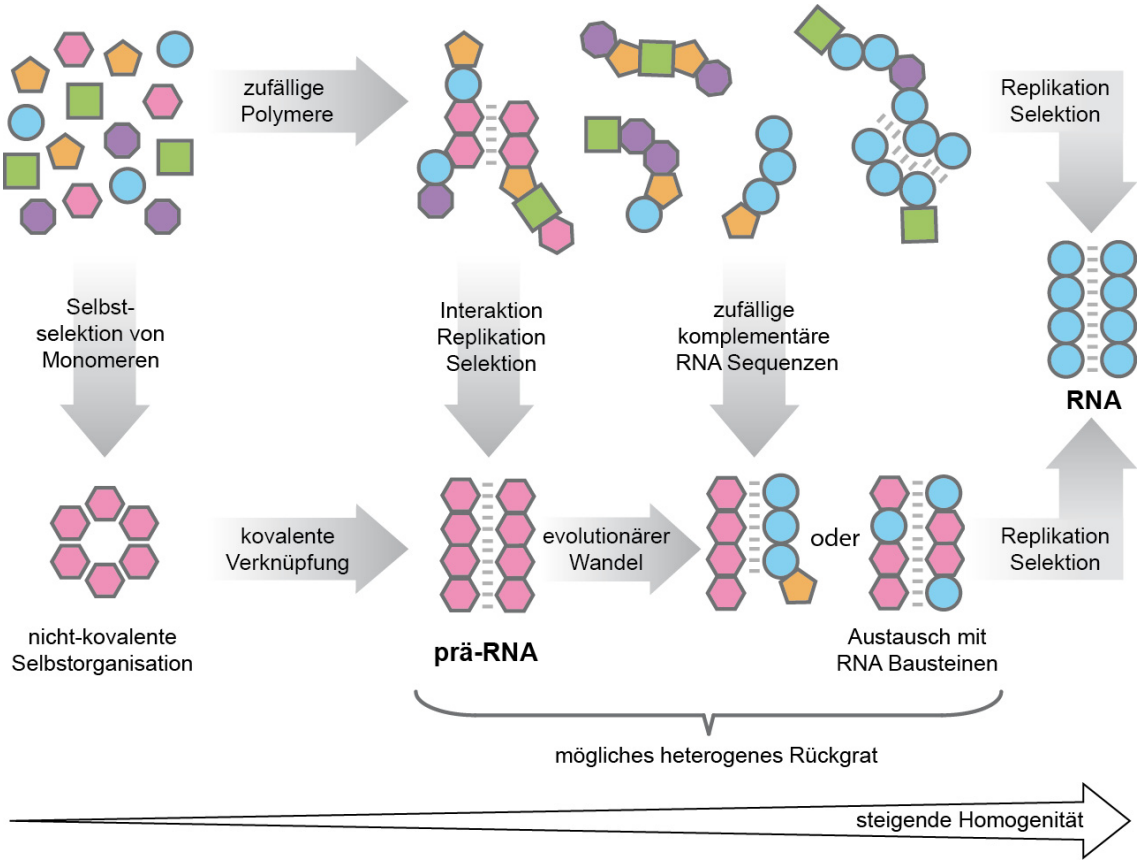


Abbildung 11. Schematische Darstellung für die mögliche Entstehung der RNA auf der frühen Erde.

3.2.1 Die Rolle von nicht-kanonischen Nukleotiden bei der Entstehung von RNA

Bevor RNA überhaupt entstehen konnte, mussten zuerst die RNA-Bausteine verfügbar sein. Um eine Polymerbildung zu begünstigen, wäre eine lokal hohe Konzentration der jeweiligen Monomere sowie deren Präorganisation von Vorteil. Auf Grund der Vielfalt an Nukleosiden/Nukleotiden auf der frühen Erde, stellt sich die Frage, wie die Monomere aus einem diversen Gemisch selektiert werden konnten, um daraus ein genetisches Polymer zu bilden. Eine Präorganisation sowie Selektion der kanonischen RNA-Monomere ist problematisch, da diese keine Basenpaare in Lösung bilden.[216] Im Gegensatz dazu sind die von Nicholas Hud beschriebenen Nukleotide mit Melamin und Barbitursäure als Nukleobase in der Lage, über Selbstassemblierung nicht-kovalente Polymere zu bilden.[195] Damit könnten sich diese nicht-kanonischen Nukleoside selbst aus einem komplexen Gemisch selektiert haben (**Abb. 11**). In diesem Zusammenhang ist es besonders interessant, dass das Barbitursäure-Nukleotid dem Pseudouridin, welches in den drei Domänen des Lebens zu finden ist, strukturell sehr ähnlich ist.
Durch die Präorganisation der Nukleotide könnte eine Verknüpfung dieser Monomere zu einem kovalenten Polymer stark begünstigt gewesen sein, welches als potentielles genetisches Material auf der frühen Erde gedient haben könnte. Interessanterweise ist der Einbau der β-Isomere in die supramolekulare Struktur mit einem Verhältnis von 2:1 bevorzugt.[195] Dies könnte ein frühes Indiz sein, weshalb unser heutiges genetisches Material β- statt α-Isomere verwendet. Über chemische Evolution mit Hilfe der vorhanden präbiotischen Bausteine könnte sich später die RNA mit ihren vier kanonischen Basen und dem Ribose-Phosphat-Rückgrat entwickelt haben (**Abb. 11**). Ob dieses Rückgrat direkt die Furanose-Zuckerkonstitution besaß, ist ungewiss. Möglich wäre, dass zuerst eine prä-RNA mit einer Pyranose- statt einer Furanose-Konfiguration existierte oder sogar ein einfacherer Zucker wie Threose verwendet wurde.[217-218] Die Threose- oder Ribopyranosyl-Monomere wären auf der frühen Erde ebenfalls verfügbar gewesen.[107, 191] Auch andere Pentosen oder Hexosen könnten in Betracht gezogen werden.[219] Wieso Ribose durch Evolution selektiert wurde, bleibt weiterhin ungewiss.
Ebenfalls ist unklar, ob RNA tatsächlich aus einem „einfacheren“ genetischen System hervorging oder ob sie direkt aus ihren präbiotisch verfügbaren Komponenten aufgebaut wurde (**Abb. 11**). Unabhängig davon hat ein genetisches Polymer während der chemischen und biologischen Evolution vermutlich eine Vielzahl von Selektionsprozessen durchlaufen, wonach dann unsere heutige RNA mit den vier kanonischen Basen entstanden ist.[208-209, 220]

3.2.2 Einfluss von Modifikationen auf Replikation oder katalytische Aktivität

Neben den kanonischen RNA Bausteinen waren wahrscheinlich auch Hydrolyseprodukte sowie einfache Modifikationen wie Methylierungen oder Thiolyierungen, aber auch Aminosäurederivate auf der frühen Erde vorhanden.[140, 199, 221] Viele dieser RNA-Derivate sind trotz ihrer Modifizierung immer noch in der Lage, eine Basenpaarung mit einer kanonischen Base einzugehen (**Abb. 12a**).[222-228] Dies impliziert, dass die Fähigkeit für Templat-basierte Replikation grundsätzlich auch für einige modifizierte Stränge vorhanden war. Generell kann man sich vorstellen, dass bei der Replikation (nicht-enzymatisch oder enzymatisch) neben den kanonischen Basen je nach präbiotischer Verfügbarkeit auch

komplementäre modifizierte Basen in den Strang eingebaut werden konnten.[229-230] Die ersten RNA-Polymere müssen dabei nicht zwingend selbstreplizierend gewesen sein, sondern könnten über nicht-enzymatische Replikation vervielfältigt worden sein, um so die chemische Evolution voranzutreiben (**Abb. 12b**).[231-232] Hierbei konnte die Gruppe um Szostak zeigen, dass die Verwendung von 2-Thio-uridin (s^2U) im Vergleich zu Uridin eine höhere Genauigkeit sowie Reaktionsgeschwindigkeit für das Kopieren von RNA-Templaten durch nicht-enzymatische Replikation besitzt. Dies ist vermutlich auf die höhere thermodynamische Stabilität des s^2U:A Basenpaares oder besseres *stacking* in der Helix zurückzuführen.[229] Eine analoge Rolle könnte auch der DA RNA-Baustein übernommen haben. DA ist in der Lage A im Erbgut komplett zu ersetzen.[233] Durch die Ausbildung von drei statt zwei Wasserstoffbrücken mit U, ist das DA im Vergleich zu A in der Lage, U durch Templat-basierte RNA-Synthese deutlich effizienter in den Gegenstrang einzubauen.[230]
In einer frühen RNA-Welt, bei der Replikation nicht-enzymatisch bzw. durch primitive RNA-Polymerasen erfolgte, konnten falsch eingebaute Basen wegen nicht vorhandener Fehlerkorrektur-Mechanismen wie *proof-reading* nicht erkannt werden. Dies führte zwangsläufig zu deutlich höherer Promiskuität bezüglich des Einbaus von nicht-kanonischen RNA-Monomeren, solange sie auf der frühen Erde verfügbar waren. Zu hohe Fehleranfälligkeit kann jedoch zum Verlust der genetischen Information führen. Dadurch entstand zwangsläufig ein Selektionsdruck für eine höhere Replikationsgenauigkeit durch effizientere Basenpaarung, um die genetische Information zu erhalten. Modifikationen wie s^2U oder DA könnten somit durchaus eine wichtige Rolle in einer frühen RNA-Welt gespielt haben. Durch die evolutive Entstehung von komplexen DNA- und RNA-Polymerasen, welche *proof-reading* Eigenschaften besitzen, ist der Selektionsdruck für effizientere Basenpaarung deutlich schwächer geworden. Daher ergibt es aus evolutionärer Sicht Sinn, dass z.B. DA im Genom nicht dauerhaft konserviert wurde, da die Identität der DA:U Paarung nach einer ungewollten Hydrolyse von DA zu G verloren gehen würde.

Dennoch sind in einigen Spezies kanonische Basen komplett durch nicht-kanonische Basen ersetzt worden. Neben dem schon beschriebenen DA wurden auch 5-Methylcytosin, 5-Hydroxymethylcytosin sowie 5-Hydroxymethyluracil als Basen identifiziert.[234-235] Weiterhin konnte die Gruppe um Benner Ende der 1980er zeigen, dass neben dem Austausch einer Base auch eine komplett artifizielle Basenpaarung zwischen Isoguanin und Isocytosin möglich ist. Eindrucksvoll konnte gezeigt werden, dass RNA- und DNA-Polymerasen selektiv auch die nicht-kanonischen Basen in Gegenwart der kanonischen Basen einbauen und replizieren können.[236-237] Dies impliziert, dass der Einbau von nicht-kanonischen Basen durch eine frühzeitliche RNA-Polymerase durchaus möglich gewesen sein könnte.

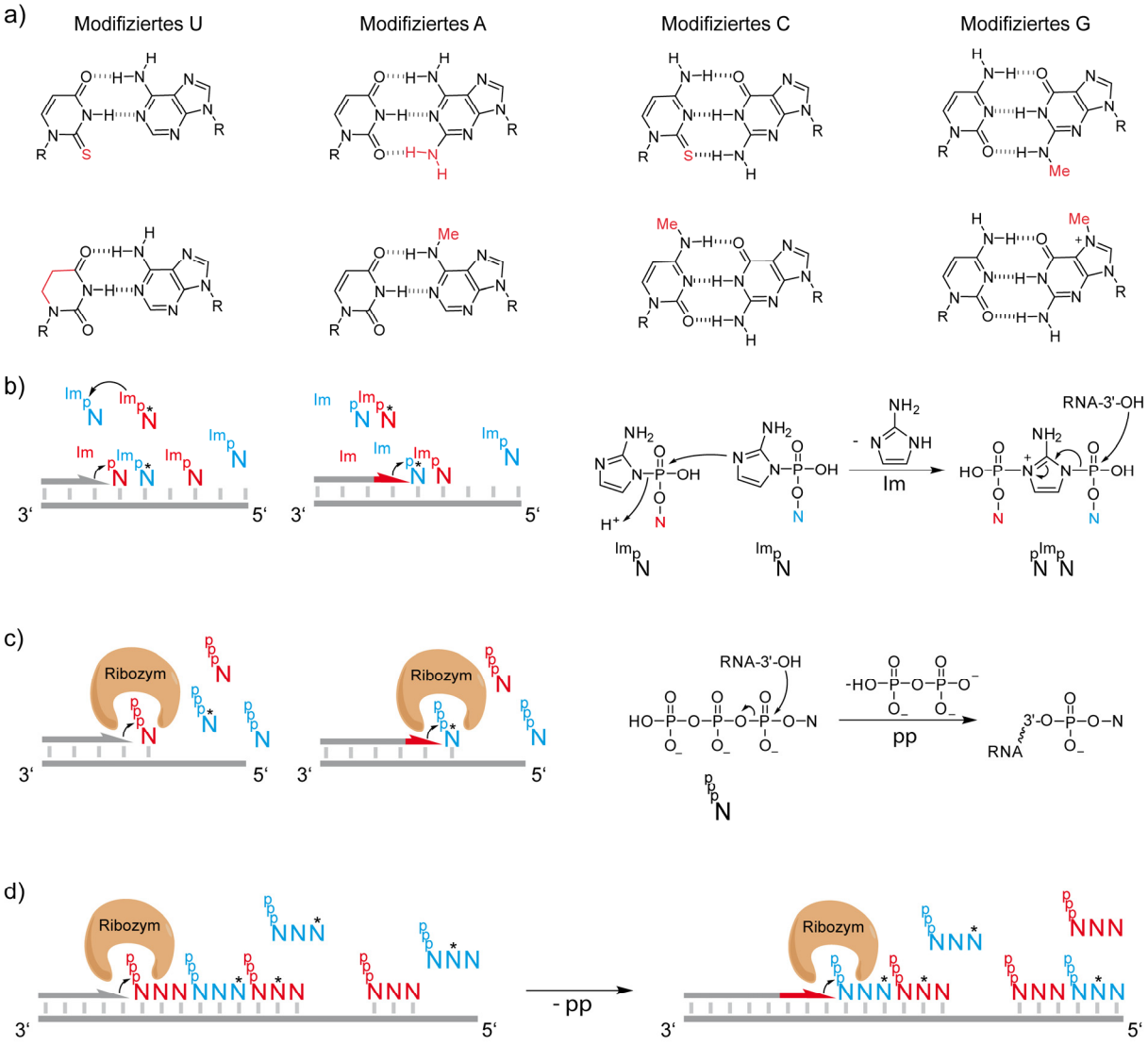


Abbildung 12. Basenpaarungen von nicht-kanonischen Nukleosiden sowie mögliche Szenarien für deren Templat-basierten Einbau. Mit (*) markierte Nukleotide (N) stellen modifizierte RNA-Basen dar. a) Biologisch gefundene Basenpaarungen von modifizierten Nukleobasen. b) Nicht-enzymatischer Einbau von aktivierten Nukleotiden. Der chemische Mechanismus des Einbaus ist rechts dargestellt. c) Enzymatischer Einbau von Nukleosidtriphosphaten. Der chemische Mechanismus des Einbaus ist rechts dargestellt. d) Enzymatischer Einbau von Trinukleotidtriphosphaten (Tripletts) der Mechanismus des Einbaus ist Äquivalent zu c).

Der enzymatische Einbau von nicht-kanonischen Basen durch RNA-Polymerase Ribozyme könnte auf zwei verschiedene Arten erfolgt sein. Die erste Möglichkeit wäre der Einbau durch Nukleosid-triphosphate (**Abb. 12c**).[134, 238-240] Hierbei muss ein modifizierter RNA-Baustein eine starke Basenpaarung mit der jeweiligen Base im Templatstrang besitzen, um so einen Einbau zu ermöglichen. Zu schwache Interaktionen hätten vermutlich nicht ausgereicht, um kompetitiv mit den kanonischen Nukleotiden mitzuhalten. Der Einbau von Modifikationen, welche nur sehr schwache Basenpaare bilden, hätte jedoch über eine andere Strategie funktionieren können. Wie die Gruppe um Holliger zeigen konnte, können statt Nukleotiden auch Trinukleotidtriphosphate (Tripletts) für die RNA-Replikation verwendet werden (**Abb. 12d**).[137] Durch Verwendung von Tripletts entsteht eine deutlich stärkere Interaktion mit dem Templatstrang im Vergleich zu einzelnen Nukleotiden. Somit hätten möglicherweise auch Tripletts mit RNA-Modifikationen eingebaut werden können. Die spezifische Interaktion von Tripletts, welche RNA-Modifizierungen tragen, kann bei der Condon-Anticodon-Wechselwirkung

während der Proteinbiosynthese beobachtet werden. Dies impliziert, dass spezifische Triplett-Interaktionen nicht nur auf rein kanonische Basen beschränkt ist.
Auch wenn der Einbau von nicht-kanonischen RNA Bausteinen höchstwahrscheinlich möglich war, bleibt die Frage, ob solche RNA-Stränge auch in der Lage waren, katalytisch aktiv zu sein. Durch *in vitro* Evolution konnte ein Ligase Ribozym erzeugt werden, welches lediglich DA und U als Nukleobasen verwendet.[241] Neben dem DA als modifizierte Adeninbase konnte auch für modifiziertes U die Entstehung eines Ribozyms gezeigt werden. Dabei konnten neue C-C Bindungen durch eine katalysierte Diels-Alder Reaktion entstehen.[242] Diese *in vitro* Experimente zeigen deutlich, dass katalytische Aktivität nicht auf rein kanonische Nukleoside beschränkt ist. Leider sind SELEX Experimente noch nicht in der Lage, auch modifizierte RNA-Bausteine zusätzlich zu den kanonischen Nukleotiden systematisch nutzbar zu machen.

3.2.3 Evolutionärer Vorteil durch RNA-Modifikationen

Schon mit den vier kanonischen Nukleosiden können 4^n (n = Anzahl der Monomere im Strang) Sequenzkombinationen entstehen. Leider wissen wir nicht, welche dieser RNA-Sequenzen nützliche Funktionen hervorbringen. Trotz vieler Erfolge haben *in vitro* Evolutionsexperimente gezeigt, dass die katalytische Fähigkeit der RNA auf Grund der limitierten chemischen Vielfalt im Vergleich zu Proteinen begrenzt ist.[136, 243-247] Der Einbau von nicht-kanonischen Basen könnte somit die Wahrscheinlichkeit für eine nützliche Funktion der RNA durch höhere chemische Variabilität deutlich erhöht haben.[248-251] Selbst wenn Modifikationen nicht aktiv über primitive RNA-Polymerasen in die RNA eingebaut wurden, könnten diese dennoch spontan z.B. durch Hydrolyse (A zu Inosin) in der RNA entstanden sein. Auf Grund fehlender Korrekturmechanismen blieben eingebaute wie auch spontan entstandene Modifikation in der RNA erhalten. Da viele RNA-Modifikationen immer noch in der Lage sind, eine Basenpaarung einzugehen, könnte die Fähigkeit für Templat-basierte Replikationen weiterhin für einige RNA-Stränge bestanden haben. Der Einbau von komplementären modifizierten Basen (z.B. s^2U oder DA)[229-230] führt dazu, dass zwar die Sequenz, aber nicht das Modifikationsmuster erhalten geblieben wäre. Somit wären bei jeder Replikation neue Modifikationsmuster entstanden, von denen einige möglicherweise neue Funktionen hervorbringen konnten, obwohl ihre Sequenz mehr oder weniger unverändert blieb. Im Prinzip könnte sich so die Entstehung des Lebens auf eine selbstreplizierende RNA-Sequenz reduzieren lassen, welche dann verschiedene funktionale RNAs auf Grund unterschiedlicher Modifikationsmuster hervorbringen konnte, ohne die zugrundeliegende Sequenz zu verändern (**Abb. 13**). Bei den primitiven RNA-Polymerasen muss zwischen zwei Typen unterschieden werden – den Selbstreplikatoren und den RNA-Polymerasen, die andere RNA-Stränge als Templat für ihre Synthese nutzen. Selbstreplikatoren replizieren nur die eigene Sequenz, wodurch diese schnell akkumuliert. Andere RNA-Polymerasen hingegen nutzten andere RNA-Stränge als Templates und amplifizierten diese. Beide Typen von primitiven RNA-Polymerasen konnten höchstwahrscheinlich komplementäre nicht-kanonische RNA-Basen erkennen und einbauen (s. Abschnitt 3.2.2), sodass sich zwar die grundsätzliche Sequenz nicht änderte, aber unter Umständen die Funktionalität auf Grund neuer Modifikationsmuster (**Abb. 13.**). Vorteilhafte Modifikationen konnten sich dann in einem

Selektionsprozess durchsetzen, wobei auch nicht-funktionale RNA-Stränge weiterhin als Template zur Verfügung gestanden hätten. Aus Ihnen hätten neue funktionale RNA-Stränge durch Replikation auf Grund von neuen Modifikationsmustern entstehen können. Zu Sequenzänderungen konnte es hingegen in beiden Fällen durch fehlerhaften Einbau eines nicht-komplementären Nukleosids kommen. Aus *in vitro* Evolutionsexperimenten konnten bislang verschiedene RNA-Replikationssysteme hervorgehen, es konnte aber bislang kein vollwertiger Selbstreplikator gezeigt werden.[134-135, 137]

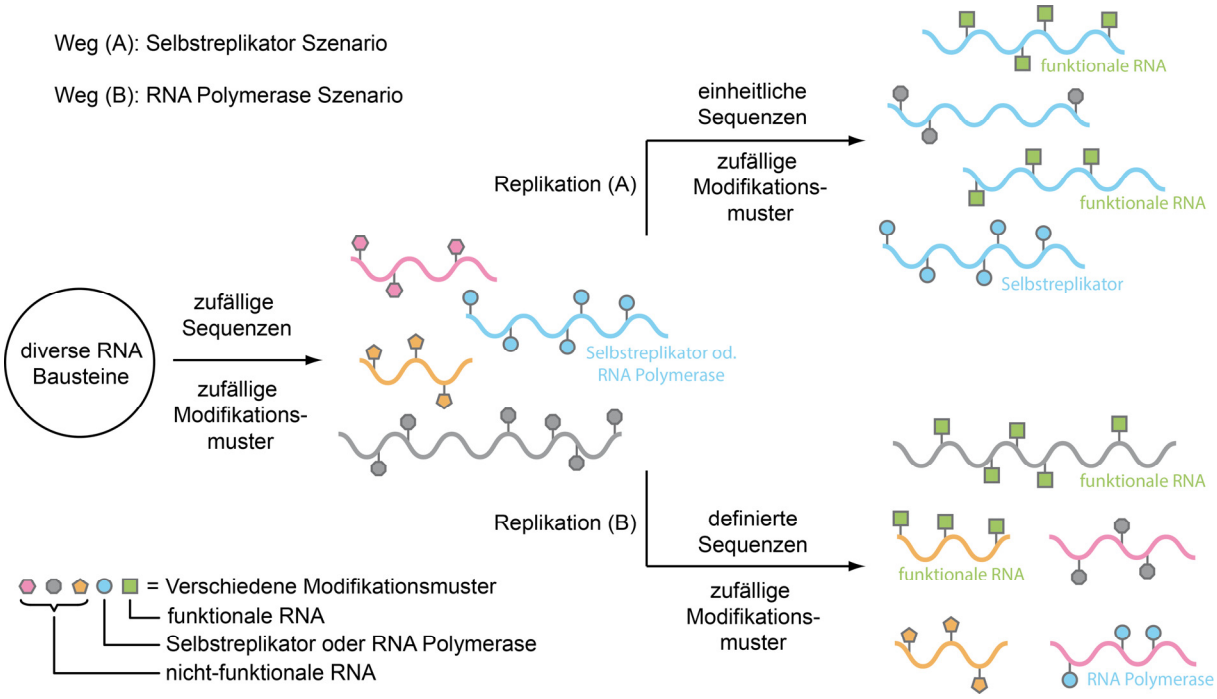


Abbildung 13. Effekt von RNA Modifikationen auf die Funktion der RNA. Aus den präbiotisch verfügbaren RNA Bausteinen entstanden zunächst zufällige Sequenzen (pinke, blaue, orange und graue Wellenlinien) mit zufälligen Modifikationsmustern durch den Einbau von nicht-kanonischen Basen (geometrische Formen in gleicher Farbe). Einige dieser Sequenzen hatten zufällige RNA-Polymerase-Eigenschaften (blaue Sequenz). Es wird zwischen Selbstreplikator (Weg A) und RNA-Polymerase (Weg B) unterschieden. Bei jeder Replikation entstanden zufällige neue Modifikationsmuster. Einige Modifikationsmuster führten ebenfalls zu funktionalen RNAs (grüne Quadrate oder blaue Kreise) während andere nicht-funktional waren (grau, orange oder pinke geometrische Formen). Per Zufall konnte auch das Modifikationsmuster reproduziert werden (gleichfarbige Sequenz und Modifikationsmuster) Weg A: Der Selbstreplikator vervielfältigt nur seine eigene Sequenz (blau), wobei zufällige neue Modifikationsmuster entstehen. Weg B: RNA-Polymerase vervielfältigt alle Sequenzen bis auf die eigene, wobei zufällige neue Modifikationsmuster entstehen.

Durch Replikation wäre die Sequenz auf Grund komplementärer Modifikationen (**Abb. 12a**) erhalten geblieben, solange man von spontanen Mutationen absieht. Dies ist wichtig, um so vorteilhafte Modifikationsmuster per Zufall zu reproduzieren. Die Wahrscheinlichkeit für eine zufällige Reproduzierbarkeit nimmt jedoch mit der Anzahl an zusätzlichen Modifikationen ab, da die möglichen Sequenzkombinationen exponentiell mit $(4+m)^n$ ansteigen (m = Anzahl zusätzlicher Modifikationen und n = Anzahl der Monomere im Strang). Dies berücksichtigt jedoch nicht, dass verschiedene Modifikationsmuster auf Grund ihrer höheren chemischen Variabilität jeweils die gleiche Funktion hervorbringen könnten, analog zu konvergenten Proteinen. Eine Modifizierung kann damit auf der einen Seite zwar die Wahrscheinlichkeit für eine funktionelle RNA erhöht haben, auf der anderen Seite aber die Wahrscheinlichkeit für eine Reproduzierbarkeit gesenkt haben. Es könnte also zu einem Trade-off zwischen höherer chemischer Vielfalt und der Anzahl an verschiedenen Modifikationen gekommen sein. Ein optimales Set an RNA-Modifikationen müsste demnach zwischen null und den in der heutigen

Biologie gefunden Modifikationen liegen. Diese Betrachtung kann jedoch weiter eingeschränkt werden. Phylogenetische Analysen legen nahe, dass LUCA ein sehr limitiertes Set an RNA-Modifikationen besaß.[1, 252] Obwohl LUCA bereits DNA, RNA und Proteine nutzte, fällt auf, dass nur die RNA-Modifikationen aber nicht die Proteinmodifikationen stark konserviert sind.[253] Dies könnte dafür sprechen, dass ein primitiver Organismus RNA-Modifikationen nutzte, noch bevor kodierte Proteine überhaupt vorhanden waren. Der Grund dafür könnte die Evolution des Translationsapparats gewesen sein, welcher die RNA-katalysierte Proteinsynthese ermöglicht hat. Hierbei fällt auf, dass die funktionalen tRNAs und rRNAs als Teil des Translationsapparats stark modifiziert sind. Da die Modifikationen eine Vielzahl an essentiellen Prozessen steuern, stellt sich die Frage, ob Translation mit rein kanonischen Basen überhaupt möglich wäre. Vielleicht ist dies auch der Grund, dass jedes bekannte Leben von modifizierten RNA-Bausteinen abhängig ist.

Die Evolution des Translationsapparats kann als Meilenstein für die Entstehung des Lebens betrachtet werden. Sie ermöglichte den Übergang von einem RNA-Organismus, welcher sich in Abwesenheit von kodierten Proteinen vermehren und einen primitiven Metabolismus aufrechterhalten musste, zu komplexeren Lebewesen wie LUCA.[214-215] Bei diesem Übergang wurden vermutlich auch die essentiellen RNA-Modifikationen übernommen, welche in den verschiedenen funktionellen RNA-Polymeren vorhanden waren.[213] Ein frühzeitlicher RNA-basierter Organismus war vermutlich noch nicht in der Lage, alle seine Bausteine selbst zu synthetisieren. Solange die Bausteine jedoch präbiotisch verfügbar waren, konnten diese aus der Umwelt aufgenommen werden und wären damit per Definition Vitamine gewesen. Ein erfolgreiches selbstreplizierendes System hätte somit exponentiell wachsen können, solange die benötigten Ressourcen in genügender Menge vorhanden waren. Auf Grund des exponentiellen Wachstums wären die Bausteine aus der Umwelt irgendwann aufgebraucht gewesen, da die präbiotische Synthese nicht mehr in der Lage war die Bausteine genügend schnell nachzuliefern. Daraus wäre ein starker Selektionsdruck für die Synthese der eigenen Bausteine entstanden. Hierbei hätte ein Organismus, welcher Peptide über ein primitives Translationssystem synthetisieren konnte, einen evolutionären Vorteil gegenüber einem reinen RNA-Metabolismus gehabt. Dieser Vorteil lässt sich durch die größere chemische Vielfalt von Proteinen und der damit verbundenen höheren Effizienz der Protein- vs. Ribozym-Katalyse erklären.[254-255] Die höhere Effizienz auf Grund größerer chemischer Vielfalt spricht ebenfalls dafür, dass Modifikationen in einem RNA-Metabolismus einen evolutionären Vorteil gebracht haben könnten. Dies kommt besonders stark bei limitierten Ressourcen zum Tragen, da bei unbegrenzter Verfügbarkeit die Effizienz des Metabolismus keine Wichtigkeit besitzt oder nicht zwingend vorhanden sein muss.

3.2.4 RNA Modifikationen und der Translationsapparat

Ein frühes Translationssystem war höchstwahrscheinlich deutlich simpler im Vergleich zur heutigen, über Milliarden von Jahren entwickelten, ribosomalen Translation. Da ein frühzeitliches Translationssystem zuerst ohne kodierte Proteine auskommen musste, war dieses bei weitem nicht so präzise, wie es biologische Systeme heute sind.[214, 256-257] Wie ein solches primitives System ausgesehen haben könnte und wie sich daraus die heutigen Ribosomen sowie tRNAs gebildet haben, soll hier nicht weiter vertieft werden. Hierbei soll auf die bekannte Literatur verwiesen werden.[214-215, 257-

[258] Vielmehr soll hier auf einige essentielle Prozesse der Translation eingegangen werden, welche durch RNA-Modifikationen beeinflusst werden. Diese Prozesse könnten so grundlegend gewesen sein, dass die Modifikationen unverzichtbar und damit dauerhaft konserviert wurden.

Um bei der Translation eine Basensequenz zuverlässig in eine Aminosäuresequenz zu überführen, ist es von fundamentaler Wichtigkeit, dass es nicht zu einer Verschiebung des Leserahmens kommt. Während ein falscher Einbau einer Aminosäure meist keinen großen Einfluss auf das Protein hat, ist eine Leserahmenverschiebung absolut schädlich für die Proteinsynthese. Um eine Leserahmenverschiebung zu verhindern, nutzt die Natur RNA-Modifikationen an den Positionen 37 und 34 der tRNA, welche die tRNA-mRNA-Interaktion während der Dekodierung stabilisieren und daher fundamental wichtig für die Proteinsynthese sind. An Position 37 ist die RNA-Modifikation m^1G in allen drei Domänen des Lebens hochkonserviert,[259] wobei auch teilweise t^6A als Modifikation an dieser Stelle vorkommt.[260] Wegen der hohen Konservierung ist anzunehmen, dass m^1G schon sehr früh Teil des Translationsapparates war.[259] Die Position 34 der tRNA ist Teil der Codon-Anticodon-Paarung und wird auch als Wobble-Position bezeichnet. Hier findet sich eine Vielzahl an Nukleosidmodifikationen wie z.B. Methylierung der Ribose, Acetylierung, Pseudouridylierung oder das Einfügen von Thiol-Derivaten.[261-262] Dies könnte darauf hinweisen, dass die Entstehung des genetischen Codes in direktem Zusammenhang mit den vorhandenen Modifikationen steht.

Neben der tRNA ist auch die rRNA teilweise stark modifiziert. Einige Hypothesen gehen davon aus, dass die tRNA sowie ein proto-Ribosom ursprünglich andere Aufgaben besaßen und die Proteinsynthese erst sekundär adaptiert wurde. Als mögliche Funktionen des Ribosoms wurden Splicing, Aminoacyl-tRNA-Synthetase oder RNA-Replikase vorgeschlagen.[263-266] Durch den Einbau von RNA-Modifikationen änderte sich dann im Laufe der chemischen Evolution die Funktion des Ribosoms hin zur Proteinsynthese. Es ist bekannt, dass viele Modifikationen tatsächlich die Faltung der rRNA sowie RNA-RNA Interaktionen beeinflussen.[267-270] Es ist also durchaus vorstellbar, dass ein verändertes Modifikationsmuster zu unterschiedlichen 3-dimensionalen Strukturen führte und damit auch zu anderen Funktionen (s. Abschnitt 3.2.3). Andere Hypothesen gehen davon aus, dass das Ribosom ursprünglich mehrere Funktionen besaß, von denen dann einige durch RNA Modifikationen inaktiviert wurden, dabei die Affinität des Substrats (tRNA) aber erhalten blieb.[251]

Abschließend bleibt zu sagen, dass RNA Modifikationen eine essentielle Rolle für viele wichtige biologische Prozesse spielen.[267-270] Relevant für die RNA-Welt sind vor allem Modifikationen, welche RNA-RNA Interaktionen oder die Faltung von RNA in seine tertiäre Struktur beeinflussen. In jedem Fall sind RNA-Modifikationen in der Lage, die chemische Vielfalt zu erhöhen, um eventuell eine effizientere Katalyse in einem RNA-Metabolismus zu gewährleisten. Es ist jedoch schwer zu sagen, ab wann genau die RNA-Modifikationen eine wichtige Rolle spielten. Wie präbiotische Experimente jedoch zeigen konnten, waren modifizierte RNA-Bausteine mit großer Sicherheit auf der frühen Erde vorhanden.[140, 199, 221, 250] Es ist anzunehmen, dass frühe Replikationssysteme weitaus ungenauer arbeiteten und somit teilweise auch modifizierte RNA-Bausteine eingebaut wurden. Auf Grund der hohen Konservierung von RNA-Modifikationen im Vergleich zu Proteinmodifikationen, scheint es plausibel, dass RNA-

Modifikationen schon eine Rolle gespielt haben mussten, noch bevor kodierte Proteine existierten. Dafür verantwortlich könnte die Evolution das Translationssystems sein, welches aus heutiger Sicht ohne die Verwendung von modifizierten Basen schwer vorstellbar erscheint. Dies könnte auch der Grund sein, wieso jedes bekannte Leben von den fundamentalen RNA-Modifikationen abhängig ist.

4 Zielsetzung

Die RNA-Welt-Hypothese legt nahe, dass das Leben zuerst aus replizierender RNA hervorging. Aus diesem genetischen Polymer entstanden integrierte chemische Systeme über einen chemischen Evolutionsprozess, gefolgt von der Entstehung des Lebens. Die Replikation von RNA erfordert die Bildung von komplementären Pyrimidin-Purin-Watson-Crick-Basenpaaren, die eine Voraussetzung für einen präzisen genetischen Informationstransfer sind. Zwar sind einige präbiotische Syntheserouten bekannt, welche RNA Bausteine liefern, jedoch existiert kein Reaktionspfad, welcher alle vier kanonischen Bestandteile der RNA in einer wässrigen Umgebung erschließen kann.[107, 190-191, 196] Dies ist problematisch, da die spontane Entstehung der Watson-Crick Basenpaare A:U und G:C auf der frühen Erde ohne die gemeinsame Bildung nicht gegeben ist. Interessanterweise besteht RNA jedoch nicht nur aus den kanonischen Bausteinen, sondern auch aus über 120 Modifikationen, welche eine Reihe an fundamentalen biologischen Prozessen steuern. Modifikationen sind ein essentieller Bestandteil des Lebens. Es ist bisher jedoch immer noch unklar, ab wann RNA-Modifikationen aus evolutionärer Sicht eine Rolle spielten. Zum einen könnten sie Teil der präbiotischen RNA-Welt gewesen sein. Damit wären sie Teil früher genetischer Polymere gewesen. Oder sie sind erst später als Teil der Biologie entstanden.

Das Ziel dieser Arbeit war demnach die Entwicklung von präbiotisch plausiblen Syntheserouten, welche die kanonischen Purin- bzw. Pyrimidin-RNA Bausteine liefern können. Außerdem sollte evaluiert werden, inwiefern es für modifizierte (nicht-kanonische) Nukleoside eine Chance für eine spontane Entstehung auf der frühen Erde gab. Die Resultate sollten weitere Einblicke in die mögliche Entstehung von kanonischen und nicht-kanonischen RNA-Bausteinen als Voraussetzung für die präbiotische Synthese von RNA schaffen. Dabei wäre es von Vorteil direkt aktivierte Nukleotide zu erhalten, welche potentiell zu RNA-Oligomeren reagieren könnten.

5 Resultate: Publizierte Arbeiten

5.1 A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway

Sidney Becker[#], Ines Thoma[#], Amrei Deutsch, Tim Gehrke, Peter Mayer, Hendrik Zipse, Thomas Carell

[#] geteilte Erstautorenschaft

Prolog

Der Ursprung des Lebens begann mit einfachsten präbiotischen Molekülen, welche entlang nicht identifizierter Reaktionspfade reagierten, um so die Schlüsselmoleküle des Lebens zu produzieren. Laut der RNA-Welt-Hypothese war RNA das zentrale Molekül während der chemischen Evolution. Wie aber konnten die Bausteine der RNA auf der frühen Erde entstehen? Bisher wurde ein einziger präbiotischer Weg zu den Purinnukleosiden vorgeschlagen. Er gilt auf Grund fehlender Regioselektivität und geringer Ausbeuten als ineffizient,. Zwar wurden in der Zwischenzeit auch andere Reaktionen für die Synthese von Purinnukleosiden durch die Gruppen um Benner und Zare gezeigt, jedoch liefern beide kein Guanosin. In dieser Arbeit wurde eine neue präbiotisch plausible Synthese für die Entstehung von Purinnukleosiden erarbeitet. Dabei wurden Formamidopyrimidine (FaPys) mit Ribose kondensiert, welche die natürlichen *N*-9-Nukleoside mit hoher Regioselektivität und guten Ausbeuten (60%) darstellen können. Die FaPys können durch Formylierung von Aminopyrimidinen entstehen, welche wiederum aus einfachen präbiotischen Molekülen wie Aminomalononitril als HCN Trimer und Guanidin entstehen können. Damit konnte ein plausibles Szenario für die Entstehung von Purinnukleosiden auf einer frühzeitlichen Erde geliefert werden.

Autorenbeitrag

Entwicklung und Optimierung der präbiotischen Synthesen. Entwicklung neuer Syntheserouten für α-pyranosyl Nukleoside als synthetische Standards. LC-MS Analysen sowie Interpretation von Daten.

Lizenz

$N52_{CDR\ H2}$, and $Y97_{CDR\ H3}$; and the pocket for V518 is defined by $G31_{CDR\ H1}$, $N52_{CDR\ H2}$, and $Y97_{CDR\ H3}$.
21. J. Huang *et al.*, *Nature* **515**, 138–142 (2014).
22. Sequences of 3943 HIV-1 strains were downloaded from the HIV database (www.hiv.lanl.gov/content/index).
23. P. D. Kwong *et al.*, *Nature* **420**, 678–682 (2002).
24. J. B. Munro *et al.*, *Science* **346**, 759–763 (2014).
25. The following nomenclature is used: "Tolerated" Env sequences contained one or more mutations within the eight residue fusion-peptide sequence, which individually showed little effect on VRC34.01 neutralization. "Affected" sequences contained one or more residues (513I, 513A, 514b_T, 515M, 518M, 518L, 519I, or 519L) that diminished VRC34.01 neutralization. "Untested" sequences contained one or more untested residues plus tolerated residues.
26. P. D. Kwong *et al.*, *Nature* **393**, 648–659 (1998).
27. A. Trkola *et al.*, *Nature* **384**, 184–187 (1996).
28. J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, S. Subramaniam, *Nature* **455**, 109–113 (2008).
29. M. J. Harvey, G. Giupponi, G. D. Fabritiis, *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).
30. While 500 ns is not enough time to sample large-scale conformational transitions associated with the prefusion HIV-1 Env trimer, it should be sufficient to assess local motions of the fusion peptide.
31. When trimers are positioned into nearest-neighbor contact, only the N-terminal three residues of the fusion peptide can extend to contact a fusion peptide on a neighboring trimer. The N-terminal three residues of the fusion peptide are Ala-Val-Gly. Although the Ala-Val are somewhat hydrophobic, the positive charges of two N termini likely repel, preventing aggregation. In terms of close-by glycans, these may help to prevent aggregation, especially glycans N88, N611, and N618. However, it appears to be primarily the recessed nature of the fusion peptide, coupled to the fact that only eight residues of the fusion peptide are flexible (Fig. 4), that prevents aggregation.
32. Materials and methods are available as supplementary materials on *Science* Online.
33. M. Pancera *et al.*, *Nat. Struct. Mol. Biol.* **20**, 804–813 (2013).
34. J. H. Lee, G. Ozorowski, A. B. Ward, *Science* **351**, 1043–1048 (2016).
35. The structure of PGT151 in complex with HIV-1 Env was recently reported (*34*). This structure shows PGT151 to recognize the fusion peptide in an exposed, extended conformation.
36. Trimer (or trimer-antibody complex) was injected for 60 s onto CD4-IgG (immunoglobulin G) captured on a Biacore CM5 chip coated with a monoclonal mouse anti-human IgG (Fc) antibody (GE Healthcare), and binding responses were measured 60 s after the injection ended. Binding to CD4-Ig normalized with binding to antibody 2G12, which was measured similarly on a parallel-flow cell.
37. TZM-bl cells were incubated with JR-FL Env pseudovirus in the presence of antibodies at 50 μg/ml or medium control and then stained with biotinylated 2G12 followed by phycoerythrin (PE)-conjugated streptavidin. MFI was obtained from at least 10,000 cell events and normalized to the medium control.
38. The HIV Env trimer is conformationally dynamic. Binding and smFRET experiments show that VRC34 do not bind to the open co-receptor binding-competent conformation. Thus, any molecule spontaneously in that conformation would not be neutralized. The conformation specificity of VRC34.01 for the closed and intermediate conformation may explain the observed incomplete neutralization.
39. Specifically, soluble SOSIP.664 trimers were incubated with soluble, two-domain CD4, either alone or following preincubation with Fab, and the incubated trimer samples were injected over a 17b-coupled surface for 30 s, with response units reported 10 s after the injection ended.

ACKNOWLEDGMENTS

We thank the members of the Flow Cytometry Core, Vaccine Research Center, and especially R. Nguyen for his assistance in cell sorting. We thank members of the Structural Biology Section, Structural Bioinformatics Core Section, and Human Immunology Section, Vaccine Research Center, for discussions and comments on the manuscript. We thank J. Baalwa, D. Ellenberger, F. Gao, B. Hahn, K. Hong, J. Kim, F. McCutchan, D. Montefiori, L. Morris, J. Overbaugh, E. Sanders-Buell, G. Shaw, R. Swanstrom, M. Thomson, S. Tovanabutra, C. Williamson, and L. Zhang for contributing the HIV-1 Envelope plasmids used in our neutralization panel. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. Nucleotide sequences of VRC34.01-VRC34.07 variable regions are available under GenBank accession numbers KU711816 to KU711829. Atomic coordinates and structure factors of the reported crystal structures have been deposited with the Protein Data Bank (PDB) under accession codes 5I8C, 5I8E, and 5I8H. EM data have been deposited with the Electron Microscopy Data Bank under accession code EMD-8125. Support for this work was provided by the Intramural Research Programs of the Vaccine Research Center and Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, and from the International AIDS Vaccine Initiative's (IAVI's) Neutralizing Antibody Consortium. IAVI's work is made possible by support from many donors, including the Bill & Melinda Gates Foundation, the Ministry of Foreign Affairs of Denmark, Irish Aid, the Ministry of Finance of Japan, the Ministry of Foreign Affairs of the Netherlands, the Norwegian Agency for Development Cooperation (NORAD), the UK Department for International Development (DFID), and the U.S. Agency for International Development (USAID). The full list of IAVI donors is available at www.iavi.org. S.C.B. was supported by NIH grants R01GM079238 and P01GM56550. W.M. was supported by NIH grants R01GM116654 and P01GM56550. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-Eng-38. The following patent applications are pending: U.S. Patent Application 13/202,351, Methods and Compositions for Altering Photopysical Properties of Fluorophores via Proximal Quenching (S.C.B.); U.S. Patent Application 14/373,402 Dye Compositions, Methods of Preparation, Conjugates Thereof, and Methods of Use (S.C.B.); and International and U.S. Patent Application PCT/US13/42249 Reagents and Methods for Identifying Anti-HIV Compounds (S.C.B. and W.M.). S.C.B. holds an equity interest in Lumidyne Technologies Corporation.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6287/828/suppl/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S3
Databases S1 to S3
References (*40*–*68*)

10 December 2015; accepted 30 March 2016
10.1126/science.aae0474

ORIGIN OF LIFE

A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway

Sidney Becker,* Ines Thoma,* Amrei Deutsch, Tim Gehrke, Peter Mayer, Hendrik Zipse, Thomas Carell†

The origin of life is believed to have started with prebiotic molecules reacting along unidentified pathways to produce key molecules such as nucleosides. To date, a single prebiotic pathway to purine nucleosides had been proposed. It is considered to be inefficient due to missing regioselectivity and low yields. We report that the condensation of formamidopyrimidines (FaPys) with sugars provides the natural *N*-9 nucleosides with extreme regioselectivity and in good yields (60%). The FaPys are available from formic acid and aminopyrimidines, which are in turn available from prebiotic molecules that were also detected during the Rosetta comet mission. This nucleoside formation pathway can be fused to sugar-forming reactions to produce pentosides, providing a plausible scenario of how purine nucleosides may have formed under prebiotic conditions.

It is assumed that life originated from a simple set of small molecules. These prebiotic molecules can be found on comets and as components of Earth's early atmosphere (*1*, *2*). The RNA-world hypothesis (*3*, *4*) posits that these molecules assembled to produce nucleosides, and later informational polymers, which are able to replicate themselves. A prebiotically plausible route to pyrimidine nucleosides provides these key building blocks in a stepwise reaction sequence in high yields (*5*). The situation is less clear for the purine nucleosides, which are not only central components of DNA and RNA but also constituents of adenosine triphosphate and guanosine triphosphate and substructures in many coenzymes (*6*).

Presently, the only prebiotic route to purine nucleosides proposes the condensation of the complete nucleobase, such as adenine (**1**, R^1 = NH_2, R^2 = H), with ribose **2** in the molten state (Fig. 1A) (*7*). This reaction provides complex mixtures of purine ribosides with yields of about 4% for adenosine (β-fA). The reaction needs adenine **1** as its hydrochloride salt and subsequent equilibration of the mixture under basic (NH_4OH) conditions (*8*). Because many of the N atoms of the purine skeleton can react, a regioselectivity problem is encountered, which is responsible for the low yields. In addition, the *N*-9 atoms of the purine heterocycles, which connect the nucleobase with the sugar in canonical purine ribosides, are particularly unreactive. This, together with the fact that the purine heterocycles themselves are available only in low yields under prebiotic

Ludwig-Maximilians-Universität München, Department für Chemie, D-81377 Munich, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: thomas.carell@cup.uni-muenchen.de

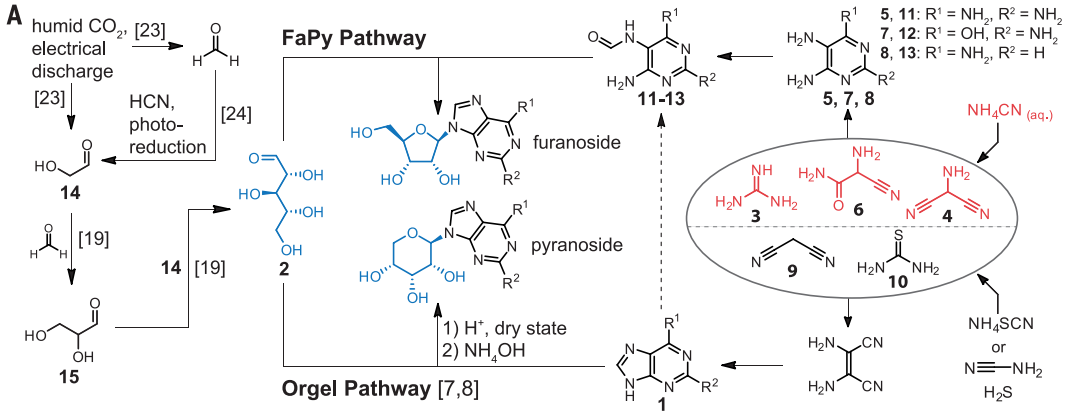


Fig. 1. Prebiotic pathways to purine ribosides. (**A**) The FaPy pathway starts with the formamidopyrimidines **11** to **13**, which are prebiotically available from simple starting materials (encircled) via multi-aminopyrimidines. The molecules in red are derived directly from NH_4CN. For comparison, Orgel's pathway is shown involving coupling of the full nucleobases. Ribose is generated from glycolaldehyde and formaldehyde, formed from an early atmosphere containing humid CO_2 (left). (**B**) Nitrogen reactivity analysis of the multi-aminopyrimidines and crystal structure of the monoprotonated 4,5,6-triaminopyrimidine **8** with bond length and the bond angles of the critical amino groups.

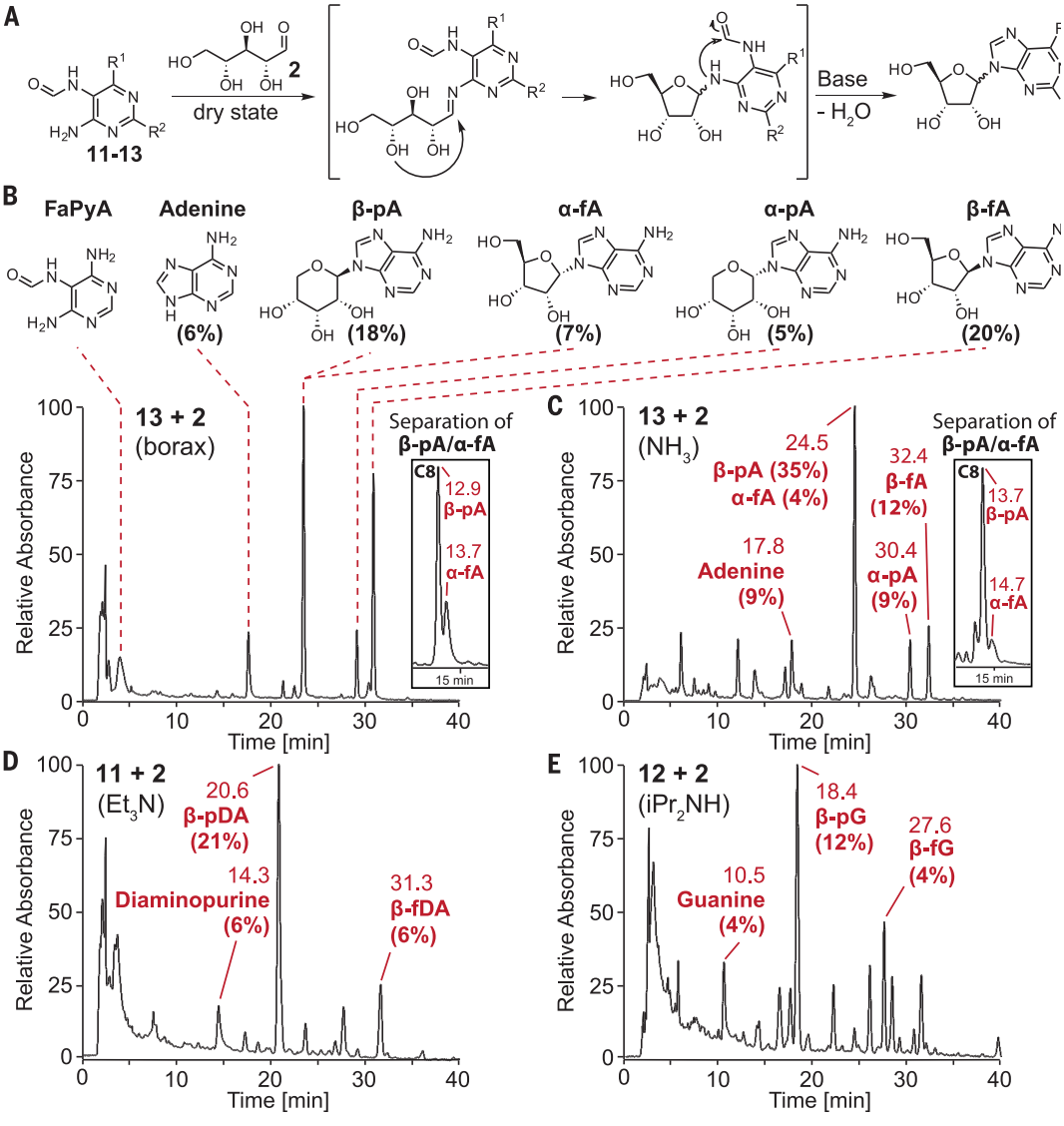


Fig. 2. Formation of the purine ribosides. The FaPy pathway provides pyranosides (p) and furanosides (f) as α- and β-isomers (A, adenine; DA, diaminopurine; G, guanine). (**A**) Mechanism of the FaPy pathway. (**B**) Depiction of the formed products analyzed by HPLC [ultraviolet (UV) detection] of the reaction of **13** (R^1, NH_2; R^2, H) with ribose **2**, followed by treatment with either borax (0.125 M at 100°C for 2 days) or (**C**) NH_3 (0.5 M for 7 days). (**D**) Reaction of **11** (R^1, NH_2; R^2, NH_2) with ribose **2**, followed by heating in Et_3N (0.5 M for 14 days). (**E**) Reaction of **12** (R^1, OH; R^2, NH_2) with ribose **2** and subsequent heating in iPr_2NH (0.5 M for 12 days).

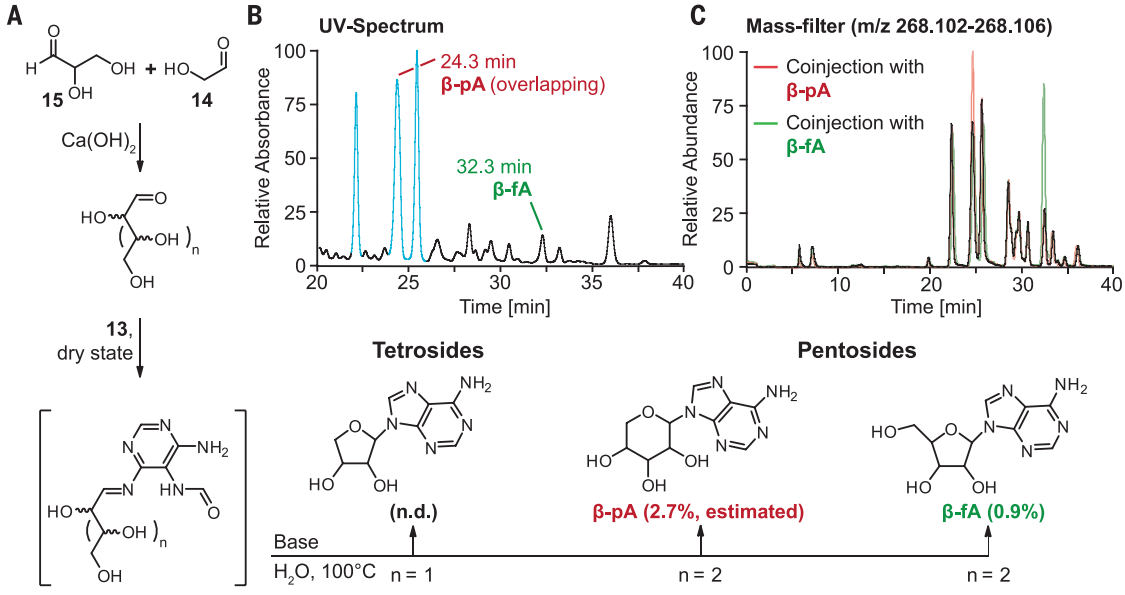


Fig. 3. Prebiotic formation of purine ribosides from sugar precursors. (**A**) Glycolaldehyde **14** and glyceraldehyde **15** react in the presence of $Ca(OH)_2$ and formamidopyrimidine **13** to tetrosides and pentosides. (**B**) HPLC analysis (UV detection) of the reaction. The four main peaks (light blue, peak at 24.30 min, two peaks overlapping) are the α- or β-pyranosyl isomers of the aldopentoses (arabinose, lyxose, ribose, and xylose). (**C**) Detection of the purine pentosides by HPLC-MS with a specific mass filter and coinjection of synthetic material.

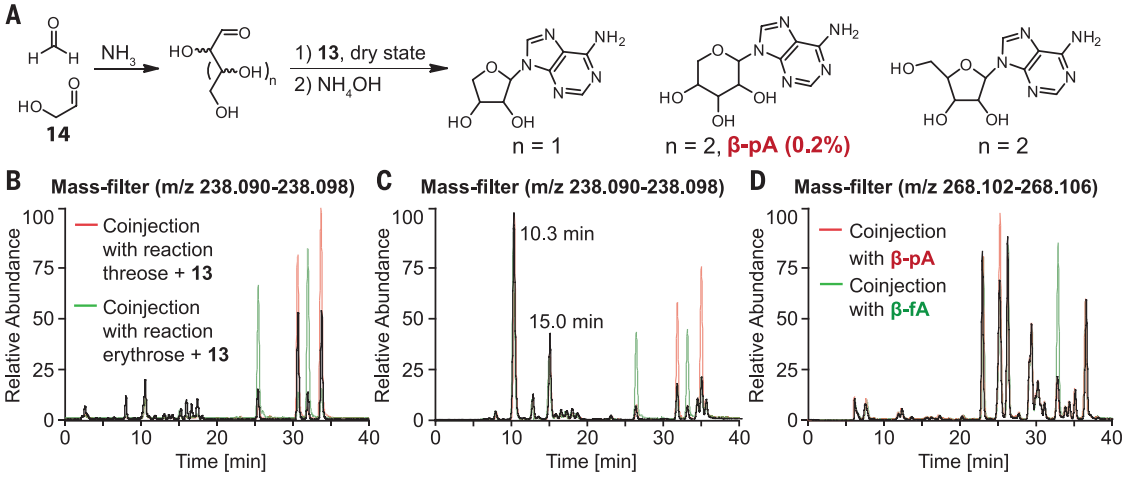


Fig. 4. Prebiotic formation of purine nucleosides in a formose-type reaction. (**A**) Reaction scheme. (**B**) Detection of the α/β-purine tetroside from HCHO (0.21 M) and glycolaldehyde (0.4 M) by HPLC-MS and coinjection (red, threoside; green, erythroside). (**C**) Increased concentrations of HCHO (2.1 M) and glycolaldehyde (4 M) results in increased formation of two unidentified side products at 10.26 and 14.99 min with decreasing tetroside yields. (**D**) Same reaction conditions as (C) produced more pentosides (n = 2) compared with tetrosides (n = 1). The β-pA (0.2%) and β-fA were identified by coinjection.

conditions (0.5%) (*9*, *10*), prompted us to search for an alternative prebiotic pathway (Fig. 1A, FaPy pathway).

The FaPy pathway starts from aminopyrimidines, which are accessible from simple molecules such as NH_4CN under prebiotically plausible conditions. Guanidine **3** (available from cyanamide and NH_3), for example reacts, with the HCN trimer aminomalononitrile **4** to produce tetraaminopyrimidine **5** (72%) (*11*). Despite the high nitrogen content of such molecules, they are very robust. The aminopyrimidines are often electron rich and hence readily oxidized in an oxygen-containing atmosphere but stable in the absence of oxygen (*2*). Other amino-substituted pyrimidines may have been formed when the HCN trimer **4** is hydrolyzed in the presence of formaldehyde, to produce aminocyanoacetamide **6** (*12*), which reacts with guanidine **3** to produce triaminopyrimidinone **7** (53%) (*12*). Finally, triaminopyrimidine **8** is available by condensation of malononitrile **9** with thiourea **10** (84%) (*13*), followed by nitrosation in water and desulfuration (Ni and H_2 or HCOOH) (see the supplementary materials) (*14*). Thiourea **10** is prebiotically available via addition of H_2S to cyanamide or directly from ammonium thiocyanate (*15*). The so-formed aminopyrimidines (**5**, **7**, **8**) feature, in contrast to the purines, high solubility in water.

Although the aminopyrimidines **5**, **7**, and **8** feature multiple nucleophilic positions, a deeper reactivity analysis and symmetry relations uncover a proton-assisted reactivity guidance (red, green, and blue N atoms in Fig. 1B) that eliminates the regioselectivity problem. The two in-ring N atoms (red, Fig. 1B) are due to the present exocyclic amino groups unusually basic (pK_a ~ 7.5), which leads to monoprotonation already under slightly acidic conditions. Due to the symmetry of the molecules, protonation of either one of the two-ring N atoms blocks the reactivity of the exocyclic amino groups (2, 4, and 6, in blue, Fig. 1B) in ortho- and para-position. As such, the proton serves as a protecting group for most amines on the ring structure, leaving only the *N*-5 amino group (green, Fig. 1B) nucleophilic. To support this analysis, we crystallized monoprotonated triaminopyrimidine **8**. Structural analysis shows short C–N bonds (1.33 Å) and planar geometry for amino groups 4 and 6, allowing full conjugation with the electron-deficient pyrimidine ring. In contrast, the *N*-5 amino group (green, Fig. 1B) has a longer C–N bond (1.41 Å) and pyramidal geometry (109°), hence showing increased sp^3 character, which results in a nucleophilic lone pair oriented at an angle of 65° to the ring plane.

When we heated aminopyrimidines **5**, **7**, and **8** with formic acid or formamide, we obtained—in agreement with this analysis—in all cases

regioselectively just the *N*-5 formylated FaPy product (**11** to **13**, Fig. 1A) in excellent yields (70 to 90%). Formylation provides the C1 unit needed for later purine formation. It also acts as a protecting group for the 5-NH_2 group. The amino groups 2, 4, and 6 are now, under neutral conditions, available for a condensation reaction with a sugar. The amino groups in ortho position to the formamide are symmetry-related by a C2 axis (in **11** and **13**), which eliminates the second regioselectivity problem. Reaction of either one with ribose provides the same product. Indeed, when we reacted the formamides **11** to **13** with ribose **2**, followed by stirring of the mixture under slightly basic conditions, we obtained in all cases regioselectively only the *N*-9–reacted α/β-furanosides and α/β-pyranosides (Fig. 2). To assign the high-performance liquid chromatography and mass spectrometry (HPLC-MS)–detected compounds, we synthesized all expected α- and β-pyranosides and furanosides (Fig. 2 and supplementary materials) and performed coinjection studies. To further support the data regarding adenosine, we also isolated this compound (β-fA) by HPLC and confirmed the correct structure by nuclear magnetic resonance (fig. S1). The β- and not the α-anomers are observed as the main products in all experiments. The reaction mechanism involves condensation of FaPy compounds **11** to **13** with ribose to produce an imine intermediate, followed by sugar-ring closure to produce the FaPy ribosides. The following second-cyclization reaction under mild basic conditions provides the purine skeleton (Fig. 2A). This cyclization can occur even without previous reaction with the sugar, which generates the corresponding (sugar-free) nucleobase as a side product. The formed heterocycles can, however, be prebiotically recycled. Formamide in the presence of TiO_2 or, alternatively, riboflavin in the presence of light was shown to allow degradation of the purines back into the corresponding FaPy compounds (Fig. 1A and fig. S2) (*16*, *17*).

We studied purine nucleoside synthesis in detail and found that slightly basic conditions gave the highest yields. Reaction of **13** with **2** in pure ammonia solutions (0.5 M) provided the main purine nucleosides β-pyranosyladenosine (β-pA) (35%) and β-fA (12%) in good yields, although ammonia will certainly react with ribose as well (Fig. 2C). We found that the presence of basic amino acids (0.5 M) (fig. S3, A and B) instead of ammonia, such as lysine (β-pA, 13% and β-fA, 4%) and arginine (β-pA, 22% and β-fA, 6%), gave satisfactory yields as well. Good yields were also obtained when we used borax. In this case, the FaPy building block **13** reacted with ribose **2** (Fig. 2B) to produce mainly furanosides (*18*) with the canonical β-adenosine (β-fA) formed as the main product (20%). Similar results were obtained with a carbonate/borate buffer (*19*). Under these conditions, the yields for β-pA and β-fA were 15% and 18%, respectively (fig. S3C). Silicate buffers, reported to stabilize pentoses, however, failed in our hands (*20*). The FaPy pathway also provides direct access to diaminopurine and guanine ribosides upon reaction of FaPy building block **11** (Fig. 2D) and **12** (Fig. 2E) with ribose **2**. Although we did not optimize the reaction conditions for these nucleosides, the purine β-furanosides are obtained in yields between 4 and 6% when the reaction is equilibrated with alkylamines. The reaction of **11** with **2** directly provides some guanosine via hydrolysis of the diaminopurine moiety. Formation of the alternative hydrolysis product isoguanosine was not observed (*21*, *22*). Importantly, the FaPy pathway provides canonical β-adenosine in yields of up to 20%. The highest total yields for *N*-9 ribosides of up to 60% was achieved using simple amines as bases that were also discovered on comet 67P/Churyumov-Gerasimenko (Fig. 2, B to E) (*1*).

We next asked whether the FaPy pathway can be linked to potential prebiotic sugar syntheses (Fig. 3). An attractive route involves the formaldehyde and glycolaldehyde **14**–based formose reaction, e.g., as a mineral-guided version (*19*). Formaldehyde and small amounts of **14** are both formed by electrical discharge of humid CO_2 (*23*). The formed formaldehyde can further react with HCN by photoreduction to produce **14** (*24*). A central reaction in the mineral-guided sugar formation is an aldol addition of **14** with formaldehyde, which produces glyceraldehyde **15**. This can react with a second molecule of **14** to pentose sugars (Fig. 1A) (*19*). When we reacted **14** and **15** (both 16.7 mM) in the presence of $Ca(OH)_2$ (23 mM) (*25*) with FaPy **13**, we indeed found purine pentosides as the main products (estimated yield, 21%). Quantification of D/L β-adenosine (D/L-β-fA, Fig. 3) gave 0.9% (based on **13**). Despite overlapping signals, we were also able to estimate a yield of 2.7% for the D/L-β-pA. When we performed the reaction in the presence of borax, the D/L-β-fA yield increased among all other aldopentoses to 1.3% ultimately (*18*). Product identification was achieved by HPLC-MS and coinjection studies (Fig. 3C). Small amounts of purine threosides and erythrosides were also found (fig. S4).

The FaPy route delivers ribosides even directly from formaldehyde and glycolaldehyde **14** (Fig. 4). We mixed formaldehyde (0.21 M) and glycolaldehyde (0.40 M) with $Ca(OH)_2$ (0.5 M), borax (0.125 M), or NH_3 (0.5 M) in water and heated the mixture for 1 hour at 65°C. The FaPy compound **13** (0.12 M) was added and the mixture was heated in an open vessel for another 8 hours at 100°C. Water (or 0.5 M NH_3) was added to the obtained solid, and the mixture was heated for another 2 to 3 days at 100°C. Reactions containing borax or $Ca(OH)_2$ gave complex spectra with only small signals for nucleosides. In the experiments with NH_3, however, purine threosides and erythrosides were formed as the main nucleoside components, with the threose nucleosides slightly favored (*19*). To identify the tetrosides, we again performed coinjection studies (Fig. 4, B and C). Besides the purine tetrosides, we were able to identify D/L-β-pA and D/L-β-fA (D/L-adenosine) in this one-pot reaction (Fig. 4D), with a preference for the β-pyranosyl compound (D/L-β-pA, 0.2% based on **13**).

Overall, we show that the reported FaPy pathway with the corresponding aminopyrimidine (*11*, *12*) precursor molecules provides a feasible pathway to purine nucleosides that is compatible with early Earth conditions and that provides adenosine under a variety of conditions in yields of up to 20%. The starting materials are small organic molecules such as HCN, NH_3, and particularly formic acid derivatives that were all discovered on comets like 67P/Churyumov-Gerasimenko (*1*).

REFERENCES AND NOTES

1. F. Goesmann *et al*., *Science* **349**, aab0689 (2015).
2. J. F. Kasting, *Science* **259**, 920–926 (1993).
3. F. H. C. Crick, *J. Mol. Biol.* **38**, 367–379 (1968).
4. L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
5. M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **459**, 239–242 (2009).
6. H. B. White 3rd, *J. Mol. Evol.* **7**, 101–104 (1976).
7. W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **67**, 25–33 (1972).
8. W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1**, 249–257 (1972).
9. J. Oró, *Biochem. Biophys. Res. Commun.* **2**, 407–412 (1960).
10. J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **94**, 217–227 (1961).
11. U. P. Trinks, A. Eschenmoser, thesis, ETH Zurich (1987), ETH 2-collection; http://dx.doi.org/10.3929/ethz-a-000413538.
12. K. E. Koch, A. Eschenmoser, thesis, ETH Zurich (1992), ETH e-collection; http://dx.doi.org/10.3929/ethz-a-000694124.
13. W. Traube, *Liebigs Ann. Chem.* **331**, 64–88 (1904).
14. J. Clark, R. K. Grantham, J. Lydiate, *J. Chem. Soc. C. Org.* **1968**, 1122 (1968).
15. M. P. Robertson, M. Levy, S. L. Miller, *J. Mol. Evol.* **43**, 543–550 (1996).
16. R. Saladino *et al*., *Chem. Bio. Chem.* **4**, 514–521 (2003).
17. T. Mori, K. Tano, K. Takimoto, H. Utsumi, *Biochem. Biophys. Res. Commun.* **242**, 98–101 (1998).
18. A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **303**, 196 (2004).
19. H. J. Kim *et al*., *J. Am. Chem. Soc.* **133**, 9457–9468 (2011).
20. J. B. Lambert, S. A. Gurusamy-Thangavelu, K. Ma, *Science* **327**, 984–986 (2010).
21. R. B. Trattner, G. B. Elion, G. H. Hitchings, D. M. Sharefkin, *J. Org. Chem.* **29**, 2674–2677 (1964).
22. M. D. Kirnos, I. Y. Khudyakov, N. I. Alexandrushkina, B. F. Vanyushin, *Nature* **270**, 369–370 (1977).
23. W. Löb, *Berichte der Dtsch. Chem. Gesellschaft* **46**, 684–697 (1913).
24. D. J. Ritson, J. D. Sutherland, *Angew. Chem. Int. Ed.* **52**, 5845–5847 (2013).
25. G. Harsch, H. Bauer, W. Voelter, *Liebigs Ann. Chem.* **1984**, 623–635 (1984).

ACKNOWLEDGMENTS

We thank the Deutsche Forschungsgemeinschaft (SFB749, SFB646, and SFB1032) and the Excellence cluster EXC114 (Center for Integrated Protein Science) for financial support. Additional figures and synthetic procedures can be found in the supplementary materials. The crystallographic data (CCDC 1416691) can be obtained from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6287/833/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1 to S3
References (*26*–*36*)

20 August 2015; accepted 15 March 2016
10.1126/science.aad2808

www.sciencemag.org/content/352/6287/833/suppl/DC1

Supplementary Materials for

A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway

Sidney Becker, Ines Thoma, Amrei Deutsch, Tim Gehrke,
Peter Mayer, Hendrik Zipse, Thomas Carell*

*Corresponding author. Email: thomas.carell@cup.uni-muenchen.de

Published 13 May 2016, *Science* **352**, 833 (2016)
DOI: 10.1126/science.aad2808

This PDF file includes:

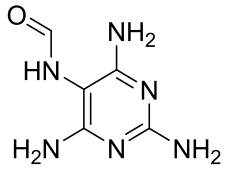
Materials and Methods

General Information

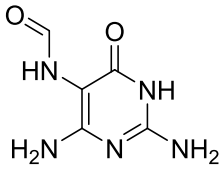
Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR, Carbosynth or Acros organics and used without further purification. Solutions were concentrated *in vacuo* on a *Heidolph* rotary evaporator. The solvents were of reagent grade or purified by distillation. Chromatographic purification of products was accomplished using flash column chromatography on *Merck* Geduran Si 60 (40-63 μm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on *Merck* 60 (silica gel F_{254}) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or standard staining solutions. ^{1}H- and ^{13}C-NMR spectra were recorded in deuterated solvents on *Varian Oxford 200, Bruker ARX 300, Varian VXR400S, Varian Inova 400, Bruker AMX 600 and Bruker AVIIIHD 400* spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. High-resolution ESI spectra were obtained on the mass spectrometers *Thermo Finnigan* LTQ FT-ICR. IR measurements were performed on *Perkin Elmer Spectrum BX FT-IR* spectrometer with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were measured on a *Büchi* B-540. For preparative HPLC purification a *Waters 1525 binary HPLC Pump* in combination with a *Waters 2487 Dual Absorbance Detector* was used, with *Macherey-Nagel VP 250/10 Nucleosil 100-7 C18* reversed phase column. The analysis of the prebiotic reactions were analyzed by LC-ESI-MS on a *Thermo Finnigan LTQ Orbitrap XL* and were chromatographed by a *Dionex Ultimate 3000 HPLC* system with a flow of 0.15 mL/min over an *Interchim Uptisphere120-3HDO C18* column or an *Macherey-Nagel EC 125/2 Nucleosil 120-3 C8* column. The column temperature was maintained at 30 °C. Eluting buffers were buffer A (2 mM $HCOONH_4$ in H_2O (pH 5.5)) and buffer B (2 mM $HCOONH_4$ in H_2O/MeCN 20/80 (pH 5.5)). The gradient for all samples containing FaPyA **13** was 0 → 45 min; 0% → 10% buffer B for the *Interchim* column and 0 → 45 min; 0% → 2% buffer B for the *Macherey-Nagel* C8 column. For samples containing FaPyDA **11** or FaPyG **12**, a gradient of 0 → 45 min; 0% → 7% buffer B for the *Interchim* column was used. The elution was monitored at 260 nm (*Dionex Ultimate 3000 Diode Array Detector).* The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of *m/z* 120-1000 with a resolution of 30000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5 μM). The parameters used in this section were sheath gas flow rate, 16 arb; auxiliary gas flow rate, 11 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200 °C; capillary voltage, 20 V, tube lens 65 V. The X-ray intensity data of 4,5,6-triaminopyrimidine hydrochloride (**8**) were collected at 173 K on an Oxford Diffraction Xcalibur3 diffractometer equipped with a Mo sealed tube (MoKα radiation). The structure was solved by direct methods with SIR97 (*26*) and refined by full-matrix least-squares techniques on F^2 with SHELXL (*27*). All hydrogen atoms have been refined freely.

Synthetic Procedures

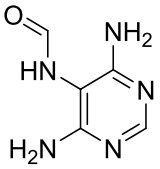
Synthesis of formamides under prebiotic conditions



11



12



13

2,4,6-triamino-5-formamidopyrimide (FaPyDA, 11)

2,4,5,6-tetraaminopyrimidine-sulfate (2.38 g, 10 mmol, 1.0 eq) was suspended in formic acid (30 mL) and sodium formiate (680 mg, 10 mmol, 1.0 eq.) was added. The mixture was heated to reflux for 2 h. The solvent was evaporated and water was added until everything was dissolved. Ammonia solution (5 M) was added to the solution until pH 8 was reached. The solution was cooled overnight at 4 °C in a fridge. The crystals were filtered off to obtain FaPyDA **11** (1.3 g, 7.7 mmol, 77%) as light orange crystals. NMR shows cis/trans rotamers of the formamide group.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.45 (d, cis *J*=1.4, 1H; NH), 8.06 (d, trans *J*=11.7, 1H; NH), 8.05 (d, cis *J*=1.4, 1H; CHO), 7.70 (d, trans *J*=11.7, 1H; CHO), 5.76 (s, cis, 4H; C4NH_2 and C6NH_2), 5.56 (s, cis, 2H; C2NH_2), 5.55 (s, trans, 4H; C4NH_2 and C6NH_2), 5.46 (s, trans, 2H; C2NH_2); **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.6 (trans, CHO), 161.67 (trans, C4 and C6), 161.23 (cis, CHO), 161.08 (trans, C2), 161.05 (cis, C2), 160.14 (cis, C4 and C6), 86.68 (trans, C5), 86.40 (cis, C5). **HRMS** (ESI+): calc. for: $[C_5H_9N_6O]^+$ 169.0832, found: 169.0831 $[MH]^+$.

2,6-diamino-4-oxo-5-formamidopyrimidine (FaPyG, 12)

2,4,5-triamino-6-hydroxypyrimidine-sulfate (2.39 g, 10 mmol, 1.0 eq) was suspended in formic acid (30 mL) and sodium formiate (680 mg, 10 mmol, 1.0 eq.) was added. The mixture was heated to reflux for 2 h. The solvent was evaporated and water was added until everything was dissolved. Ammonia solution (5 M) was added to the solution until pH 8 was reached. The solution was cooled for 4-6 h at 4 °C in a fridge. The precipitate was filtered off to obtain FaPyG (1.45 g, 7.7 mmol, 85%) as light beige solid. NMR shows cis/trans rotamers of the formamide group.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.01 (s, cis/trans 2H; $NH_{arom.}$), 8.47 (d, cis *J*=1.6, 1H; NH), 8.00 (d, cis *J*=1.6, 1H; CHO), 7.83 (d, trans *J*=11.6, 1H; NH), 7.73 (d, trans *J*=11.6, 1H; CHO), 6.21 (s, trans, 2H; C2NH_2), 6.14 (s, cis, 2H; C2NH_2), 6.01 (s, trans, 2H; C4NH_2), 5.78 (s, cis, 2H, C4NH_2). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.82 (trans, CHO), 161.47 (trans, C6), 160.61 (cis, CHO), 160.58 (cis, C6), 159.97 (trans, C4), 159.43(cis, C4), 153.41 (trans, C2), 153.25 (cis, C2), 88.67 (trans, C5), 88.57 (cis, C5). **HRMS** (ESI-): calc. for: $[C_5H_6N_5O_2]^-$ 168.0527, found: 168.0527 $[M-H]^-$.

4,6-diamino-5-formamidopyrimidine (FaPyA, 13)

The reaction was carried out slightly modified according to *C. Haley et al.* (*28*).

4,5,6-triaminopyrimidine-sulfate (620 mg, 2.8 mmol, 1 eq.) was suspended in 1 M NaOAc solution (14 mL, pH 1.1, acidified with HCl). Formic acid (0.4 mL, 10.6 mmol, 3.8 eq.) was added and stirred for 4 days at room temperature. The pH was adjusted with ammonia solution (5 M) to pH 8-9. The solution was kept at 4°C overnight and the crystals were filtered off and washed with cold ethanol to give the product (313 mg, 2.04 mmol, 73%) as colorless crystals.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.78 (d, *J*=1.3, 1H; NH), 8.08 (d, *J*=1.3, 1H; CHO), 7.74 (s, 1H, C2H), 5.98 (s, 4H; C4NH_2 and C6NH_2). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 161.31 (CO), 159.92 (C4 and C6), 156.14 (C2), 94.22 (C5). **HRMS** (ESI+): calc. for: $[C_5H_8N_5O]^+$ 154.0723, found: 154.0725 $[M+H]^+$.

Alternative pathway *via* 4,6-diamino-2-thiolpyrimidine to 4,6-diamino-5-formamidopyrimidine (FaPyA, 13)

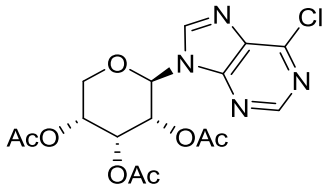
Formation of 4,6-Diamino-5-nitroso-2-thiolpyrimidine from 4,6-diamino-2-thiolpyrimidine was already described elsewhere (*29*).

4,6-Diamino-5-nitroso-2-thiolpyrimidine (20 mg, 0,12 mmol, 1 eq.) was suspended in HCOOH in H_2O (50%, 2 mL). Solid nickel powder (50 mg) was added and stirred in a sealed tube (*ACE* 15 mL pressure tube) at 100 °C for 3 d. The light green precipitate that formed during the reaction was removed by filtration through celite and washed with little amount of water. The solvent was removed *in vacuo* and the product was purified by column chromatography (EtOAc/Acetone/EtOH/H_2O 15:3:4:3) to give FaPyA **13** (9 mg, 0,06 mmol, 50%) as a colourless solid. The analytical data matched the ones described above for FaPyA **13**.

Synthesis of adenine nucleosides

α-adenosine and β-adenosine were bought from *Sigma-Aldrich*. The α/β-pyranosyl isomers were synthesized according to the following procedures.

6-chloro-9-(2',3',5'-tri-*O*-acetyl-*β*-D-ribopyranosyl)-9*H*-purine (16)

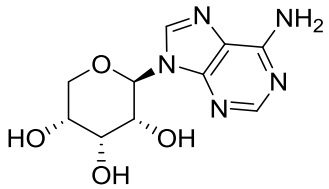


The reaction was carried out slightly modified to *Vorbrüggen et al*. (*30*).

Tetraacetyl-D-ribopyranose (*31*) (161 mg, 0.506 mmol, 2 eq.) and 6-chloropurine (40 mg, 0.259 mmol, 1.00 eq.) were suspended in acetonitrile (1 mL) under nitrogen. Bis(trimethylsilyl)acetamide (0.11 mL, 0.467 mmol, 1.80 eq) was added and the reaction mixture was heated to 60 °C. After the suspension turned into a clear solution TMSOTf (0.09 mL, 0.505 mmol, 2.00 eq) was added. After stirring at 60 °C for 2 h ethyl acetate was added and the reaction mixture was washed three times with sat. $NaHCO_3$, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude mixture was purified *via* column chromatography (ethyl acetate/isohexane=1:1 → 4:1) to yield nucleoside **16** (73 mg, 44%) as a white solid.

mp: 75–77 °C; **^{1}H-NMR** (400 MHz, Chloroform-*d*): δ = 8.78 (s, 1H; C2), 8.26 (s, 1H; C8), 6.06 (d, *J*=9.7, 1H; HC1'), 5.85 (dd, *J*=2.5, 2.3, 1H; HC3'), 5.66 (dd, *J*=9.7, 2.8, 1H; HC2'), 5.25 (ddd, *J*=10.8, 5.78, 2.8, 1H; HC4'), 4.10 (ddd, *J*=11.1, 5.7, 0.8, 1H; HaC5'), 4.05 – 3.99 (m, 1H; HbC5'), 2.27 (CH_3), 2.05 (CH_3), and 1.78 ppm (CH_3); **^{13}C-NMR** (101 MHz, $CDCl_3$): δ=169.7 (CO), 169.3 (CO), 168.7 (CO), 152.4 (C2), 151.7 (C4), 151.6 (C6), 143.2 (C8), 131.7 (C5), 79.0 (C1'), 68.0 (C3'), 67.8 (C2'), 65.7 (C4'), 63.8 (C5'), 20.8 (CH_3), 20.5 (CH_3), and 20.2 ppm (CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 1746 (s), 1591 (m), 1562 (m), 1370 (m), 1241 (s), 1196 (vs), 1084 (m) 1040 (s), 731 (w); **HRMS** (ESI+): calc. for: $[C_{16}H_{18}ClN_4O_7]^+$ 413.0859, found: 413.0859 $[M+H]^+$.

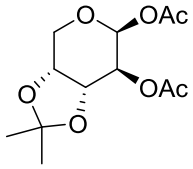
9-*β*-D-ribopyranosyladenine (β-pA, 17)



Ammonia was condensed (ca. 3 mL) and stirred in a sealed tube together with nucleoside **16** (20 mg, 0.048 mmol) at 25 °C for 41 h. The crude mixture was purified *via* HPLC to obtain β-pA **17** (8 mg, 62%) as a white solid.

mp: 241 °C (decomp.); **^{1}H-NMR** (600 MHz, Methanol-d_4): δ = 8.24 (s, 1H; C8), 8.18 (s, 1H; C2), 5.75 (d, *J*=9.3, 1H; HC1'), 4.27 (dd, *J*=9.3, 2.8, 1H; HC2'), 4.21 – 4.20 (m, 1H; HC3'), 3.94 (ddd, *J*=10.6, 4.9, 2.6, 1H; HC4'), 3.89 (dd, *J*=10.7, 10.4, 1H; HaC5'), and 3.77 ppm (dd, *J*=5.2, 10.4, 1H; HbC5'); **^{13}C-NMR** (101 MHz, Methanol-d_4): δ=155.9 (C4), 152.4 (C2), 149.6 (C6), 140.2 (C8), 118.8 (C5), 80.8 (C1'), 71.4 (C3'), 68.8 (C2'), 66.7 (C4'), and 65.2 ppm (C5'); **IR** (cm^{-1}): $\tilde{\nu}$ = 3215 (br, m), 2939 (w), 1663 (s), 1609 (m), 1577 (m), 1490 (w), 1417 (m), 1333 (m), 1291 (w), 1255 (m), 1150 (w), 1114 (m), 1088 (s), 1047 (vs), 1045 (s), 977 (m), 885 (m), 795 (w), 729 (m); **HRMS** (ESI+): calc. for: $[C_{10}H_{14}N_5O_4]^+$ 268.1040, found: 268.1039 $[M+H]^+$.

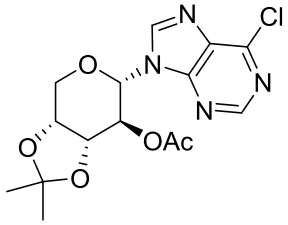
1,2-di-*O*-acetyl-*3,4*-*O*-isopropylidene-*β*-D-arabinopyranose (18)



3,4-O-Isopropylidene-D-ribose (*32*) (10.3 g, 54.2 mmol, 1 eq.) and NaOAc (22.2 g, 271 mmol, 5 eq.) was dissolved in Ac_2O (81.3 mL, 860 mmol, 16 eq.). The mixture was stirred for 0.5 h at 100 °C. Ice cold water was added to the mixture and stirred for 1 h. It was extracted with DCM (3 x 20 mL) and washed with brine (1 x 20 mL). The combined organic layers were dried over Na_2SO_4 and the solvent evaporated *in vacuo*. The product was recrystallized from EtOH (10 mL). The crystals were filtered off and dried in high vacuum to yield the β-anomer **18** (6.56 g, 23.9 mmol, 44%) as colorless needles.

mp: 106 °C; **^{1}H-NMR** (400 MHz, DMSO-d_6): δ = 5.55 (d, *J*=7.8, 1H; HC1‘), 4.92 (t, *J*=7.3, 1H; HC2‘), 4.31 – 4.23 (m, 2H; HC3‘, HC4‘), 4.09 – 3.90 (m, 2H; HaC5‘, HbC5‘), 2.06 (s, 3H; CH_3), 2.04 (s, 3H; CH_3), 1.43 (s, 3H; CH_3) und 1.28 ppm (s, 3H; CH_3); **^{13}C-NMR** (101 MHz, DMSO-d_6): δ=169.3 (CO), 169.2 (CO), 109.4 (C_q), 91.2 (C1‘), 75.2 (C3‘), 72.5 (C4‘), 71.1 (C2‘), 63.0 (C5‘), 27.6 (CH_3), 26.1 (CH_3), 20.6 (CH_3) und 20.6 ppm (CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 2983 (w), 2934 (w), 1739 (s), 1371 (m), 1213 (vs), 1132 (m), 1039 (vs), 968 (m), 846 (s), 804 (m); **HRMS** (ESI+): calc. for: $[C_{12}H_{18}NaO_7]^+$ 297.0945, found: 297.0945 $[M+Na]^+$.

6-chloro-9-(2’-*O*-acetyl-3’,4’-*O*-isopropylidene-*a*-D-arabinopyranosyl)-9*H*-purine (19)

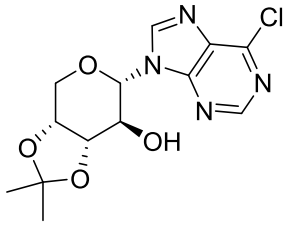


6-chloropurine (155 mg, 1 mmol, 1 eq.) was suspended in dry MeCN (8 mL) under inert atmosphere. Bis(trimethylsilyl)acetamide (480 µL, 1.9 mmol, 1.9 eq.) was added and the mixture was heated to 40°C until the solution became clear. TMSOTf (250 µL, 1.40 mmol, 1.4 eq.) was added. The protected sugar **18** (358 mg, 1.3 mmol, 1.3 eq.) was separately dissolved in dry MeCN (2.5 mL) under inert atmosphere and slowly dropped (within 1.5 h) into the silylated nucleobase solution and stirred additional 1.5 h at 40 °C. The orange solution was quenched with sat. $NaHCO_3$ (15 mL) and extracted with DCM (3 x 15 mL) and washed with brine (1 x 15 mL). The combined organic layers were dried over Na_2SO_4 and the solvent evaporated *in vacuo*. The product was purified by column chromatography (*i*Hex/acetone = 7:3 → 1:1) to yield the nucleoside **19** (192 mg, 0.52 mmol, 52%) as a pale yellowish foam.

^{1}H-NMR (400 MHz, Chloroform-*d*) δ = 8.72 (s, 1H; HC2), 8.39 (s, 1H; HC8), 5.71 (d, *J*=8.7, 1H; HC1‘), 5.50 (dd, *J*=8.7, 6.7, 1H; HC2‘), 4.54 – 4.32 (m, 3H; HC3‘, HC4‘, HaC5‘), 4.07 (dd, *J*=13.8, 2.5, 1H; HbC5‘), 1.82 (s, 3H; $COCH_3$), 1.61 (s, 3H; CH_3), 1.39

(s, 3H; CH_3). **^{13}C-NMR** (101 MHz, Chloroform-*d*) δ = 169.33 (CO), 152.28 (C2), 151.65 (C4), 151.34 (C6), 143.49 (C8), 131.40 (C5), 110.92 (C_q), 80.61 (C1‘), 75.92 (C3‘), 72.97 (C4‘), 71.45 (C2‘), 66.25 (C5‘), 27.62 (CH_3), 26.01 (CH_3), 20.47 (CH_3). **IR** (cm^{-1}): $\tilde{\nu}$ = 2987 (w), 2933 (w), 1739 (m), 1591 (m), 1563 (m), 1374 (m), 1340 (m), 1215 (vs), 1052 (vs), 956 (m), 832 (vs); **HRMS** (ESI+): calc. for: $[C_{15}H_{18}ClN_4O_5]^+$ 369.0960, found: 369.0961 $[M+H]^+$.

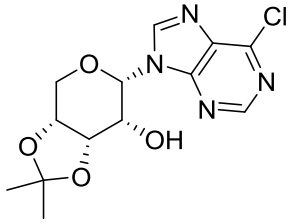
6-chloro-9-(3’,4’-*O*-isopropylidene-*a*-D-arabinopyranosyl)-9*H*-purine (20)



The protected nucleoside **19** (829 mg, 2.25 mmol, 1 eq.) was dissolved in methanolic ammonia (1 M, 20 mL) and stirred for 24 h at room temperature. The solvent was evaporated *in vacuo* and the product purified by column chromatography (*i*Hex/Aceton = 4:1 → 1:1) to yield the deacetylated nucleoside **20** (450 mg, 1.4 mmol, 62%) as a white powder.

mp: 198 °C; **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 8.95 (s, 1H; HC8), 8.82 (s, 1H; HC2), 5.63 (d, *J*=5.7, 1H; OH), 5.53 (d, *J*=9.5, 1H; HC1‘), 4.37 – 4.22 (m, 3H; HC2‘, HC4‘, HaC5‘), 4.20 (dd, *J*=7.2, 5.6, 1H; HC3‘), 4.09 (dd, *J*=13.6, 2.6, 1H; HbC5‘), 1.55 (s, 3H; CH_3), 1.34 (s, 3H; CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 152.51 (C4), 152.36 (C2), 149.71 (C6), 146.72 (C8), 131.50 (C5), 109.36 (C_q), 83.41 (C1‘), 79.20 (C3‘), 73.63 (C4‘), 70.48 (C2‘), 65.82 (C5‘), 28.45 (CH_3), 26.68 (CH_3). **IR** (cm^{-1}): $\tilde{\nu}$ = 3453 (m), 2979 (w), 1589 (m), 1557 (m), 1338 (m), 1200 (s), 1131 (vs), 1080 (s), 952 (m), 846 (s), 826 (s); **HRMS** (ESI+): calc. for: $[C_{13}H_{16}ClN_4O_4]^+$ 327.0855, found: 327.0854 $[M+H]^+$.

6-chloro-9-(3’,4’-*O*-isopropylidene-*a*-D-ribopyranosyl)-9*H*-purine (21)

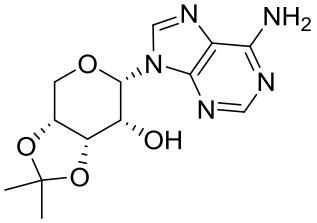


Nucleoside **20** (250 mg, 0.76 mmol, 1 eq.) was dissolved in dry DCM (5 mL) under inert atmosphere. It was added $NaHCO_3$ (255 mg, 3.0 mmol, 4 eq.) and Dess-Martin periodinane (967 mg, 2.3 mmol, 3 eq.). The mixture was stirred for 2 h at 50 °C. The mixture was diluted with EtOAc (5 mL). A solution containing $Na_2S_2O_3$ (0.7 g) and $NaHCO_3$ (0.7 g) in H_2O (10 mL) was added and rigorously stirred until the organic phase became clear. The organic phase was separated and the water phase extracted with EtOAc (2 x 20 mL). The combined organic phases were washed with sat. $NaHCO_3$ and

brine and dried over Na_2SO_4. The solvent was evaporated *in vacuo* to yield the nucleoside as white solid.
The crude nucleoside was directly dissolved in a mixture of DCM/EtOAc/MeOH (2:1:1, 15 mL) and $NaBH_4$ (43 mg, 1.1 mmol, 1.5 eq.) was added. The mixture was stirred for 0.5 h at 0 °C. The reaction was quenched with sat. NH_4Cl and the organic phase separated. The water phase was extracted with EtOAc (1 x 20 mL) and the combined organic layers were washed with sat. $NaHCO_3$ and brine. The organic phase was dried with Na_2SO_4 and the solvent evaporated *in vacuo*. The product was purified by column chromatography (EtOAc/*i*Hex = 2:1 → 4:1) to yield the riboside **21** (110 mg, 0.34 mmol, 45%) as white solid.

mp: 158 °C; **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 8.92 (s, 1H; HC8), 8.81 (s, 1H; HC2), 6.22 (d, *J*=4.5, 1H; HC1'), 5.70 (d, *J*=5.7, 1H; OH), 4.54 (dd, *J*=6.5, 4.1, 1H; HC3'), 4.36 (dt, *J*=6.9, 3.9, 1H; HC4'), 4.26 (q, *J*=5.2, 4.6, 1H; HC2'), 3.95 (dd, *J*=12.8, 3.9, 1H; HC5'), 3.83 (dd, *J*=12.7, 3.9, 1H; HC5'), 1.49 (s, 3H; CH_3), 1.35 (s, 3H; CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 151.91 (C4), 151.83 (C2), 149.09 (C6), 147.39 (C8), 130.39 (C5), 109.31 (C_q), 79.62 (C1'), 72.90 (C3'), 71.14 (C4'), 64.29 (C2'), 63.89 (C5'), 26.40 (CH_3), 25.49 (CH_3). **IR** (cm^{-1}): $\tilde{\nu}$ = 3134 (br, w), 2984 (w), 2900 (w), 1591 (m), 1564 (m), 1381 (m), 1339 (m), 1204 (vs), 1041 (vs), 950 (s), 844 (s), 829 (m); **HRMS** (ESI+): calc. for: $[C_{13}H_{16}ClN_4O_4]^+$ 327.0855, found: 327.0854 $[M+H]^+$.

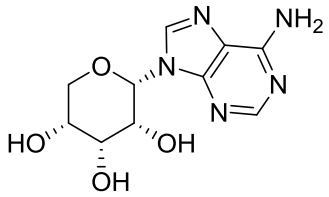
6-Amino-9-(3',4'-*O*-isopropylidene-*a*-D-ribopyranosyl)-9*H*-purine (22)



Nucleoside **20** (30 mg, 0.1 mmol, 1 eq.) was dissolved in EtOH saturated with ammonia (5 mL) and stirred in a sealed tube for 24 h at 65 °C. The solvent was evaporated *in vacuo* and the product purified by column chromatography ($CHCl_3$/MeOH = 100:1 → 20:1) to yield **22** (27 mg, 0.09 mmol, 90%) as a white solid.

mp: 219 °C; **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 8.44 (s, 1H; HC8), 8.15 (s, 1H; HC2), 7.29 (s, 2H, NH_2), 5.97 (d, *J*=4.0, 1H; HC1'), 5.57 (d, *J*=5.8, 1H; OH), 4.49 (dd, *J*=6.3, 4.3, 1H; HC3'), 4.30 (td, *J*=7.1, 3.7, 1H; HC4'), 4.19 – 4.09 (m, 1H; HC2'), 3.90 (dd, *J*=12.8, 4.0, 1H; HC5'), 3.83 (dd, *J*=12.8, 3.9, 1H; HC5'), 1.50 (s, 3H; CH_3), 1.34 (s, 3H; CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 155.97 (C6), 152.58 (C2), 149.51 (C4), 140.97 (C8), 117.90 (C5), 109.19 (C_q), 78.84 (C1'), 73.12 (C3'), 71.07 (C4'), 64.78 (C2'), 63.77 (C5'), 26.40 (CH_3), 25.66 (CH_3). **IR** (cm^{-1}): $\tilde{\nu}$ = 3317 (w), 3208 (m), 3143 (m), 2992 (w), 2901 (w), 1644 (vs), 1610 (vs), 1576 (m), 1383 (m), 1338 (m), 1241 (s), 1215 (s), 1048 (vs), 846 (s), 753 (s); **HRMS** (ESI+): calc. for: $[C_{13}H_{18}N_5O_4]^+$ 308.1353, found: 308.1351 $[M+H]^+$.

9-α-D-ribopyranosyl-adenosine (α-pA, 23)



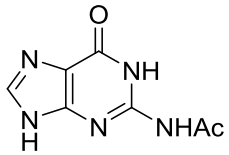
The isopropylidene-protected nucleoside **22** (17 mg, 0.055 mmol, 1 eq.) was dissolved in AcOH/H_2O (7:3) and heated to 60 °C for 5 h. The solvent was evaporated and co-evaporated with EtOH (5 x). The crude product was purified by HPLC to yield α-pA **23** (9 mg, 0,037 mmol, 61%) as a white powder.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.37 (s, 1H; HC8), 8.15 (s, 1H; HC2), 7.31 (s, 2H; NH_2), 5.76 (s, 1H; HC1'), 5.36 (br, 3H; 3 OH), 3.93 (dd, *J*=12.4, 2.6, 1H; HbC5'), 3.89 – 3.86 (m, 1H; HC2'), 3.84 (t, *J*=3.1, 1H; HC3'), 3.81 – 3.74 (m, 2H; HC4' and HaC5'). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 156.00 (C6), 152.57 (C2), 148.86 (C4), 139.79 (C8), 117.84 (C5), 81.14 (C1'), 71.48 (C2'), 69.55 (C5'), 68.53 (C4'), 67.52 (C3'). **HRMS** (ESI+): calc. for: $[C_{10}H_{14}N_5O_4]^+$ 268.1040, found: 268.1039 $[M+H]^+$.

Synthesis of Guanosine nucleoside

β-guanosine was bought from Sigma-Aldrich β-pyranosyl-guanosine was synthesized by the following procedures.

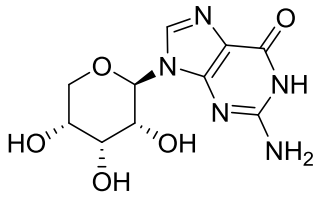
***N*-2-acetylguanine (24)**



Guanine (2 g, 13.2 mmol, 1.0 eq.) was suspended in dimethylacetamide (15 mL) together with Ac_2O (3.3 mL, 35 mmol, 2.6 eq.) and stirred for 20 h at 160 °C. After cooling to room temperature a precipitate was formed that was filtered off and washed with EtOH. The crude product was suspended in EtOH/H_2O (1:1, 15 mL) and heated under reflux for 2 h. The solvent was evaporated *in vacuo* to yield a light brown solid. The solvent was removed and the remaining solid dried to yield the protected base **24** (2.1 g, 10.8 mmol, 81%) as light brown solid.

mp: 324 °C; **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 12.03 (s, 1H; NH), 11.56 (s, 1H; NH), 8.03 (s, 1H; HC8), 2.16 (s, 3H; CH_3); **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 173.37 (NHCO), 23.79 (CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 3058 (w), 1682 (s), 1613 (s) und 1380 (m); **HRMS** (ESI+): calc. for: $[C_7H_8N_5O_2]^+$ 194.0673, found: 194.0673 $[M+H]^+$.

9-*β*-D-ribopyranosyl-guanosine (β-pG, 25)

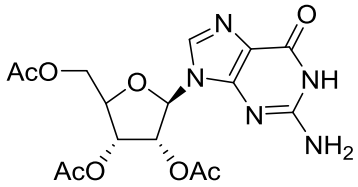


Tetraacetyl-D-ribopyranose (*31*) (198 mg, 0.522 mmol, 1 eq.) and *N*-2-acetylguanine **24** (100 mg, 0.518 mmol, 1 eq.) were suspended in dry MeCN (3 mL) under inert atmosphere. Bis(trimethylsilyl)acetamide (0.25 mL, 1.035 mmol, 2.00 eq) was added and the reaction mixture was heated to 60 °C. After the suspension turned into a clear solution TMSOTf (0.24 mL, 1.295 mmol, 2.50 eq) was added. After stirring at 60 °C for 2 h ethyl acetate was added and the reaction mixture was extracted three times with $NaHCO_3$ (sat.), dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude mixture was dissolved in $NH_{3\ (l)}$ in a sealed tube and stirred at 25 °C for 41 h. The crude mixture was purified *via* HPLC to obtain β-pG **25** (47 mg, 32%) as white solid.

mp.: 230 °C (decomp.); **^{1}H-NMR** (600 MHz, DMSO-d_6): δ = 10.59 (s, 1H; NH), 7.82 (s, 1H; C8), 6.49 (s, 2H; NH_2), 5.45 (d, *J*=9.4 Hz, 1H; HC1ˈ), 5.04 – 5.02 (m, 2H; 2 OH), 4.85 (d, *J*=6.30, 1H; OH), 4.04 – 4.00 (m, 2H; HC2ˈ, HC3ˈ), 3.69 – 3.67 (m, 1H; HC4ˈ), and 3.63 – 3.52 ppm (m, 2H; H_2C5ˈ); **^{13}C-NMR** (101 MHz, DMSO-d_6): δ = 157.2 (C6), 154.1 (C2), 152.3 (C4), 136.1 (C8), 116.8 (C5), 79.3 (C1ˈ), 71.7 (C3ˈ),68.7 (C2ˈ), 67.0 (C4ˈ), and 65.5 ppm (C5ˈ); **IR** (cm^{-1}): $\tilde{\nu}$ = 3337 (br, m), 2931 (w), 1635 (s), 1590 (s), 1489 (m), 1436 (m), 1327 (m), 1187 (m), 1132 (m), 1072 (s), 1037 (s), 954 (w), 919 (m), 884 (w), 839 (w), 787 (w); **HRMS** (ESI+): calc. for: $[C_{10}H_{14}N_5O_4]^+$ 268.1040, found: 268.1039 $[M+H]^+$

Synthesis of 2,6-diaminopurine nucleosides

9-(2',3',5'-tri-*O*-acetyl-*ß*-D-ribofuranosyl)-9*H*-guanine (26)

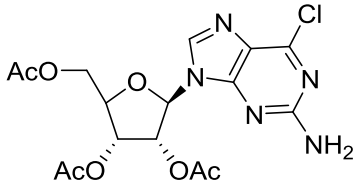


Guanosine (1 g, 3.53 mmol, 1 eq.) was dissolved in Ac_2O (50 mL) together with DMAP (4 mg, 0.04 mmol, 0.01 eq.). The mixture was stirred for 48 h at 40 °C. The solvent was evaporated *in vacuo* to yield the protected guanosine **26** (1.37 g, 95%) as colorless solid without further purification.

^{1}H-NMR (400 MHz, DMSO-d_6): δ = 10.74 (s, 1H; NH), 7.92 (s, 1H; HC8), 6.53 (s, 2H; NH_2), 5.98 (d, *J*=6.04, 1H; HC1ˈ), 5.80 – 5.78 (m, 1H; HC2ˈ), 5.49 (dd, *J*=6.0, 4.1, 1H; HC3ˈ), 4.39 – 4.23 (m, 3H; HC4ˈ, C5ˈH2), 2.11 (s, 3H; CH_3), 2.04 (s, 3H; CH_3), 2.03 ppm (s, 3H; CH_3); **^{13}C-NMR** (101 MHz, DMSO-d_6): δ=170.5 (CO), 169.9 (CO), 169.7 (CO), 157.5 (C2), 154.3 (C6), 151.5 (C4), 136.1 (C8); 117.2 (C5), 84.8 (C1ˈ), 80.0 (C4ˈ),

72.5 (C2'), 70.7 (C3'), 63.5 (C5'), 21.0 (CH_3), 20.8 (CH_3), 20.6 ppm (CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 1746 (s, CO), 1699 (s, NH_2), 1218 (s, CH_2) und 1072 (m, NH_2); **HRMS** (ESI+): calc. for: $[C_{16}H_{20}N_5O_8]^+$ 410.1306, found: 410.1308 $[M+H]^+$.

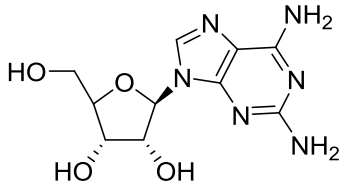
6-chloro-9-(2',3',5'-tri-*O*-acetyl-*ß*-D-ribofuranosyl)-9*H*-guanine (27)



The protected guanosine **26** (1 g, 2.4 mmol, 1eq.) was dissolved in MeCN (5 mL) together with tetraethylammoniumchloride (810 mg, 4.9 mmol, 2 eq.), N,N-dimethylaniline (0.31 mL, 2.4 mmol, 1 eq.) and phosphoryl chloride (1.34 mL, 14.7 mmol, 6 eq.). It was stirred for 10 min at 100 °C. The solvent was evaporated in vacuo and the residue taken up in H_2O and extracted with $CHCl_3$ (2 x). The combined organic layers were dried over $MgSO_4$ and the solvent evaporated *in vacuo*. The product was purified by column chromatography (DCM/MeOH = 50:1) to yield the chlorinated nucleoside **27** (912 mg, 87%) as a pale yellow foam.

mp: 76 °C; **^{1}H-NMR** (600 MHz, Chloroform-*d*): δ = 7.88 (s, 1H; HC8), 5.99 (d, *J*=5.0, 1H; HC1'), 5.93 – 5.92 (m, 1H; HC2'), 5.71 (t, *J*=5.1, 1H; HC3'), 4.44 – 4.33 (m, 3H; HC4', H_2C5'), 2.12 (s, 3H; CH_3), 2.08 (s, 3H; CH_3), 2.06 ppm (s, 3H; CH_3); **^{13}C-NMR** (151 MHz, Chloroform-*d*): δ=170.4 (CO), 169.5 (CO), 169.3 (CO), 160.0 (C2), 153.0 (C4), 151.7 (C6), 140.7 (C8); 125.6 (C5), 86.6 (C1'), 80.0 (C4'), 72.7 (C2'), 70.4 (C3'), 62.9 (C5'), 20.6 (CH_3), 20.5 (CH_3), 20.4 ppm (CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 1742 (s, CO), 1611 (s, NH_2), 1214 (s, CH_2) und 1045 (m, NH2); **HRMS** (ESI+): calc. for: $[C_{16}H_{19}ClN_5O_7]^+$ 428.0968, found: 428.0970 $[M+H]^+$.

9-*ß*-D-ribofuranosyl-2,6-diaminopurine (β-fDA, 28)

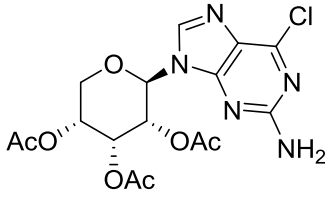


Compound **27** (50 mg, 0.12 mmol) was dissolved in $NH_{3\ (l)}$ (ca. 3 mL) and stirred at room temperature for 48 h in a sealed tube. After evaporation of the solvent, the crude mixture was purified by HPLC to obtain β-fDA **28** (22 mg, 67%) as a white solid.

mp: 233 °C**; ^{1}H-NMR** (400 MHz, DMSO-d_6): δ = 7.91 (s, 1H; HC8), 6.76 (s, 2H; NH2), 5.71 (s, 2H; NH2), 5.70 (d, *J*=6.32, 1H; HC1'), 4.50 (dd, *J*=6.3, 5.1, 1H; HC2'), 4.09 (dd, *J*=5.0, 2.9, 1H; HC3'), 3.91 – 3.88 (m, 1H; HC4'), 3.63 (dd, *J*=12.1, 3.6, 1H; HaC5'), 3.53 ppm (dd, *J*=12.0, 3.5, 1H; HbC5'); **^{13}C-NMR** (101 MHz, DMSO-d_6): δ=160.5 (C2), 156.7 (C4), 151.9 (C6), 136.6 (C8); 114.0 (C5), 87.4 (C1'), 85.9 (C4'), 73.6 (C2'), 71.1

(C3ˈ), 62.2 ppm (C5ˈ); **IR** (cm^{-1}): $\tilde{\nu}$ = 3175 (m), 1653 (s), 1596 (s), 1111 (m) and 1059 (s); **HRMS** (ESI+): calc. for: $[C_{10}H_{15}N_6O_4]^+$ 283.1149, found: 283.1150 $[M+H]^+$.

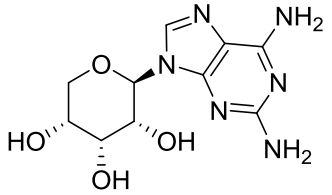
6-chloro-9-(2',3',4'-tri-*O*-acetyl-ß-D-ribopyranosyl)-9*H*-guanine (29)



Tetraacetyl-D-ribopyranose (*31*) (420 mg, 1.3 mmol, 0.8 eq.) and 2-amino-6-chloroguanine (300 mg, 1.8 mmol, 1 eq.) were suspended in MeCN (6 mL) under inert atmosphere. *N*,*O*-Bis-(trimethylsilyl)-acetamid (1.1 mL, 4.4 mmol, 2.5 eq.) was added and the mixture stirred for 0.5 h at 60 °C until everything was dissolved. TMSOTf (0.96 mL, 5.3 mmol, 3 eq.) was added. After 2 h tetraacetyl-D-ribopyranose (*31*) (420 mg, 1.3 mmol, 0.8 eq.) was added and further stirred for 19 h at 60 °C. The reaction mixture was taken up in EtOAc and washed with sat. $NaHCO_3$ solution. The combined H_2O phases were extracted with EtOAc once. The combined organic layers were dried over $MgSO_4$ and the solvent evaporated *in vacuo*. The crude product was purified by column chromatography (EtOAc/*i*Hex = 1:2 → 5:1) to yield the nucleoside **29** (332 mg, 44%) as pale yellow solid.

mp: 102 °C; **^{1}H-NMR** (400 MHz, DMSO-d_6): δ = 8.41 (s, 1H; HC8), 5.95 (dd, *J* =9.6, 3.0, 1H; HC2ˈ), 5.83 (d, *J*=9.7, 1H; HC1ˈ), 5.76 – 5.74 (m, 1H; HC3ˈ), 5.10 – 5.15 (m, 1H; HC4ˈ), 4.04 (dd, *J*=11.0, 4.8, 1H; HaC5ˈ), 3.96 – 3.90 (m, 1H; HbC5ˈ), 2.23 (s, 3H; CH_3), 2.01 (s, 3H; CH_3), 1.80 ppm (s, 3H; CH_3); **^{13}C-NMR** (101 MHz, DMSO-d_6): δ = 170.3 (CO), 169.7 (CO), 169.2 (CO), 160.4 (C2), 154.4 (C4), 150.2 (C6), 141.4 (C8), 123.5 (C5), 78.0 (C1ˈ), 67.8 (C3ˈ), 67.6 (C2ˈ), 66.0 (C4ˈ), 63.1 (C5ˈ), 21.1 (CH_3), 20.9 (CH3), 20.5 ppm (CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 3441 (w), 3210 (w), 1753 (m), 1735 (m), 1612 (m), 1377 (w), 1256 (s), 1230 (s), 1170 (s), 1070 (m), 1035 (vs), 975 (m); **HRMS** (ESI+): calc. for: $[C_{16}H_{19}ClN_5O_7]^+$ 428.0968, found: 428.0966 $[M+H]^+$.

9-ß-D-ribopyranosyl-2,6-diaminopurine (β-pDA, 30)



The nucleoside **29** (27 mg, 0.06 mmol) was dissolved in $NH_{3\ (l)}$ (ca. 3 mL) and stirred at room temperature for 48 h in a sealed tube. After evaporation of the solvent, the crude mixture was purified by HPLC to obtain the product **30** (11 mg, 68%) as a white solid.

mp: 241 °C; **^{1}H-NMR** (400 MHz, Water-d_2): δ = 8.09 (s, 1H; HC8), 5.62 (d, J=9.04, 1H; HC1ˈ), 4.34 – 4.31 (m, 2H; HC2ˈ and HC3ˈ) und 4.07 – 4.02 ppm (m, 1H; HC4ˈ), 3.91 – 3.81 (m, 2H; C5ˈH2); **^{13}C-NMR** (101 MHz, Water-d_2): δ = 153.6 (C2), 151.1 (C4), 151.0

(C6), 140.1 (C8), 111.7 (C5), 80.1 (C1'), 80.7 (C3'), 68.1 (C2'), 65.9 (C4'), 64.8 ppm (C5'); **IR** (cm^{-1}): $\tilde{\nu}$ = 3322 (br), 1688 (m), 1650 (s) und 1039 (s); **HRMS** (ESI+): calc. for: $[C_{10}H_{15}N_6O_4]^+$ 283.1149, found: 283.1148 $[M+H]^+$.

Pebiotic nucleoside formation procedure

Nucleoside formation using ribose and FaPy (11 - 13)

Ribose (56.5 mg, 0.375 mmol, 15 eq.) was thoroughly ground up with the corresponding FaPy compound **11 - 13** (0.025 mmol, 1 eq.) and heated to 100 °C for 8 h in an oven. The remaining solid was taken up in the corresponding basic solution (3 mL) (see text) and heated in a sealed tube (*ACE* 15 mL pressure tube) at 100 °C for several days (1-14 d, see text). At different time intervals 100 µl samples were removed and diluted to 1 mL. The samples were used for LC-MS analysis.

Nucleoside formation using prebiotic sugar synthesis with glycolaldehyde (14) and glyceraldehyde (15)

The synthesis of the sugar was carried out according to Harsch et al. (*33*).

Glycolaldehyde (20 mg, 0.33 mmol, 1 eq.) and glyceraldehyde (30 mg, 0.33 mmol, 1 eq.) was reacted in $Ca(OH)_2$ solution (20 mL, 23 mM) at 23 °C for 0.5 h. The solution was neutralized to pH 7 with H_2SO_4 (0.5 M). FaPyA (3.8 mg, 0.025 mmol, 0.075 eq.) was added. Because of the large amount of solvent the mixture was concentrated by freeze drying to about 2 mL. The insoluble $CaSO_4$ was removed by filtration and the remaining solvent removed by freeze drying again. The remaining solid was heated for 8 h at 100 °C in an oven. The samples were taken up in water and divided equally. To each sample was either added NH_3 to give a final concentration of 0.5 M or borax to give a final concentration of 0.125 M (final volume for each sample 1.5 mL). Both samples were heated in a sealed Eppendorf tube at 100 °C for 2-3 d. 100 µl were removed and diluted to 1 mL for LC-MS analysis according to the general procedure.

Nucleoside formation using Formose-"type" reaction with Formaldehyde and glycolaldehyde (14)

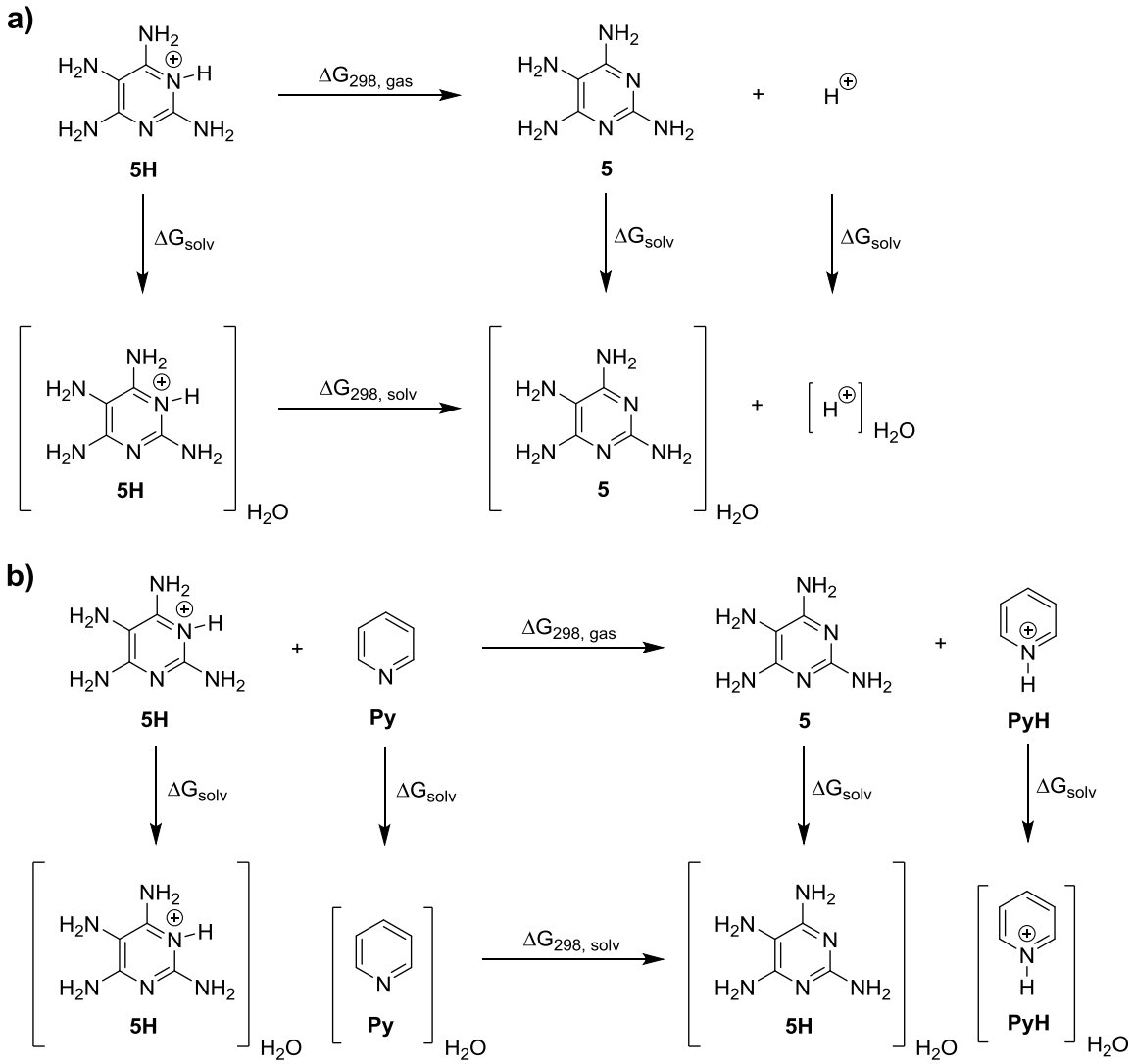
A methanol-free 16% formaldehyde solution in H_2O (39 µl, 0.21 mmol, 1 eq.) was reacted with glycolaldehyde (24 mg, 0.40 mmol, 2 eq.) in 1 mL of NH_3 solution (0.5 M) for 1 h at 65 °C in an *Eppendorf* tube (1.5 mL). Afterwards FaPyA (20 mg, 0.13 mmol, 0.6 eq.) was dissolved and the reaction solution was placed on a small petri dish and reacted for 8 h at 100°C in an oven. The remaining solid was taken up in 3 mL of NH_3 solution (0.5 M) and stirred for several days (usually 3-7 d) in a sealed tube (*ACE* 15 mL pressure tube) at 100 °C. At different time intervals 100 µl samples were removed from

the reaction mixture and diluted to exactly 25 mL with a volumetric flask. The diluted samples were used for LC/MS analysis according to the general procedure.
For reactions with higher concentrations we used twice the amounts of formaldehyde (78 µl, 0.42 mmol, 1 eq.) and glycolaldehyde (48 mg, 0.80 mmol, 2 eq.) in only 0.2 mL of NH_3 solution (0.5 M). The remaining procedure was kept the same.

Supplementary Text

Calculations of pK_a for aminopyrimidines

The highly basic nature of **5** implies that it will mainly exist in its *N*-protonated form in aqueous solution. The respective pK_a value can be calculated using the thermochemical cycle shown in Scheme 1a, where the free energy for protonation in aqueous solution $\Delta G_{298,solv}$ is obtained as the sum of the respective gas phase free energy and differences in solvation free energy of all components: $\Delta G_{298,sol} = \Delta G_{298,gas} - \Delta G_{solv}(\mathbf{5H}) + \Delta G_{solv}(\mathbf{5}) + \Delta G_{solv}(H^+)$. The acidity of pyrimidinium ion **5H** is then obtained according to pK_a(**5H**) = $\Delta G_{298,solv}/2.303RT$. With R = 8.31451 $JK^{-1}mol^{-1}$, each pK_a unit amounts to a free energy difference of 5.709 kJ/mol at 298.15 K.
Using the G3(MP2)B3(+) value for the gas phase basicity of **5** of $\Delta G_{298,gas}$ = +951.1 kJ/mol, the experimental free energy of solvation of the proton of $\Delta G_{solv}(H^+)$ = -1104.5 kJ/mol (*34*), corrections of 7.925 kJ/mol for all species for moving from the gas to the solution phase standard state of 1 mol/l at 298.15 K, and using solvation free energies for **5** and **5H** obtained from single point calculations at SMD/B3LYP/6-31+G(d) level (*35*) yields $\Delta G_{298,sol}$ = +43.74 kJ/mol and therefore pK_a(**5H**) = +7.7. In terms of its basicity aminopyrimidine **5** thus ranges between pyridine (pK_a(**Py**) = +5.2) and 4-aminopyridine (pK_a = +9.1) (*36*).



Scheme S1. (a) Protonation of **5** in aqueous solution; and (b) proton transfer between aminopyrimidine **5** and pyridinium ion **PyH** in aqueous solution.

The basicity of aminopyrimidine **5** can also be assessed by direct comparison to that of pyridine (**Py**) as graphically shown in Scheme 1b. The free energy for proton exchange with pyridinium ion **PyH** is directly proportional to the difference in pK_a values of the two bases according to ΔpK_a(**5H-PyH**) = $\Delta G_{298,sol}$/2.303RT. Again, the free energy difference in solution can be calculated as a combination of the gas phase reaction free energy $\Delta G_{298,gas}$ and differences in the respective solvation free energies. Using the same theoretical methods as before we obtain $\Delta G_{298,gas}$ = +52.89 kJ/mol at G3(MP2)B3(+) level for the gas phase proton exchange. Addition of aqueous solvation effects obtained from SMD/B3LYP/6-31+G(d) calculations yields $\Delta G_{298,sol}$ = +12.38 kJ/mol, and thus ΔpK_a(**5H-PyH**) = +2.17. Together with the literature value for pyridine of pK_a(**PyH**) = +5.2 (*36*), this yields pK_a(**5H**) = +7.4. The thermochemical cycles shown in Scheme 1a and 1b thus yield closely similar pK_a values for protonated aminopyrimidine **5H** (+7.7 vs. +7.4), and the recommended value can simply be taken as the arithmetic mean of pK_a(**5H**) = +7.5±0.2.

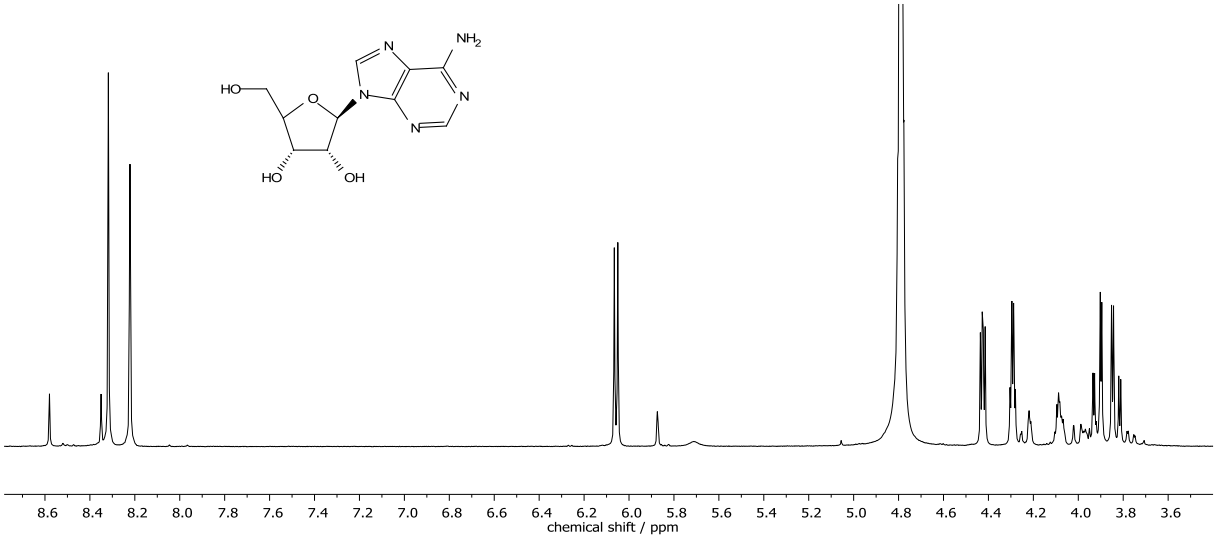


Fig. S1

^{1}H-NMR of β-adenosine purified by HPLC from a prebiotic reaction. The ^{1}H-signal of HC2' overlaps with the solvent peak, confirmed by COSY. **^{1}H-NMR** (400 MHz, Water-d2) δ = 8.32 (s, 1H; HC2), 8.22 (s, 1H; HC8), 6.06 (d, J=6.2, 1H; HC1'), 4.43 (dd, J=5.3, 3.3, 1H; HC3'), 4.29 (q, J=3.2, 1H; HC4'), 3.91 (dd, J=12.9, 2.8, 1H; HC5'), 3.83 (dd, J=12.9, 3.6, 1H; HC5').

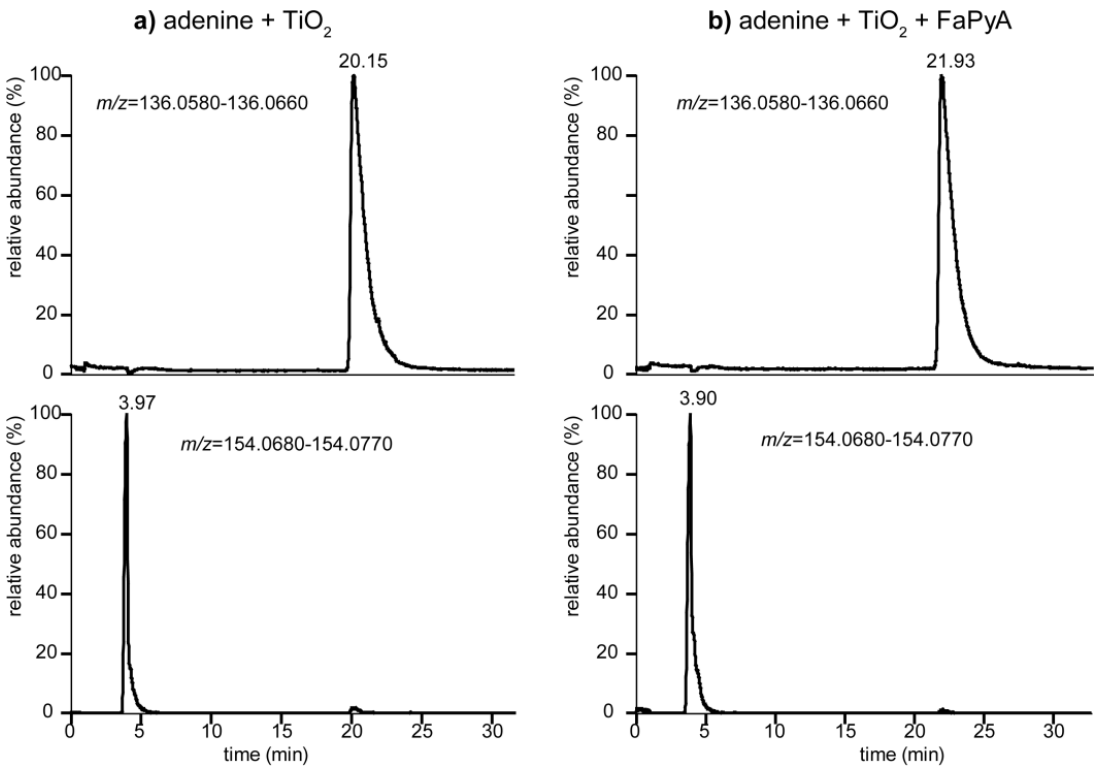


Fig. S2

Formation of FaPyA *via* TiO_2. In a sealed tube adenine (30 mg, 0.222 mmol, 1 eq) and TiO_2 (9 mg, 0.111 mmol, 0.5 eq) were suspended in water (1 mL) and stirred at 120 °C for 20 h. HPLC-MS analysis with co-injection experiments show FaPyA **13** formation. a) MS-spectrum with selected mass-filter for adenine (top, 20.15 min) and FaPyA **13** (bottom, 3.97 min) of the crude reaction mixture. b) MS-spectrum with selected mass-filter for adenine (top, 21.93 min) and FaPyA (bottom, 3.90 min). Same reaction mixture but co-injected with FaPyA, gives still one signal in the corresponding mass-filter (*m/z* 154.068-154.077).

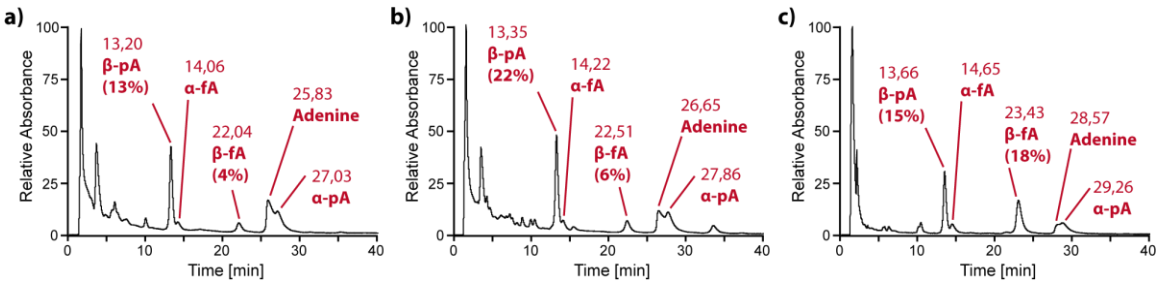


Fig. S3

For all three samples HPLC-MS analysis was conducted according to the general synthesis procedure using a C8 column. Identified UV-signals are labeled and yields are given in brackets. a) HPLC analysis (UV-detection) of the reaction of FaPyA **13** with ribose **2** and subsequent heating in lysine solution (0.5 M) for 4 d. b) HPLC analysis (UV-detection) of the reaction of FaPyA **13** with ribose **2** and subsequent heating in arginine solution (0.5 M) for 4 d. c) HPLC analysis (UV-detection) of the reaction of FaPyA **13** with ribose **2** and subsequent heating in carbonate buffer (1.1 M) containing boric acid (0.14 M) of pH 10.3 for 1 d.

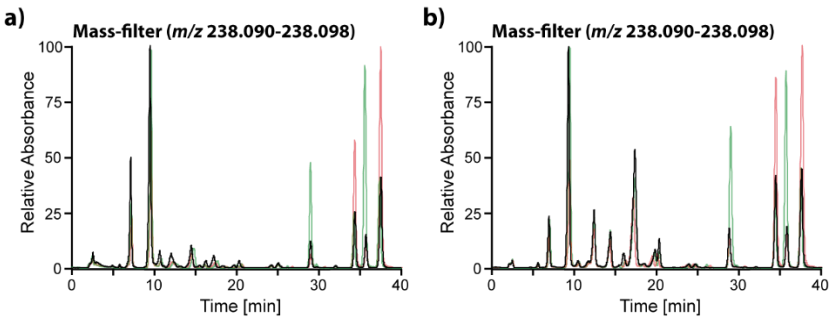


Fig. S4

Co-injection experiments with a reaction of threose + **13** (red) and erythrose + **13** (green). Selection of the MS-spectrum with tetrose mass-filter (m/z 238.090-238.098). A reaction using **14** and **15** (both 16.7 mM) in the presence of $Ca(OH)_2$ (23 mM) and addition of **13**. a) Detection and identification of α/β-tetrosides after heating in NH_3 solution (0.5 M) for 2 d. b) Detection and identification of of α/β-tetrosides after heating the reaction mixture in borax solution (0.125 M) for 2 d.

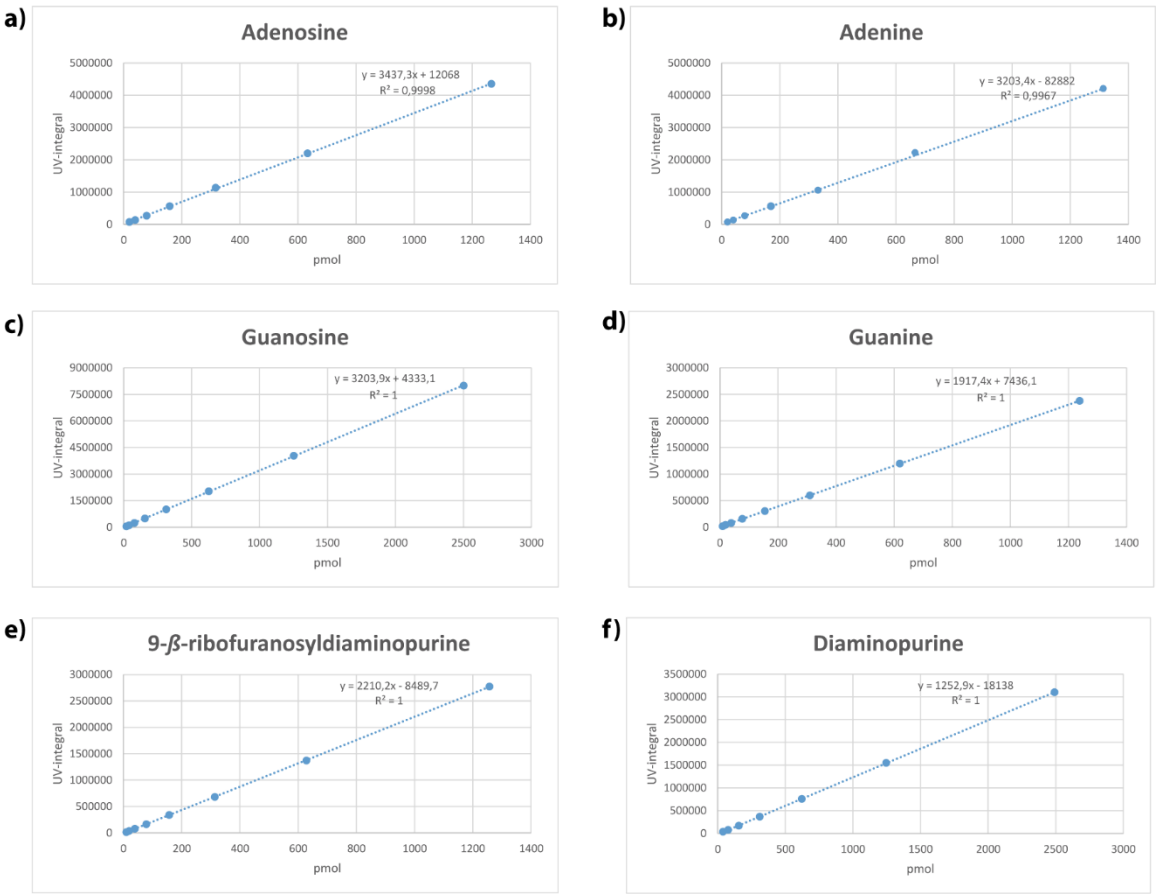


Fig. S5

Calibration curves at 260 nm for quantification of nucleosides from UV-signals. For each *N*-9-nucleoside we assumed the same extinction coefficient for all the four expected isomers.

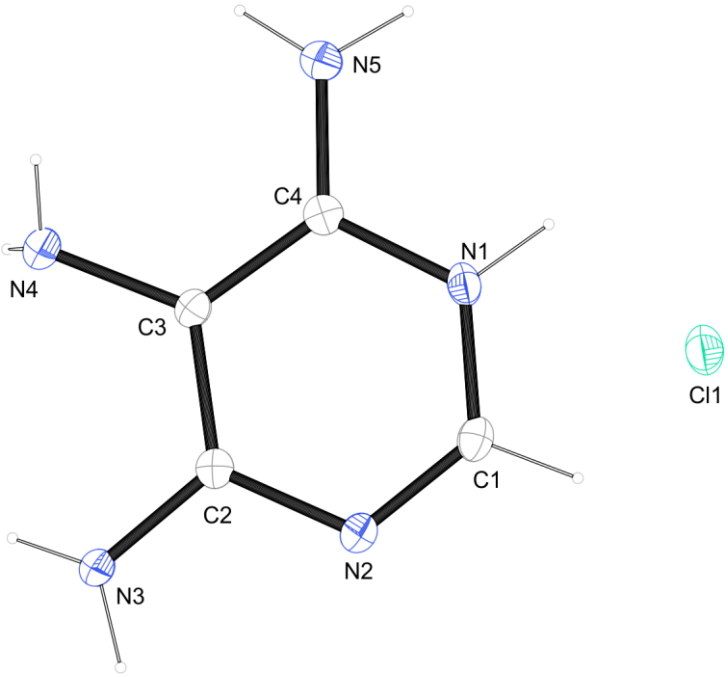


Fig. S6

Crystal structure of monoprotonated 4,5,6-triaminopyrimidine **8** (CCDC 1416691). Chloride ion (light green), nitrogen atoms (blue) and carbon atoms (white) are represented by large spherical structures. Hydrogen atoms (white) are represented by small spherical structures.

Table S1
Crystallographic data for 4,5,6-triaminopyrimidine hydrochloride (**8**)

CCDC	1416691
net formula	$C_4H_8ClN_5$
M_r/g mol^{-1}	161.60
crystal size/mm	0.266 × 0.183 × 0.182
T/K	173(2)
radiation	MoKα
diffractometer	'Oxford XCalibur'
crystal system	monoclinic
space group	'P 21/n'
a/Å	4.4527(3)
b/Å	16.7106(9)
c/Å	9.4509(6)
α/°	90
β/°	100.377(6)
γ/°	90
V/Å^3	691.72(8)
Z	4
calc. density/g cm^{-3}	1.552
μ/mm^{-1}	0.478
absorption correction	'multi-scan'
transmission factor range	0.98120–1.00000
refls. measured	3678
R_{int}	0.0367
mean σ(I)/I	0.0471
θ range	4.265–26.368
observed refls.	1104
x, y (weighting scheme)	0.0347, 0.2784
hydrogen refinement	refall
refls in refinement	1412
parameters	123
restraints	0
$R(F_{obs})$	0.0380
$R_w(F^2)$	0.0947
S	1.075
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.269
min electron density/e Å^{-3}	−0.334

Table S2
Bond distances, bond angles and torsion angles of monoprotonated 4,5,6-triaminopyrimidine **8**.

Bond distances (in Å):	Bond angles (in °):	Torsion angles (in °):
C1 N2 1.295(3)	N2 C1 N1 124.2(2)	N3 C2 C3 C4 -179.3(2)
C1 N1 1.345(3)	N2 C1 H1A 120.4(14)	N2 C2 C3 C4 2.1(3)
C1 H1A 0.95(2)	N1 C1 H1A 115.3(13)	N3 C2 C3 N4 2.5(3)
C2 N3 1.333(3)	N3 C2 N2 115.8(2)	N2 C2 C3 N4 -176.12(19)
C2 N2 1.376(3)	N3 C2 C3 122.0(2)	C2 C3 C4 N5 179.8(2)
C2 C3 1.400(3)	N2 C2 C3 122.2(2)	N4 C3 C4 N5 -2.0(3)
C3 C4 1.390(3)	C4 C3 C2 117.7(2)	C2 C3 C4 N1 -1.3(3)
C3 N4 1.410(3)	C4 C3 N4 122.8(2)	N4 C3 C4 N1 176.9(2)
C4 N5 1.332(3)	C2 C3 N4 119.51(19)	N2 C1 N1 C4 0.3(3)
C4 N1 1.368(3)	N5 C4 N1 117.5(2)	N5 C4 N1 C1 179.2(2)
N1 H1 0.90(3)	N5 C4 C3 124.6(2)	C3 C4 N1 C1 0.2(3)
N3 H31 0.89(3)	N1 C4 C3 117.8(2)	N1 C1 N2 C2 0.5(3)
N3 H32 0.81(3)	C1 N1 C4 121.0(2)	N3 C2 N2 C1 179.6(2)
N4 H41 0.83(3)	C1 N1 H1 121.0(17)	C3 C2 N2 C1 -1.7(3)
N4 H42 0.83(3)	C4 N1 H1 118.0(17)	
N5 H51 0.84(3)	C1 N2 C2 117.08(19)	
N5 H52 0.86(3)	C2 N3 H31 117.0(16)	
	C2 N3 H32 119.6(18)	
	H31 N3 H32 121(2)	
	C3 N4 H41 113(2)	
	C3 N4 H42 112(2)	
	H41 N4 H42 108(3)	
	C4 N5 H51 122.6(18)	
	C4 N5 H52 118.8(19)	
	H51 N5 H52 118(3)	

Table S3
Hydrogen bond geometry (D donor, H hydrogen, A acceptor, D-H donor-hydrogen distance in Å, H…A hydrogen-acceptor distance in Å, D…A donor-acceptor distance in Å, D-H…A angle in °, symm(A) symmetry code of acceptor):

D	H	A	D-H	H...A	D...A	D-H.A	symm(A)
N1	H1	Cl1	0.90(3)	2.21(3)	3.087(2)	164(2)	x-1, y, z
N3	H31	N2	0.89(3)	2.10(3)	2.989(3)	174(2)	1-x, 1-y, -z
N3	H32	Cl1	0.81(3)	2.58(3)	3.349(2)	161(2)	x-1, y, z-1
N4	H41	N3	0.83(3)	2.65(3)	3.336(3)	141(3)	x-1, y, z
N5	H51	Cl1	0.84(3)	2.48(3)	3.298(2)	165(2)	x-3/2,1/2-y,z-1/2
N5	H52	N4	0.86(3)	2.48(3)	3.095(3)	129(2)	x-1/2,1/2-y,1/2+z
N5	H52	Cl1	0.86(3)	2.62(3)	3.383(3)	147(2)	x-1, y, z

References

1. F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Krüger, L. Le Roy, A. J. MacDermott, S. McKenna-Lawlor, U. J. Meierhenrich, G. M. Muñoz Caro, F. Raulin, R. Roll, A. Steele, H. Steininger, R. Sternberg, C. Szopa, W. Thiemann, S. Ulamec, Organic compounds on comet 67P/Churyumov-Gerasimenko revealed by COSAC mass spectrometry. *Science* **349**, aab0689 (2015). Medline doi:10.1126/science.aab0689

2. J. F. Kasting, Earth's early atmosphere. *Science* **259**, 920–926 (1993). Medline doi:10.1126/science.11536547

3. F. H. C. Crick, The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968). Medline doi:10.1016/0022-2836(68)90392-6

4. L. E. Orgel, Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004). Medline doi:10.1080/10409230490460765

5. M. W. Powner, B. Gerland, J. D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009). Medline doi:10.1038/nature08013

6. H. B. White 3rd, Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976). Medline doi:10.1007/BF01732468

7. W. D. Fuller, R. A. Sanchez, L. E. Orgel, Studies in prebiotic synthesis: VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67**, 25–33 (1972). Medline doi:10.1016/0022-2836(72)90383-X

8. W. D. Fuller, R. A. Sanchez, L. E. Orgel, Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides. *J. Mol. Evol.* **1**, 249–257 (1972). Medline doi:10.1007/BF01660244

9. J. Oró, Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* **2**, 407–412 (1960). doi:10.1016/0006-291X(60)90138-8

10. J. Oró, A. P. Kimball, Synthesis of purines under possible primitive Earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys.* **94**, 217–227 (1961). Medline doi:10.1016/0003-9861(61)90033-9

11. U. P. Trinks, A. Eschenmoser, thesis, ETH Zurich (1987), ETH 2-collection http://dx.doi.org/10.3929/ethz-a-000413538.

12. K. E. Koch, A. Eschenmoser, thesis, ETH Zurich (1992), ETH e-collection, http://dx.doi.org/10.3929/ethz-a-000694124.

13. W. Traube, Der Aufbau der Xanthinbasen aus der Cyanessigsäure. Synthese des Hypoxanthins und Adenins. *Liebigs Ann. Chem.* **331**, 64–88 (1904). doi:10.1002/jlac.19043310106

14. J. Clark, R. K. Grantham, J. Lydiate, Heterocyclic studies. Part V. Desulphuration of heterocyclic thiols with nickel boride. *J. Chem. Soc. C. Org.* **1968**, 1122 (1968). doi:10.1039/j39680001122

15. M. P. Robertson, M. Levy, S. L. Miller, Prebiotic synthesis of diaminopyrimidine and thiocytosine. *J. Mol. Evol.* **43**, 543–550 (1996). Medline doi:10.1007/BF02202102

16. R. Saladino, U. Ciambecchini, C. Crestini, G. Costanzo, R. Negri, E. Di Mauro, One-pot TiO_2-catalyzed synthesis of nucleic bases and acyclonucleosides from formamide: Implications

for the origin of life. *Chem. Bio. Chem.* **4**, 514–521 (2003). Medline doi:10.1002/cbic.200300567

17. T. Mori, K. Tano, K. Takimoto, H. Utsumi, Formation of 8-hydroxyguanine and 2,6-diamino-4-hydroxy-5-formamidopyrimidine in DNA by riboflavin mediated photosensitization. *Biochem. Biophys. Res. Commun.* **242**, 98–101 (1998). Medline doi:10.1006/bbrc.1997.7916

18. A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, Borate minerals stabilize ribose. *Science* **303**, 196 (2004). Medline doi:10.1126/science.1092464

19. H. J. Kim, A. Ricardo, H. I. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye, S. A. Benner, Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J. Am. Chem. Soc.* **133**, 9457–9468 (2011). Medline doi:10.1021/ja201769f

20. J. B. Lambert, S. A. Gurusamy-Thangavelu, K. Ma, The silicate-mediated formose reaction: Bottom-up synthesis of sugar silicates. *Science* **327**, 984–986 (2010). Medline doi:10.1126/science.1182669

21. R. B. Trattner, G. B. Elion, G. H. Hitchings, D. M. Sharefkin, Deamination studies on pyrimidine and condensed pyrimidine systems. *J. Org. Chem.* **29**, 2674–2677 (1964). doi:10.1021/jo01032a047

22. M. D. Kirnos, I. Y. Khudyakov, N. I. Alexandrushkina, B. F. Vanyushin, 2-aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature* **270**, 369–370 (1977). Medline doi:10.1038/270369a0

23. W. Löb, Über das Verhalten des Formamids unter der Wirkung der stillen Entladung Ein Beitrag zur Frage der Stickstoff-Assimilation. *Berichte der Dtsch. Chem. Gesellschaft* **46**, 684–697 (1913). doi:10.1002/cber.19130460193

24. D. J. Ritson, J. D. Sutherland, Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew. Chem. Int. Ed.* **52**, 5845–5847 (2013). Medline doi:10.1002/anie.201300321

25. G. Harsch, H. Bauer, W. Voelter, Kinetik, Katalyse und Mechanismus der Sekundärreaktion in der Schlußphase der Formose-Reaktion. *Liebigs Ann. Chem.* **1984**, 623–635 (1984). doi:10.1002/jlac.198419840402

26. A. Altomare, M. C. Burla, M. Camalli, G. L. Cascarano, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, G. Polidori, R. Spagna, SIR 97: A new tool for crystal structure determination and refinement. *J. Appl. Cryst.* **32**, 115–119 (1999). doi:10.1107/S0021889898007717

27. G. M. Sheldrick, A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008). Medline doi:10.1107/S0108767307043930

28. C. A. C. Haley, P. Maitland, 697. Organic reactions in aqueous solution at room temperature. Part I. The influence of pH on condensations involving the linking of carbon to nitrogen and of carbon to carbon. *J. Chem. Soc.* 3155 (1951). doi:10.1039/jr9510003155

29. A. Bendich, J. F. Tinker, G. B. Brown, A synthesis of isoguanine labeled with isotopic nitrogen. *J. Am. Chem. Soc.* **70**, 3109–3113 (1948). Medline doi:10.1021/ja01189a081

30. H. Vorbrüggen, K. Krolikiewicz, B. Bennua, Nucleoside syntheses, XXII[1)] Nucleoside synthesis with trimethylsilyl triflate and perchlorate as catalysts. *Chem. Ber.* **114**, 1234–1255 (1981). doi:10.1002/cber.19811140404

31. H. Zinner, Die Acetate der d-Ribose. *Chem. Ber.* **86**, 817–824 (1953). doi:10.1002/cber.19530860630

32. M. Kiso, A. Hasegawa, Acetonation of some pentoses with 2,2-dimethoxypropane-N,N-dimethylformamide-p-toluenesulfonic acid. *Carbohydr. Res.* **52**, 95–101 (1976). doi:10.1016/S0008-6215(00)85950-9

33. G. Harsch, H. Bauer, W. Voelter, Kinetik, Katalyse und Mechanismus der Sekundärreaktion in der Schlußphase der Formose-Reaktion. *Liebigs Ann. Chem.* **1984**, 623–635 (1984). doi:10.1002/jlac.198419840402

34. M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, T. R. Tuttle, The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. *J. Phys. Chem. A* **102**, 7787–7794 (1998). doi:10.1021/jp982638r

35. A. V. Marenich, C. J. Cramer, D. G. Truhlar, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009). Medline doi:10.1021/jp810292n

36. D. R. Lide, *CRC Handbook of Chemistry and Physics* (CRC Taylor & Francis, ed. 89., 2008).

5.2 Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis

Sidney Becker, Christina Schneider, Hidenori Okamura, Antony Crisp, Tynchtyk Amatov, Milan Dejmek, Thomas Carell

Prolog

Die Moleküle des Lebens entstanden durch einen kontinuierlichen physikalisch-chemischen Prozess auf einer frühen Erde. Hierbei wurde die Entstehung neuer chemischer Verbindungen durch Schwankungen der natürlich gegebenen physikalischen Parameter vorangetrieben. Solche Schwankungen könnten durch Tag-Nacht, Nass-Trocken oder saisonalen Zyklen entstanden sein. Im Gegensatz zu diesen frühzeitlichen Bedingungen werden die präbiotischen Synthesen im Labor oftmals aus hoch reinen Ausgangsstoffen unter streng kontrollierten Bedingungen durchgeführt sowie die Zwischenprodukte mit technischen Aufreinigungsverfahren isoliert. Daher war das Ziel dieser Arbeit aufzuzeigen, dass die Synthese von RNA-Bausteinen unter einfachsten Bedingungen möglich ist, wobei nur Fluktuationen von physikalischen Parametern verwendet wurden, um die jeweiligen Zwischenprodukte zu isolieren. Außerdem konnte gezeigt werden, dass die verwendete Chemie so robust ist, dass komplett reine Intermediate nicht benötigt werden. Interessanterweise stellte sich hierbei heraus, dass nicht nur die kanonischen Nukleoside zugänglich sind, sondern eine Reihe weiterer modifizierter RNA-Bausteine entstanden. Die gefundenen Modifikationen sind bis heute Teil des genetischen Apparates eines jeden Lebewesens. Daher kann vermutet werden, dass modifizierte RNA-Bausteine möglicherweise schon während der chemischen Evolution eine Rolle gespielt haben könnten, welche auf Grund ihrer Funktionen konserviert wurden.

Autorenbeitrag

Siehe Manuskript.

Lizenz

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-017-02639-1 OPEN

Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis

Sidney Becker [1], Christina Schneider [1], Hidenori Okamura[1], Antony Crisp [1], Tynchtyk Amatov [1], Milan Dejmek[2] & Thomas Carell [1]

The molecules of life were created by a continuous physicochemical process on an early Earth. In this hadean environment, chemical transformations were driven by fluctuations of the naturally given physical parameters established for example by wet-dry cycles. These conditions might have allowed for the formation of (self)-replicating RNA as the fundamental biopolymer during chemical evolution. The question of how a complex multistep chemical synthesis of RNA building blocks was possible in such an environment remains unanswered. Here we report that geothermal fields could provide the right setup for establishing wet-dry cycles that allow for the synthesis of RNA nucleosides by continuous synthesis. Our model provides both the canonical and many ubiquitous non-canonical purine nucleosides in parallel by simple changes of physical parameters such as temperature, pH and concentration. The data show that modified nucleosides were potentially formed as competitor molecules. They could in this sense be considered as molecular fossils.

[1] Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. [2] Institute of Organic Chemistry and Biochemistry ASCR, 16610 Prague 6, Czech Republic. Correspondence and requests for materials should be addressed to T.C. (email: thomas.carell@lmu.de)

The molecules of life originated around 4 billion years ago under conditions governed by the composition of the Earth's crust and atmosphere at that time[1, 2]. Molecules such as nucleic acids and amino acids must have formed by a continuous physicochemical process, in which greater structural complexity was generated based on fluctuations of the naturally given physical parameters[3]. Geothermal fields, for example, could have established such fluctuations by wet–dry cycles that may have driven chemical transformations, which ultimately allowed the emergence of life[4–8]. The appearance of (self)-replicating RNA was certainly of central importance for the transition from an abiotic world to biology[9–11]. We need to consider, however, that an early genetic polymer might have been structurally different from contemporary RNA. This involves differences regarding the sugar configuration (e.g., pyranosyl RNA) or the presence of other nucleobases[12, 13]. Selection pressure led in this scenario to the chemical evolution of RNA. Contemporary RNA molecules contain four canonical nucleosides (A, G, C, U), which establish the sequence information. In addition, >120 non-canonical nucleosides are present, which govern a diverse set of properties such as correct folding, e.g., to enable catalysis[14]. In fact, the genetic system of all known life is dependent on modified nucleosides. Many of these non-canonical nucleosides are found today in all three domains of life, which indicates that they were present early on during the development of life. For the ubiquitous non-canonical nucleosides we may assume that they were already formed as competitors in parallel with the canonical ones on the early Earth[15]. So far, however, a geochemical scenario that would allow for the parallel formation of canonical and non-canonical RNA building blocks by a continuous process is not known. All reported multistep chemical models so far rely on tightly controlled laboratory conditions and the isolation and purification of central reaction intermediates by sophisticated methods[1, 16–19].

Herein, we report a robust synthetic pathway, which is purely based on fluctuations of physicochemical parameters such as pH, concentration, and temperature, driven by wet–dry cycles. These fluctuations enable the direct enrichment or purification of all reaction intermediates that are directly used for the next synthetic steps. As such, a continuous synthesis is established. Our results show that RNA building blocks can indeed be formed in a prebiotically plausible geochemical environment without sophisticated isolation and purification procedures. The chemical scenario presented here supports the hypothesis that life may have originated in a hydrothermal milieu on land rather than in a deep sea environment. The key assembly step in our pathway is the formation of variously substituted 5-nitroso-pyrimidines (nitrosoPys) that can be converted into formamidopyrimidines (FaPys) in the presence of formic acid and elementary metals (Ni or Fe). When combined with ribose, the FaPy compounds react to give a set of purine nucleosides. This chemical pathway delivers not only the canonical purine nucleosides but concomitantly many of the ubiquitously present non-canonical relatives, suggesting their origin as prebiotic competitor nucleosides (A, ms^2A, m^2A, DA, G, m^2G, $m^2{}_2G$, m^1G). Since chemical evolution depended on those molecules that were available on early Earth, these non-canonical nucleosides may be considered to be molecular fossils, which maintained their essential life-supporting character until the present day.

Results

Selective crystallization of an organic salt. The chemical scenario that leads to a continuous synthesis of RNA building blocks by just fluctuations of physical parameters is shown in Figs. 1 and 2a. The scenario starts with an aqueous solution of malononitrile **1** and different amidinium salts **2a-d** (HCl or H_2SO_4 salts, 400 mM), both recognized prebiotic compounds[18]. In addition,

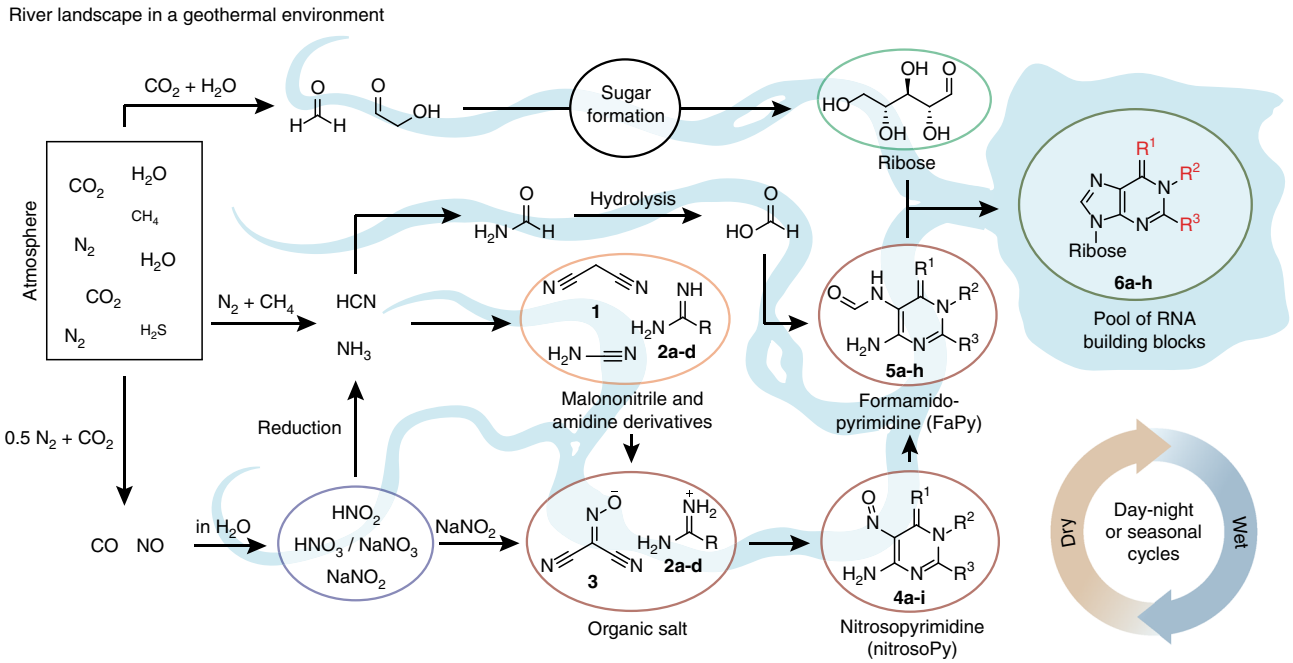


Fig. 1 RNA nucleoside formation pathway. A geothermal environment provides the right set up for the depicted transformations by establishing wet–dry cycles. The prebiotic starting materials are produced from a prebiotic atmosphere and washed into an aqueous environment (e.g. by rain). Major atmospheric components are written in larger letters, whereas minor components are written in smaller letters. Transformations are taking place in different environments, illustrated by various rivers (in light blue). Each environment provides the right setup for different chemistries, leading to several different chemical transformations. This geochemical setup leads to a set of canonical and non-canonical RNA building blocks by continuous synthesis (**6a**, m^1G: $R^1 = O$, $R^2 = Me$, $R^3 = NH_2$; **6b**, ms^2A: $R^1 = NH$, $R^2 = H$, $R^3 = SMe$; **6c**, A: $R^1 = NH$, $R^2 = H$, $R^3 = H$; **6d**, m^2G: $R^1 = O$, $R^2 = H$, $R^3 = NHMe$; **6e**, $m^2{}_2G$: $R^1 = O$, $R^2 = H$, $R^3 = N(Me)_2$; **6f**, G: $R^1 = O$, $R^2 = H$, $R^3 = NH_2$; **6g**, DA: $R^1 = NH$, $R^2 = H$, $R^3 = NH_2$; **6h**, m^2A: $R^1 = NH$, $R^2 = H$, $R^3 = Me$)

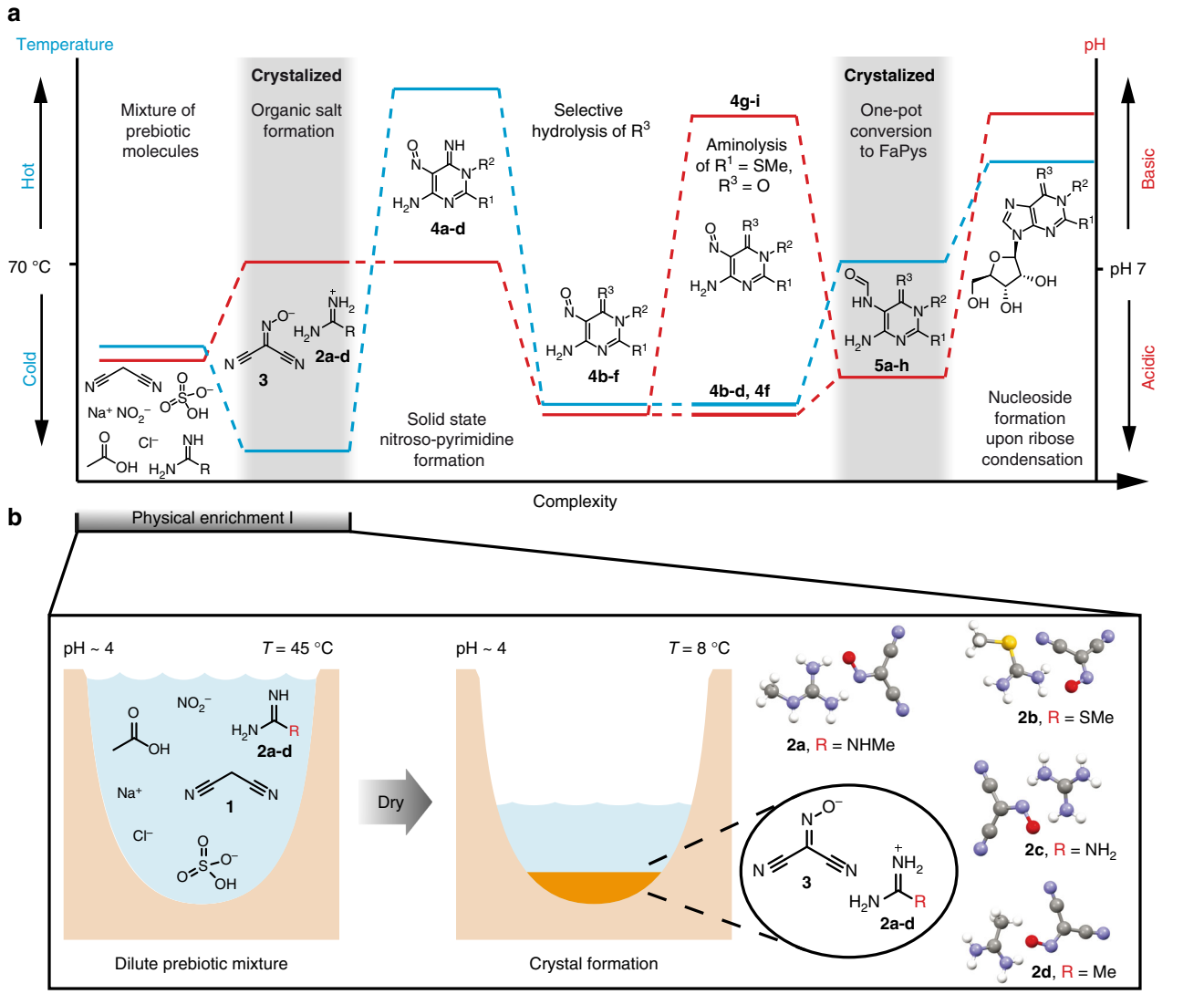


Fig. 2 Chemical complexity created by physical fluctuations. **a** Relative changes of temperature (in blue) and pH (in red) are shown for each synthetic step for the continuous synthesis of purine RNA building blocks from small organic and inorganic molecules. Several wet-dry cycles establish fluctuations of the depicted physical parameters that enable the physical enrichment of intermediates. Gray backgrounds denote compounds that are enriched by crystallization from an aqueous solution. **b** Formation of an organic salt consisting of amidine derivatives **2a-d** and (hydroxyimino)malononitrile **3**. The salt is selectively crystalized by concentrating a dilute mixture of organic and inorganic compounds by slow evaporation. The crystal structures of the four crystalized organic salts are depicted (Supplementary Tables 1–4)

sodium nitrite and acetic acid are present to establish a slightly acidic pH of around 4. Under these conditions the amidine molecules (**2a-d**) are protonated, which leads to their chemical deactivation. This allows selective nitrosation of malononitrile **1** to give (hydroxyimino)malononitrile **3** in situ. Slow evaporation of water under ambient conditions, followed by gentle cooling to 8–10 °C resulted in crystallization of a salt from the ca. 1 M amidinium solution. This crystallization is very robust and resembles naturally occurring concentration processes. The resultant crystals had excellent quality for X-ray analysis, which showed that the salts are formed from the amidinium cations **2a-d** and the (hydroxyimino)malononitrile anion of **3** (Fig. 2b). Interesting is the distance between the negatively charged oxygen in **3** and the positively charged H-bond donor centre of the amidininium units **2a-d**. We determined distances between 1.85–1.95 Å, which is long for a salt bridge but right in the regime for a typical hydrogen bond. This is important because it is supposedly the reason for the comparably low melting temperatures of the salts, which we determined between 110 and 160 °C. The robustness and ease of crystallization establishes a first physical enrichment step that finishes the initial wet–dry phase with the deposition of these salt materials (Fig. 2b).

Nitroso-pyrimidine formation. When the obtained salts containing **2a-d** and **3** are subsequently heated to their respective melting temperatures, transformation into the corresponding nitroso-pyrimidines (**4a-d**, Fig. 3a) occurs. The required temperatures between 110 and 160 °C could have been readily accessible under early Earth conditions, due to, for example, volcanic activity in geothermal fields or sunlight shining on dark surfaces. In order to investigate whether the nitroso-compounds **4a-d** would form in parallel despite their varying structures and different melting points, the different salts were combined in a reaction flask and a temperature gradient (1 °C/5 min, from 100–160 °C) was applied to simulate soil that would slowly heat up. Subsequent ^{1}H-NMR analysis indicated successful formation of the anticipated nitroso-pyrimidines **4a-d** (Supplementary Fig. 1).

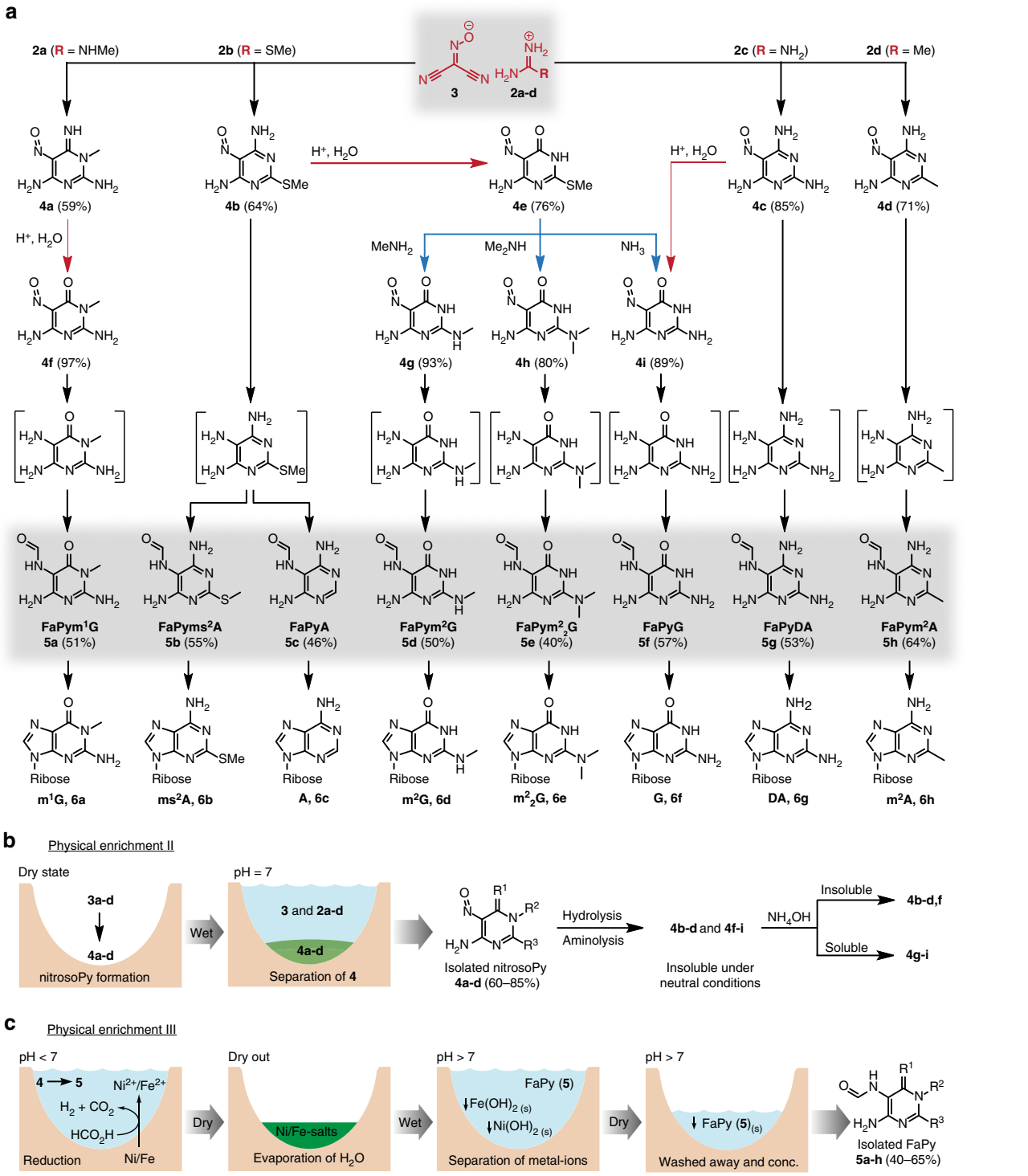


Fig. 3 Reaction scheme and physical enrichment of intermediates. **a** Dry-state reactions of salts containing **2a-d** and **3** provide nitroso-pyrimidines **4a-d**, which can be further diversified by hydrolysis (red arrows) or aminolysis (blue arrows) to give a set of nitroso-pyrimidines (nitrosoPys) **4a-i**. In the presence of elementary Fe and Ni and dilute formic acid, formation of the formamidopyrimidines (FaPys) **5a-h** as direct purine base precursors takes place. In square brackets: non-isolated reaction intermediates. **b** Second physical enrichment of the nitroso-pyrimidines isolated in high purity and yield. **c** Third physical enrichment of the formed FaPys **5a-h** as nucleoside precursors from nitroso-pyrimidines

The resultant nitroso-pyrimidines are stable compounds with melting points typically >250 °C without decomposition. In addition we noted that the nitroso-pyrimidines are rather insoluble in water, which offers the possibility for a second physical enrichment step. Addition of water to the reaction mixture dissolves unreacted starting materials, leaving the nitroso-pyrimidines in basically NMR-pure form behind (Supplementary Fig. 2). In this model, one wet–dry cycle and two physical enrichment steps with a final rain shower or flooding would be sufficient to deposit a mixture of stable nitroso-pyrimidines (**4a-d**) in excellent purities and good chemical yields between 60 and 85% (Fig. 3a).

Diversification by hydrolysis and aminolysis. Depending on the composition and pH of the aqueous environment, which may or may not contain different amines, the nitroso-pyrimidines could undergo further hydrolysis and aminolysis reactions (Fig. 3a). Because these reactions are very slow under neutral conditions, we used dilute HCl to accelerate the processes for investigation. Importantly, we noted a high regioselectivity. Upon treatment overnight at room temperature with 0.5 M HCl, compounds **4a** and **4c** for example are hydrolyzed to afford the oxo-nitroso-pyrimidines **4f** and **4i** in near quantitative yields. Hydrolysis of **4b** to product **4e** was comparitively slower, and under our accelerated conditions a mixture of **4b** and **4e** was obtained. This inefficient conversion would be advantageous in a prebiotic context given that from **4b** the canonical nucleoside adenosine (A) and its 2-thiomethyl derivative (ms^2A) are derived later, whereas **4e** gives rise to guanosine derivatives (G, m^2G, $m^2{}_2G$, Fig. 3a). This allows for the simultaneous formation of canonical and non-canonical bases from the same precursor. In contrast to the 2-amino (**4a,c**) or 2-methyl (**4d**) substituted nitroso-pyrimidines, we noted that the 2-thiomethyl functionality in **4b** and **4e** was prone to undergo selective nucleophilic substitution. Reaction of **4e** with different amines leads to efficient formation of the nitroso-pyrimidines **4g-i** with the concomitant release of methanethiol. Due to its insolubility under basic conditions, nucleophilic substitutions of **4b** are very inefficient. To confirm this, we partially hydrolyzed **4b** to **4e** in the presence of methylamine (300 mM) and dimethylamine (100 mM). The pH was carefully adjusted with Na_2CO_3 to about pH 10. Compound **4b** precipitated, while **4e** stayed in solution, consequently protecting **4b** from further reactions. It is in this context interesting that nucleosides that would form via aminolysis of **4b** have not yet been found in nature. In contrast, **4e** reacts efficiently and after 3–4 days at room temperature **4e** is almost completely converted into **4g** and **4h**, which are direct precursors to the ubiquitous non-canonical RNA bases m^2G and $m^2{}_2G$ (Fig. 3a, Supplementary Fig. 3).

Thus, a few simple chemoselective and regioselective hydrolysis and aminolysis reactions affords a diverse mixture of differently substituted nitroso-pyrimidines (**4b-d, f-i**), all of which possess the right substitution pattern for the synthesis of naturally occuring canonical and non-canonical RNA nucleosides. Because all the formed nitroso-pyrimidines are poorly soluble in water at neutral pH, neutralizing the solutions leads to their efficient precipitation, providing a naturally occurring purification step (Fig. 3b). Importantly, all nitroso-pyrimidines that later give adenosine-derived nucleosides (**4b-d**) are soluble in water under acidic conditions, while the nitroso-compounds that are converted into guanosine-derived nucleosides (**4g-i**, except for **4f**) are soluble under basic pH conditions. These properties allow for potentially divergent chemical pathways leading to A-derived and G-derived nucleosides (Fig. 3b, Supplementary Fig. 4).

Formamidopyrimidine formation as nucleobase precursor. The next wet–dry cycles allow for the formation and isolation of formamidopyrimidines (FaPys) **5a-h**, from nitroso-pyrimidines **4** that are after their precipitation exposed to acidic conditions like dilute formic acid in the presence of elementary Fe or Ni, which are components of the Earth's crust. This leads to reduction of the nitroso-pyrimidines **4** to aminopyrimidines as non-isolated reaction intermediates (Fig. 3a, in square brackets), which are immediately formylated to give the water soluble formamidopyrimidines (FaPys) **5a-h** in a one-pot reaction. During the wet phase, Ni^0 and Fe^0 are converted into the biologically relevant Ni^{2+}/Fe^{2+} ions, while formic acid decomposes into CO_2 and H_2 (Fig. 3c). In the reaction formic acid has a dual function. It provides the H-atoms needed for the reduction and it subsequently reacts with the formed aminopyrimidines to give FaPy compounds that were already shown to be prebiotically valid precursors to purine nucleosides[18]. The Ni/Fe/formic acid environment converts quantitatively all nitroso-compounds **4b-d, f-i** into the corresponding FaPy compounds **5a-h** (Fig. 3a). The water soluble FaPy compounds (under dilute basic conditions) can now be separated from unreacted Ni^0/Fe^0 and from the formed Ni^{2+}/Fe^{2+} byproducts. Under slightly basic conditions (pH ≈ 9–10) the latter compounds precipitate as insoluble carbonate or hydroxide salts. The FaPys **5a-h** are thus washed away, while the transition metal compounds sediment out. Final evaporation of water concentrates the reaction mixture, leading to the crystallization of the FaPy molecules. This third physical enrichment step, involving a wet–dry cycle, leads to the NMR-clean formation of FaPy-derivatives **5a-h** (Fig. 3c).

The 2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (**4b**) gives after treatment with formic acid and elementary Ni two different FaPy products depending on the reaction conditions. One of the products (**5b**) contains a thiomethyl group, while the other (**5c**) is desulfurated. The desulfurization reaction is simply controlled by time and can be promoted when H_2 is bubbled through the solution prior to reaction. Compound **5c** is always generated in a stepwise reaction cascade via compound **5b** which was confirmed by reacting **4b** for 2 h and isolating the only product formed (**5b**, Fig. 3c). The isolated product was immediately subjected to the same conditions, which provided **5c** after 7 days in pure form. This pathway via nitroso-pyrimidines thus affords **5c**, the precursor for the canonical base A under plausible prebiotic conditions[20]. These conditions also lead to the parallel formation of the precursor to the ubiquitous 2-thiomethyl modification (ms^2A), which is today found in all three domains of life.

Formation of canonical and non-canonical nucleosides. All of the prepared FaPy compounds undergo rapid and regioselective condensations with ribose when they are present in the same dry-state environment (Fig. 4). We do not assume that ribose was formed at the same location together with the FaPy compounds since the required carbohydrate chemistry may be incompatible. Several models are available, however, that show ribose formation in different physical environments[21–24]. Even though ribose and FaPys might have formed separately, the water solubility of the FaPys and of ribose allows them to be washed into the same environment by rain or flooding. Evaporation of water in the last wet–dry cycle would enable a condensation reaction under dry-state conditions. Indeed, the physically enriched FaPy compounds (**5a-h**) engage in a rapid reaction with ribose to give the corresponding FaPy-ribosides. Upon dissolution in water and subsequent heating under basic conditions, all four expected purine α/β-ribofuranosides (α/β–f) and α/β-pyranosides (α/β–p) are obtained (**6a-h**, Fig. 4a), completing the last wet–dry cycle. The LC-MS traces of the reactions using both UV- and MS-detection are shown in Fig. 4b. To ensure correct structural assignment we chemically synthesized some of the expected products and performed co-injection studies (Supplementary Methods). These experiments show that the major isomers are the naturally occurring β-configured pyranosides and furanosides. Pyranosides are building blocks for pyranosyl-RNA, which was suggested to be a potential RNA predecessor[12]. Therefore, our scenario delivers the building blocks for this pre-RNA and for RNA. As such it provides the basis for the chemical transition from one genetic polymer to the other directed by selection pressure. Importantly, our continuous synthetic pathway provides next to the canonical bases A and G also the non-canonical β-furanosyl-nucleosides (β–f) m^2G, $m^2{}_2G$, m^1G, ms^2A and m^2A (in red, Fig. 4b), arguing

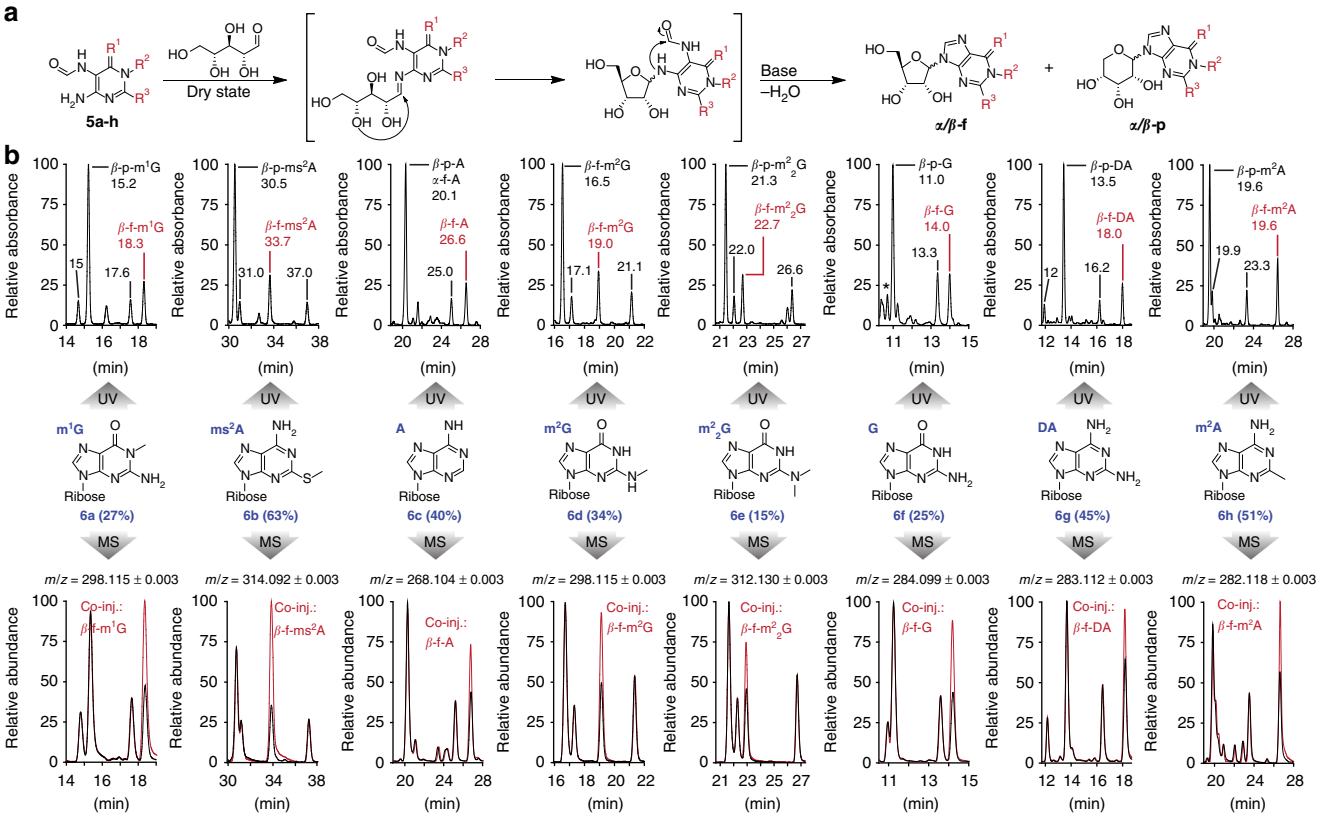


Fig. 4 Formation of RNA nucleosides from nucleobase precursors. **a** Reaction mechanism for the formation of canonical and non-canonical RNA nucleosides **6a-h** from formamidopyrimidines (FaPys) **5a-h** and ribose. The reaction provides the four expected isomers α/β ribopyranosides (α/β–p) and α/β–ribofuranosides (α/β–f). **b** LC-MS analysis of the reaction products from **5a-h** and ribose. Compounds identified by MS detection are labeled (β–p or α/β–f). The UV- and MS-chromatograms show all four expected isomers for each compound (labeled with the retention time or asterisk (*) in the UV-chromatogram). The structural assignment was assisted by co-injection (Co-inj.) studies. The MS traces show in red the co-injection signal obtained with the naturally occurring isomer (β–f)

that the early RNA polymer was structurally already more complex regarding the nucleobases. The ribosylation of the FaPys leading to non-canonical nucleosides is equally efficient to the formation of A and G, with yields between 15 and 60%. Interestingly, we noted that for some A derivatives (m^2A, DA and A) other regioisomers were found as well. These isomers were not formed when pure FaPy starting materials were used that were not derived from our continuous synthesis. We believe that these isomers might be the N3-connected nucleosides, previously proposed by Wächtershäuser for homo-purine RNA[25]. Despite the presence of these side products, we observe efficient N9-nucleoside formation with remarkable yields of up to 60% for the canonical and the non-canonical nucleosides. This work demonstrates that the non-canonical compounds could plausibly have formed as companion and potential competitor compounds in parallel to the canonical nucleosides.

Discussion

Life on earth certainly did not start in a chemists' laboratory, where the relevant compounds are assembled in a step-by-step process from pure starting materials under tightly controlled conditions. Even if the individual reaction steps are performed under plausibly prebiotic conditions, the controlled assembly over many steps with sophisticated isolation and purification procedures of reaction intermediates is an unlikely scenario for chemical synthesis under early Earth conditions. For the process of chemical evolution on the early Earth, we may rather envision a more continuous synthesis, in which small organic molecules, initially formed by volcanic action or lightning, reacted to give increasingly more complex structures (Fig. 1). Here, chemical transformations may have been driven by physical fluctuations, established for example by day-night, seasonal or wet–dry cycles. Such fluctuating parameters might include temperature, concentration, and pH, which have triggered selective precipitation and crystallization to purify and concentrate reaction intermediates (Fig. 2a).

Regarding the central nucleoside building blocks of life, we believe that the four canonical nucleosides were finally selected from a more diverse prebiotic nucleoside pool. These canonical bases today establish the sequence information. The synthesis of the canonical purine (A, G)[18] and pyrimidine (U, C)[16] RNA building blocks has been previously demonstrated in aqueous environments. It is questionable, however, if these multistep synthesis pathways are able to provide all four canonical bases at the same time, which fuels the development of new prebiotically plausible nucleoside formation reactions[17, 26, 27]. Recently, all four canonical nucleosides (A, G, U, C) were accessed in low yields via a one-pot procedure from pure formamide[28]. However, in order to establish a functional genetic system a number of non-canonical nucleosides is required as well that provide other functions related to folding and catalysis[29–34]. Since many of these non-canonical bases are present in all three domains of life, it is likely that they have been early on part of the abiotic chemical selection process[15]. We report here the discovery of a continuous synthesis pathway that enables the efficient production of

canonical and non-canonical purine bases in parallel. Our data show that formation of the many nucleosides needed to establish a functional genetic system is in fact an unavoidable event if we assume the presence of simple starting materials such as formic acid, acetic acid, sodium nitrite, malononitrile (**1**), amidinium compounds (**2a-d**), as well as transition metals like Ni or Fe. These simple compounds react in several successive wet–dry phases, leading to physical enrichment (I, II, and III) of reaction intermediates to finally give RNA building blocks. Wet–dry cycles have already been shown to be a plausible geological scenario especially for polymerization reactions[35, 36]. Our reported chemical scenario shows now that such a geological setup can also result in a diverse set of purine nucleosides by continuous synthesis (**6a-h**, Fig. 4). These nucleosides can be converted into the phosphorylated nucleotides based on recent advances in prebiotic phosphorylation reactions[37, 38]. So far, however, we are not yet able to include this step into our continuous synthesis.

Importantly, all here-reported non-canonical bases are known to exist in the three domains of life. Many are postulated components of the early genetic system of the last universal common ancestor (LUCA), suggesting that they were indeed present already at the onset of biological evolution[39, 40]. Based on the continuous synthesis pathway reported here, we hypothesize that the canonical and at least these non-canonical nucleosides could have formed side by side, dating the formation of the first non-canonical nucleosides back to the origin of chemical evolution around 4 billion years ago. As such, they could have been part of the chemical evolution process that established the putative RNA world[41]. Our chemistry invokes that methylated and thiomethylated nucleosides could particularly have been integral components of the first instructional (pre)-RNA molecules, likely to stabilize folded structures in order to accelerate catalytic processes[29–34]. We therefore propose that these nucleosides could be vestiges and molecular fossils of an early Earth, as it was suggested for cofactors[42].

Methods

1-methylguanidine (2a) salt of (hydroxyimino)malononitrile (3). 1-methylguanidine (**2a**) hydrochloride salt (10.95 g, 100 mmol, 1 eq.) and malononitrile (**1**) (6.65 g, 100 mmol, 1 eq.) was dissolved in H_2O (230 mL, containing 6 mL of AcOH) in a 500 mL beaker. A solution of $NaNO_2$ (7.00 g, 101 mmol, 1.01 eq., in 20 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 100 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (6.70 g, 40 mmol, 40%).

Mp.: 108 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 7.23 (br m, 5H), 2.72 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 157.85, 119.50, 113.31, 107.18, 28.23. **IR** (cm^{-1}): 3405 (m), 3351 (br, m), 3197 (m), 2977 (br, m), 2229 (s), 2218 (s) 1675 (s), 1635 (s), 1465 (w) 1428 (m), 1344 (s), 1269 (s) 1226 (s), 1172 (w), 1098 (m), 915 (m), 765 (m).

Methylthioamidine (2b) salt of (hydroxyimino)malononitrile (3). *S*-Methylisothiourea (**2b**) hemisulfate salt (27.8 g, 200 mmol, 1 eq.) and malononitrile (**1**) (13.3 g, 200 mmol, 1 eq.) was dissolved in H_2O (460 mL, containing 12 mL of AcOH) in a 600 mL beaker. A solution of $NaNO_2$ (14.0 g, 202 mmol, 1.01 eq., in 40 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 200 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (16.7 g, 90 mmol, 45%).

Mp.: 126 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ 8.90 (s, 4H), 2.56 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ 171.62, 119.46, 113.29, 107.17, 13.71. **IR** (cm^{-1}): 3282(m), 3146 (br, m), 2742 (br, w), 2530 (w), 2222 (s), 2213 (s) 1698 (m), 1663 (s), 1643 (s) 1549 (m), 1450 (m), 1424 (s) 1375 (w), 1335 (s), 1269 (s), 1223 (s), 1180 (w), 1099 (m), 1076 (w), 982 (m), 970 (w), 960 (w), 897 (br, m), 801 (m), 736 (m).

Guanidine (2c) salt of (hydroxyimino)malononitrile (3). Guanidine (**2c**) hydrochloride salt (9.55 g, 100 mmol, 1 eq.) and malononitrile (**1**) (6.65 g, 100 mmol, 1 eq.) was dissolved in H_2O (230 mL, containing 6 mL of AcOH) in a 500 mL beaker. A solution of $NaNO_2$ (7.00 g, 101 mmol, 1.01 eq., in 20 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 100 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (8.20 g, 53 mmol, 53%).

Mp.: 159 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 6.91 (s, 6H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 158.32, 119.52, 113.32, 107.18. **IR** (cm^{-1}): 3473 (m), 3373 (m), 3172 (w), 3087 (w), 2815 (br, w), 2223 (s), 2217 (s), 1668 (m), 1641 (s), 1578 (w), 1552 (w), 1369 (w), 1343 (s), 1294 (w), 1264 (s), 1220 (s), 1140 (w), 975 (w), 792 (w), 757 (w).

Acetamidine (2d) salt of (hydroxyimino)malononitrile (3). Acetamidine (**2d**) hydrochloride salt (9.45 g, 100 mmol, 1 eq.) and malononitrile (**1**) (6.65 g, 100 mmol, 1 eq.) was dissolved in H_2O (230 mL, containing 6 mL of AcOH) in a 500 mL beaker. A solution of $NaNO_2$ (7.00 g, 101 mmol, 1.01 eq., in 20 mL of H_2O) was slowly added at room temperature. After stirring at room temperature for 2 h the reaction mixture was kept at 45 °C in an oil bath for 3–4 days open to the air until the mixture was concentrated to about 100 mL. The reaction mixture was placed in a fridge overnight at 8–10 °C. The formed yellow crystals were filtered off to give the desired product (8.80 g, 60 mmol, 60%).

Mp.: 142 °C. **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 8.60 (s, 4H), 2.11 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 168.11, 119.50, 113.31, 107.18, 18.72. **IR** (cm^{-1}): 3282 (m), 3140 (br, m), 2781 (m), 2395 (w), 2236 (s), 2217 (s) 1708 (s), 1661 (s), 1592 (br, w) 1513 (m), 1373 (s), 1351 (s) 1259 (s), 1200 (s), 1191 (s), 1160 (w), 1125 (m), 969 (w), 907 (w), 883 (w), 857 (w), 790 (s), 691 (s), 684 (s).

Synthesis of nitrosopyrimidine 4a-d from salts containing 2a-d and 3. The reaction time of the following procedures for the formation of **4a-d** depend on the crystal size of the organic salts and the heating source. Usually conversions were done in a beaker open to the air in an oil bath to simulate hot soil. Large crystals sometimes already convert into nitroso-pyrimidines by a solid-state reaction without melting. Then much longer reaction times of up to 7 days are required because of unequally distributed temperature. Alternatively, the organic salt can be converted by heating in an oven where temperature is equally distributed within the sample.

6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine (4a). 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**) (5.00 g, 29.5 mmol, 1 eq.) was heated slowly to its melting temperature of 108 °C and kept overnight. The compound melts suddenly but becomes a dark red solid again after leaving it overnight. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**) (100 mg, 0.59 mmol, 1 eq.) was reacted as described above. The reaction mixture was suspended well in water (2–3 mL) and the dark red solid was filtered off to give 6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine (59 mg, 0.35 mmol, 59%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 11.44 (s, 1H), 8.41 (s, 1H), 8.08 (br, 1H), 7.67 (br, 1H), 7.49 (s, 1H), 3.26 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 165.44, 157.64, 146.95, 137.22, 27.64. **HRMS** (ESI +): calc. for $[C_5H_9N_6O]^+$ 169.0832, found: 169.0832 $[M + H]^+$

2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (4b). Methylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**) (5.00 g, 27 mmol, 1 eq.) was heated slowly to its melting temperature of 126 °C and kept overnight. Caution: if the product is heated too quickly above the melting temperature it decomposes with the release of MeSH! The compound melts suddenly but becomes a dark green solid again. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of methylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**) (100 mg, 0.54 mmol, 1 eq.) was reacted as described above. The reaction mixture was suspended in water (2–3 mL) and the dark green solid was filtered off to give 2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (64 mg, 0.35 mmol, 64%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 10.18 (d, J = 4.2 Hz, 1H), 9.00 (s, 1H), 8.42 (d, J = 4.2 Hz, 1H), 8.02 (s, 1H), 2.46 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ 179.05, 164.73, 146.22, 139.43, 14.08. **HRMS** (ESI+): calc. for $[C_5H_8N_5OS]^+$ 186.0444, found: 186.0444 $[M + H]^+$

5-nitrosopyrimidine-2,4,6-triamine (4c). Guanidine (**2c**) salt of (hydroxyimino) malononitrile (**3**) (5.00 g, 32 mmol, 1 eq.) was heated slowly to its melting temperature of 159 °C and kept overnight. The compound melts suddenly but becomes a red/pinkish solid again after leaving it overnight. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of guanidine (**2c**) salt of (hydroxyimino) malononitrile (**3**) (100 mg, 0.65 mmol, 1 eq.) was reacted as described above. The reaction mixture was suspended in water (2–3 mL) and the solid was filtered off to give NMR clean 5-nitrosopyrimidine-2,4,6-triamine (85 mg, 0.55 mmol, 85%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.26 (d, J = 5.1 Hz, 1H), 8.15 (s, 1H), 7.75 (d, J = 5.1 Hz, 1H), 7.35 (s, 1H), 7.19 (s, 2H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.52, 165.32, 151.43, 138.04. **HRMS** (ESI+): calc. for $[C_4H_7N_6O]^+$ 155.0676, found: 155.0676 $[M + H]^+$

2-methyl-5-nitrosopyrimidine-4,6-diamine (4d). Acetamidine (**2d**) salt of (hydroxyimino)malononitrile (**3**) (5.00 g, 32.5 mmol, 1 eq.) was heated slowly to its melting temperature of 142 °C and kept overnight. The compound melts suddenly but becomes a red/pinkish solid again after leaving it overnight. The quantitative reaction mixture can be directly used for the next step without purification.

For analytical reasons a small batch of acetamidine (**2d**) salt of (hydroxyimino) malononitrile (**3**) (100 mg, 0.65 mmol, 1 eq.) was reacted as described above. After suspension in H_2O (2–3 mL) the solid is filtered off to give NMR clean 2-methyl-5-nitrosopyrimidine-4,6-diamine (71 mg, 0.46 mmol, 71%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.04 (d, J = 3.2 Hz, 1H), 8.97 (s, 1H), 8.36 (d, J = 3.2 Hz, 1H), 7.96 (s, 1H), 2.20 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 175.20, 166.38, 146.76, 139.94, 26.83. **HRMS** (ESI-): calc. for $[C_5H_6N_5O]^-$ 152.0578, found: 152.0578 $[M-H]^-$

Data availability. All data generated or analyzed during this study are presented in this article and its Supplementary Information File, or are available from the corresponding author upon reasonable request. X-ray crystallographic data were also deposited at the Cambridge Crystallographic Data Centre (CCDC) under the following CCDC deposition numbers: 1574226 for 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**); 1574223 for methylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**); 1574225 for guanidine (**2c**) salt of (hydroxyimino)malononitrile (**3**); 1574224 for acetamidine (**2d**) salt of (hydroxyimino) malononitrile (**3**). These can be obtained free of charge from CCDC via www.ccdc.cam.ac.uk/data_request/cif.

Received: 10 October 2017 Accepted: 14 December 2017
Published online: 11 January 2018

References

1. Benner, S. A., Kim, H.-J. & Carrigan, M. A. Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc. Chem. Res.* **45**, 2025–2034 (2012).
2. Budin, I. & Szostak, J. W. Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* **39**, 245–263 (2010).
3. Pace, N. R. Origin of life-facing up to the physical setting. *Cell* **65**, 531–533 (1991).
4. Mulkidjanian, A. Y., Bychkov, A. Y., Dibrova, D. V., Galperin, M. Y. & Koonin, E. V. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc. Natl. Acad. Sci. USA* **109**, E821–E830 (2012).
5. Djokic, T., Van Kranendonk, M. J., Campbell, K. A., Walter, M. R. & Ward, C. R. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. *Nat. Commun.* **8**, 15263 (2017).
6. Damer, B. & Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life* **5**, 872 (2015).
7. Deamer, D., Singaram, S., Rajamani, S., Kompanichenko, V. & Guggenheim, S. Self-assembly processes in the prebiotic environment. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361**, 1809–1818 (2006).
8. Deamer, D., Van Kranendonk, M. & Djokic, T. The new origins of life. *Sci. Am.* **317**, 30–35 (2017).
9. Higgs, P. G. & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17 (2015).
10. Wachowius, F., Attwater, J. & Holliger, P. Nucleic acids: function and potential for abiogenesis. *Q. Rev. Biophys.* **50**, e4 (2017).
11. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2010).
12. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
13. 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).
14. Carell, T. et al. Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.* **51**, 7110–7131 (2012).
15. Rios, A. C. & Tor, Y. On the origin of the canonical nucleobases: an assessment of selection pressures across chemical and early biological evolution. *Isr. J. Chem.* **53**, 469–483 (2013).
16. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
17. Stairs, S. et al. Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun.* **8**, 15270 (2017).
18. Becker, S. et al. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).
19. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
20. Fiore, M. & Strazewski, P. Bringing prebiotic nucleosides and nucleotides down to earth. *Angew. Chem. Int. Ed.* **55**, 13930–13933 (2016).
21. Kim, H.-J. et al. Synthesis of carbohydrates in mineral-guided prebiotic cycles. *J. Am. Chem. Soc.* **133**, 9457–9468 (2011).
22. Kofoed, J., Reymond, J.-L. & Darbre, T. Prebiotic carbohydrate synthesis: zinc-proline catalyzes direct aqueous aldol reactions of α-hydroxy aldehydes and ketones. *Org. Biomol. Chem.* **3**, 1850–1855 (2005).
23. Harsch, G., Bauer, H. & Voelter, W. Kinetik, katalyse und mechanismus der sekundärreaktion in der schlußphase der formose-reaktion. *Liebigs Ann.* **1984**, 623–635 (1984).
24. Meinert, C. et al. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **352**, 208–212 (2016).
25. Wächtershäuser, G. An all-purine precursor of nucleic acids. *Proc. Natl. Acad. Sci. USA* **85**, 1134–1135 (1988).
26. Kim, H.-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci. USA* **114**, 11315–11320 (2017).
27. Saladino, R. et al. Proton irradiation: a key to the challenge of N-glycosidic bond formation in a prebiotic context. *Sci. Rep.* **7**, 14709 (2017).
28. Saladino, R. et al. Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proc. Natl. Acad. Sci. USA* **112**, E2746–E2755 (2015).
29. Björk, G. R. et al. A primordial tRNA modification required for the evolution of life? *EMBO J.* **20**, 231–239 (2001).
30. Grosjean, H. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (Taylor & Francis, Boca Raton, FL, 2009).
31. Engel, J. & von Hippel, P. H. Effects of methylation on the stability of nucleic acid conformations. *J. Biol. Chem.* **253**, 927–934 (1978).
32. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem. Sci.* **27**, 344–351 (2002).
33. Motorin, Y. & Helm, M. tRNA stabilization by modified nucleotides. *Biochemistry* **49**, 4934–4944 (2010).
34. Helm, M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **34**, 721–733 (2006).
35. Forsythe, J. G. et al. 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).
36. Da Silva, L., Maurel, M.-C. & Deamer, D. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *J. Mol. Evol.* **80**, 86–97 (2015).
37. Burcar, B. et al. Darwin's warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angew. Chem. Int. Ed.* **55**, 13249–13253 (2016).
38. Kim, H.-J. et al. Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: borate as a multifaceted problem solver in prebiotic chemistry. *Angew. Chem. Int. Ed.* **55**, 15816–15820 (2016).
39. Anantharaman, V., Koonin, E. V. & Aravind, L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30**, 1427–1464 (2002).
40. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).
41. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
42. Jadhav, V. R. & Yarus, M. Coenzymes as coribozymes. *Biochimie* **84**, 877–888 (2002).

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support through the research funding schemes SFB749, SFB1032, SPP1784, GRK2062, and the DFG Normalverfahrensprogram CA275-11/1. We thank the European Union Horizon 2020 program for funding the ERC Advanced project EPiR (741912). Further we thank the excellence cluster CiPS-M. We thank Florian Steinmann for synthetic help, Dr Markus Müller for further discussions and Dr Peter Mayer for providing X-ray structures.

Author contributions

S.B. developed the chemistry for the nitrosopyrimidine route, helped designing the study, analyzed and interpreted results and helped writing the manuscript. C.S., H.O., A.C., T.A. and M.D. performed supportive chemistry. T.C. designed the study, supervised all work, interpreted data and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02639-1.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

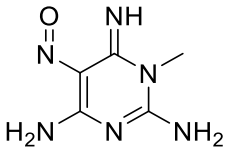
Supplementary Methods

General Information

Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR, Carbosynth or Acros organics and used without further purification. Solutions were concentrated *in vacuo* on a Heidolph rotary evaporator. The solvents were of reagent grade or purified by distillation. Chromatographic purification of products was accomplished using flash column chromatography on Merck Geduran Si 60 (40-63 µm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on Merck 60 (silica gel F254) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or standard staining solutions. ^{1}H- and ^{13}C-NMR spectra were recorded in deuterated solvents on Varian Oxford 200, Bruker ARX 300, Varian VXR400S, Varian Inova 400, Bruker AMX 600 and Bruker AVIIIHD 400 spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. High-resolution ESI spectra were obtained on the mass spectrometers Thermo Finnigan LTQ FT-ICR. IR measurements were performed on Perkin Elmer Spectrum BX FT-IR spectrometer with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were measured on a Büchi B-540. For preparative HPLC purification a Waters 1525 binary HPLC Pump in combination with a Waters 2487 Dual Absorbance Detector was used, with Macherey-Nagel VP 250/10 Nucleosil 100-7 C18 reversed phase column. The prebiotic reactions were analyzed by LC-ESI-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Interchim Uptisphere120-3HDO C18 column. The column temperature was maintained at 30 °C. Eluting buffers were buffer A (2 mM $HCOONH_4$ in H_2O (pH 5.5)) and buffer B (2 mM $HCOONH_4$ in H_2O/MeCN 20/80 (pH 5.5)). The gradient for reactions of FaPy **5a-h** with ribose was 0 → 45 min; 0% → 20% or 0% → 10% buffer B. The elution was monitored at 260 nm (Dionex Ultimate 3000 Diode Array Detector). The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 120-1000 with a resolution of 30000. Parameters of the mass spectrometer were tuned with a freshly mixed aqueous solution of adenosine (5 µM). The synthetic standards for the coinjection experiments were synthesized in our lab (see synthesis of the synthetic standards) or purchased from Sigma-Aldrich, Fluka, ABCR, Carbosynth or Acros organics. The X-ray intensity data was measured at a temperature of 100 K on a Bruker D8 Venture TXS system equipped with a multilayer mirror optics monochromator and a Mo Kα rotating-anode X-ray tube (λ = 0.71073 Å). The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. Data were corrected for absorption effects using the Multi-Scan method (SADABS). The structures were solved and refined using the Bruker SHELXTL software package.

Synthetic Procedures

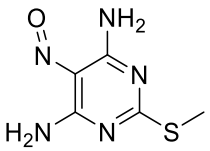
6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine (4a)



Synthetic reference: To 6-imino-1-methyl-1,6-dihydropyrimidine-2,4-diamine hydrogen iodide[1] (6.00 g, 22.5 mmol, 1 eq.) in H_2O was added CF_3COOAg (5.00 g, 22.5 mmol, 1 eq.). The flask was vigorously shaken for 5-10 min. The precipitate was filtered off over celite and washed with H_2O (15 mL). CF_3COOH (17.5 mL, 225 mmol, 10 eq.) was added to the filtrate and the solution was cooled in an ice/H_2O bath. $NaNO_2$ (1.63 g, 23.6 mmol, 1.05 eq.) dissolved in H_2O (15 mL) was added dropwise. After stirring for 30 min at room temperature, the formed precipitate was filtered off and washed with H_2O and Acetone (each 15 mL) to give 6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine triflouroacetic acid (4.13 g, 20.4 mmol, 65%). For comparison with the prebiotic product, a sample of the triflouroacetic acid salt was obtained as a free base by treatment with 3 M NH_4OH. After sonication the product was filtered off to give a reddish solid that turned pink again after drying.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 11.44 (s, 1H), 8.41 (s, 1H), 8.08 (br, 1H), 7.67 (br, 1H), 7.49 (s, 1H), 3.26 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 165.44, 157.64, 146.95, 137.22, 27.64. **HRMS** (ESI+): calc. for $[C_5H_9N_6O]^+$ 169.0832, found: 169.0832 $[M+H]^+$

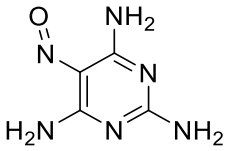
2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (4b)



Synthetic reference: 2-(methylthio)pyrimidine-4,6-diamine (2.00 g, 12.8 mmol, 1 eq.) was dissolved in H_2O (50 mL) and AcOH (2.8 mL) and cooled in an ice/H_2O bath. It was added a solution of $NaNO_2$ (1.92 g, 27.8 mmol, 2.2 eq.) in H_2O (20 mL). The reaction was kept for 2 h at 0 °C and the turquoise precipitate was filtered off to give 2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (2.20 g, 11.9 mmol, 93%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ 10.18 (d, J = 4.2 Hz, 1H), 9.00 (s, 1H), 8.42 (d, J = 4.2 Hz, 1H), 8.02 (s, 1H), 2.46 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ 179.05, 164.73, 146.22, 139.43, 14.08. **HRMS** (ESI+): calc. for $[C_5H_8N_5OS]^+$ 186.0444, found: 186.0444 $[M+H]^+$

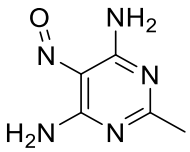
5-nitrosopyrimidine-2,4,6-triamine (4c)



Synthetic reference: The reaction was performed slightly modified according to literature.[2] 6-aminopyrimidine-2,4-diamine (900 mg, 7.20 mmol, 1eq.) was suspended in H_2O (10 mL). AcOH (660 µl, 11.5 mmol, 1.6 eq.) was added and the reaction mixture cooled in an ice/H_2O bath. A solution of $NaNO_2$ (520 mg, 7.50 mmol, 1.05 eq.) in H_2O (3 mL) was added dropwise. After stirring for about 30 min at room temperature, the pink precipitate that formed was filtered off to give 5-nitrosopyrimidine-2,4,6-triamine (987 mg, 6.40 mmol, 89%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.26 (d, *J* = 5.1 Hz, 1H), 8.15 (s, 1H), 7.75 (d, *J* = 5.1 Hz, 1H), 7.35 (s, 1H), 7.19 (s, 2H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.52, 165.32, 151.43, 138.04. **HRMS** (ESI+): calc. for $[C_4H_7N_6O]^+$ 155.0676, found: 155.0676 $[M+H]^+$

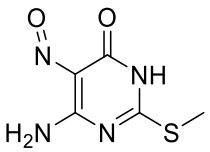
2-methyl-5-nitrosopyrimidine-4,6-diamine (4d)



Synthetic reference: Acetamidine (**2d**) salt of (hydroxyimino)malononitrile (**3**) (2.00 g, 13.5 mmol) was heated in 5-Ethyl-2-methylpyridine (10 mL) at 180°C for 20 min. After cooling to room temperature, the mixture was diluted with EtOH. After standing the product precipitated as green solid that was filtered off to give 2-methyl-5-nitrosopyrimidine-4,6-diamine (1.70 g, 11.5 mmol, 85%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.04 (d, *J* = 3.2 Hz, 1H), 8.97 (s, 1H), 8.36 (d, *J* = 3.2 Hz, 1H), 7.96 (s, 1H), 2.20 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 175.20, 166.38, 146.76, 139.94, 26.83. **HRMS** (ESI-): calc. for $[C5H6N5O]^-$ 152.0578, found: 152.0578 $[M-H]^-$

6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (4e)



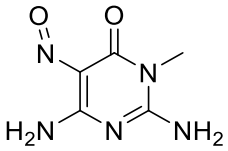
Prebiotic Synthesis: 2-(methylthio)-5-nitrosopyrimidine-4,6-diamine (1.00 g, 5.40 mmol, 1 eq.) was stirred in HCl (6 M, 500 mL) for 7 days at 0-8 °C. The sample was filtered and freeze dried and the yellow residue was taken up in a minimal amount of H_2O. The color of the residue changed to blueish/violet immediately and was filtered off to give 6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (760 mg, 4.10 mmol, 76%).

For analytic reasons 20 mg of the sample was purified by dissolving it in NH_4OH (30%, 3 mL). AcOH was slowly dropped into the solution until pH ~8-9 was reached. The product crystalized and was filtered off to give pink colored shiny crystals (12.0 mg). These crystals are treated with a minimal volume of AcOH (10%). The now blueish/violet crystals are filtered off. The sample is NMR clean but shows still 15% of starting material.

Synthetic standard: 6-amino-2-(methylthio)pyrimidin-4(3H)-one (1.00 g, 6.40 mmol, 1 eq.) was dissolved in H_2O containing NaOH (210 mg, 5.30 mmol, 0.8 eq.). To the dissolved pyrimidine was added $NaNO_2$ (480 mg, 7.00 mmol, 1.1 eq.). The solution was acidified with AcOH (660 µl, 11.6 mmol, 1.8 eq.), causing immediately a white precipitate that turned blue overnight upon standing. The blue precipitate was filtered off to give 6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (850 mg, 4.60 mmol, 72%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 12.72 (s, 1H), 11.26 (s, 1H), 9.08 (s, 1H), 2.53 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 168.88, 161.57, 147.29, 143.44, 13.52. **HRMS** (ESI-): calc. for $[C_5H_5N_4O_2S]^-$ 185.0139, found: 185.0138 $[M-H]^-$

2,6-diamino-3-methyl-5-nitrosopyrimidin-4(3H)-one (4f)



Prebiotic synthesis: 6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine (500 mg, 3.00 mmol, 1eq.) was dissolved in 0.5 M HCl (50 mL) and stirred at room temperature overnight. The reaction mixture was neutralized with NH_4OH upon which an reddish precipitate formed immediately. The reddish product was filtered off after cooling to give 2,6-diamino-3-methyl-5-nitrosopyrimidin-4(3H)-one (330 mg, 1.95 mmol, 65% over 2 steps from 1-methylguanidine (**2a**) salt of (hydroxyimino)malononitrile (**3**)).

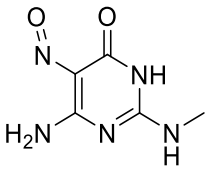
The yield of **4f** could only be determined over 2 steps, because we used impure **4a** directly from the prebiotic reaction. Therefore, the yield from **4a** to **4f** was determined by using synthetic

6-imino-1-methyl-5-nitroso-1,6-dihydropyrimidine-2,4-diamine triflouro-acetic acid (500 mg, 1.80 mmol) that was dissolved in 0.5 M HCl (30 mL) and stirred at room temperature overnight. The reaction mixture was neutralized with NH_4OH upon which a pink precipitate formed immediately. The pink product was filtered off after cooling to give 2,6-diamino-3-methyl-5-nitrosopyrimidin-4(3H)-one (295 mg, 1.74 mmol, 97%).

Synthetic standard: 6-amino-2-methoxy-3-methyl-5-nitrosopyrimidin-4(3*H*)-one (2.50 g, 13.6 mmol) was suspended in 20% NH_4OH (35 mL) and stirred at room temperature for 1 hour and 45 min. The pink solid was collected by filtration and washed with H_2O, EtOH and Et_2O, to give 2,6-diamino-3-methyl-5-nitrosopyrimidin-4(3H)-one (2.07 g, 12.2 mmol, 90%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.83 (d, *J* = 4.9 Hz, 1H), 8.25 (s, 1H), 8.07 (d, *J* = 4.9 Hz, 1H), 7.80 (s, 1H), 3.32 (s, 3H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 162.10, 157.33, 151.10, 142.39, 28.35. **HRMS** (ESI+): calc. for $[C_5H_8N_5O_2]^+$ 170.0673, found: 170.0672 $[M+H]^+$

6-amino-2-(methylamino)-5-nitrosopyrimidin-4(3H)-one (4g)

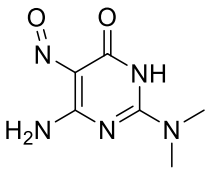


6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (**4e**, 400 mg, 2.16 mmol, 1 eq.) was dissolved in $MeNH_2$ (aq.) (1%, 21 mL) and stirred for 4 h at 8°C. The color changed from blueish/violet to dark red. The mixture was then neutralized with AcOH. The product precipitated as an orange/brownish solid and was filtered off to give 6-amino-2-(methylamino)-5-nitrosopyrimidin-4(3H)-one (258 mg, 1.52 mmol, 70%).

To determine the yield for the transformation of **4e** to **4g** from completely pure material, synthetic 6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (100 mg, 0.54 mmol, 1 eq.) was dissolved in $MeNH_2$ (aq.) (1%, 6 mL) and stirred for 4 h at 8°C. The color changed from violet to dark red. The mixture was then neutralized with AcOH. The product precipitated as an orange solid and was filtered off to give 6-amino-2-(methylamino)-5-nitrosopyrimidin-4(3H)-one (84 mg, 0.50 mmol, 93%).

Due to insolubility, a **^{13}C-NMR** was not possible **^{1}H-NMR** (400 MHz, DMSO-d_6) δ = 11.25 (s, 1H), 10.90 (s, 1H), 8.53 (s, 1H), 7.21 (q, *J* = 4.7 Hz, 1H), 2.85 (d, *J* = 4.7 Hz, 4H). **HRMS** (ESI+): calc. for $[C_5H_8N_5O_2]^+$ 170.0673, found: 170.0672 $[M+H]^+$

6-amino-2-(dimethylamino)-5-nitrosopyrimidin-4(3H)-one (4h)

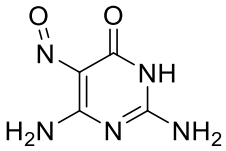


6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (**4e**, 350 mg, 1.90 mmol, 1 eq.) was dissolved in Me_2NH (aq.) (1%, 21 mL) and stirred for 4 h at 8°C. The color changed from blueish/violet to dark red. The mixture was then neutralized with AcOH. The product precipitated to form a reddish/pinkish paste like mixture. The precipitate was filtered off to give 6-amino-2-(methylamino)-5-nitrosopyrimidin-4(3H)-one (256 mg, 1.46 mmol, 73%).

To determine the yield for the transformation of **4e** to **4h** from completely pure material, synthetic 6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (1.00 g, 5.40 mmol, 1 eq.) was stirred in Me_2NH (aq.) (1%, 60 mL) for 4 h at 8°C. The mixture was then neutralized with AcOH. The product precipitated to form a pink paste like mixture. The precipitate was filtered off to give 6-amino-2-(dimethylamino)-5-nitrosopyrimidin-4(3H)-one (790 mg, 4.30 mmol, 80%). If desired, the product can be recrystallized from H_2O (50 mg in ca. 10 mL)

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.96 (s, 2H), 8.29 (d, J = 5.2 Hz, 1H), 3.13 (s, 6H). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 163.01, 154.49, 151.52, 141.75, 37.93. **HRMS** (ESI+): calc. for $[C_6H_{10}N_5O_2]^+$ 184.0829, found: 184.0829 $[M+H]^+$

2,6-diamino-5-nitrosopyrimidin-4(3H)-one (4i)

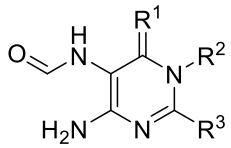


From synthetic 6-amino-2-(methylthio)-5-nitrosopyrimidin-4(3H)-one (**4e**): (40 mg, 0.21 mmol, 1 eq.) was heated in NH_4OH (2 M, 3 mL) for 3 days at 60 °C. The suspension was neutralized with AcOH and the product was filtered off, to give pink 2,6-diamino-5-nitrosopyrimidin-4(3H)-one (25.0 mg, 0.16 mmol, 76%).

From 5-nitroso-1,6-dihydropyrimidine-2,4,6-triamine (**4c**): (500 mg, 3.25 mmol, 1 eq.) was suspended in HCl solution (0.5 M, 50 mL) and stirred overnight at room temperature. The mixture was neutralized with NH_4OH and the product was filtered off, to give 2,6-diamino-5-nitrosopyrimidin-4(3H)-one (450 mg, 2.90 mmol, 89%).
Note: In case the conversion is inefficient at room temperature, the solution can also be heated up to 90°C.

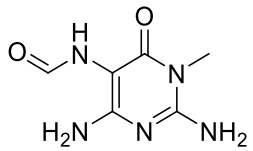
NMR and ESI experiments are not possible due to insolubility even in warm DMSO. Therefore the product was confirmed by further reaction to the FaPyG compound as described below. **IR** (cm^{-1}): 3284 (s), 3151 (s), 3111 (s), 1701 (s), 1625 (s), 1580 (w), 1497 (m), 1459 (m), 1359 (m), 1314 (s), 1258 (s), 1145 (s), 1032 (m), 988 (m), 810 (m) 782 (s), 733 (m), 702 (m), 683 (s). **HRMS** (EI+): calc. for $[C_4H_5N_5O_2]^+$ 155.0438, found: 155.0436 $[M]^+$

General procedure for the formation and isolation of FaPy compounds from nitrosopyrimidines under prebiotic conditions (5a-h)



The corresponding nitrosopyrimidine (250 mg) was suspended in dilute HCOOH (20%, 30 mL) in water in the presence of elementary Ni powder (1.00 g). (Note: all tranformations except for **5b** were also done with Fe powder) The reaction mixture was stirred at 70 °C in a sealed 100 mL Ace pressure tube. The reaction time is indicated below for the corresponding compound. The solvent was then evaporated completely until dryness. The remaining solid was taken up in water (50 mL) and the pH was adjusted with solid K_2CO_3 until pH 9-10. Precipitation occurred immidiately. The mixture was further stirred for about 30 min at rt and the precipitate filtered off through celite. Note: if the pH is < 8 after filtration add some more K_2CO_3 and filter again. The clear solution is then concentrated (usually to about <1/10 of the volume) until the product precipitated. After cooling the product was filtered off to give the yield as indicated below. All FaPy compounds can be recrystallized from H_2O (50-100 mg in 5 mL of H_2O depending on the FaPy compound).

***N*-(2,4-diamino-1-methyl-6-oxo-1,6-dihydropyrimidin-5-yl)formamide (FaPym^{1}G, 5a)**

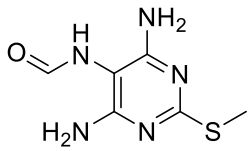


The nitrosopyrimidine **4f** (250 mg, 1.48 mmol) was reacted for 45 min. Isolated yield: 137 mg (0.75 mmol, 51%). The NMR shows cis/trans rotamers (~1.5:1)

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.54 (d, cis, *J* = 1.6 Hz, 1H; NH), 8.02 (d, cis, *J* = 1.6 Hz, 1H; CHO), 7.93 (d, trans, *J* = 11.7 Hz, 1H; NH), 7.73 (d, trans, *J* = 11.7 Hz, 1H; CHO), 6.78 (s, trans, 2H; $C2NH_2$), 6.71 (s, cis, 2H; $C2NH_2$), 5.92 (s, trans 2H; $C4NH_2$), 5.69 (s, 2H; $C4NH_2$), 3.17 (s, cis/trans, 6H; 2 CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 167.45 (trans, CHO), 161.07

(cis, CHO), 160.82 (trans, C6), 160.12 (cis, C6), 159.43 (trans, C4), 158.55 (cis C4), 154.45 (trans, C2), 154.22 (cis, C2), 88.98 (trans, C5), 88.94 (cis, C5), 28.35 (trans, CH_3), 28.19 (cis, CH_3). **HRMS** (ESI+): calc. for $[C_6H_{10}N_5O_2]^+$ 184.0829, found: 184.0830 $[M+H]^+$.

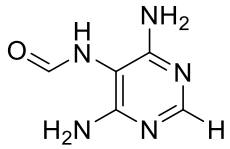
***N*-(4,6-diamino-2-(methylthio)pyrimidin-5-yl)formamide (FaPyms^{2}A, 5b)**



The nitrosopyrimidine **4b** (250 mg, 1.35 mmol) was reacted for 2 h in the presence of Ni. Isolated yield: 148 mg (0.74 mmol, 55%, over 2 steps from ethylthioamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**)). The NMR shows cis/trans rotamers (~5:1).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.66 (s, cis, 1H; NH), 8.23 (d, trans, *J* = 11.5 Hz, 1H; NH), 8.07 (s, cis, 1H; CHO), 7.74 (d, trans, *J* = 11.3 Hz, 1H; CHO), 6.18 (s, trans, 4H: 2 NH_2), 6.01 (s, cis, 4H; 2 NH_2), 2.35 (s, cis/trans, 6H, 2 CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 167.46 (trans, C2), 166.92 (cis, C2), 166.25 (trans, CHO), 161.52 (cis, CHO), 161.09 (trans, C4 and C6), 159.77 (cis, C4 and C6), 91.39 (trans, C5), 91.00 (cis, C5), 13.78 (cis/trans, 2 CH_3). **HRMS** (ESI+): calc. for $[C_6H_9N_5OS]^+$ 200.0601, found: 200.0600 $[M+H]^+$.

***N*-(4,6-diaminopyrimidin-5-yl)formamide (FaPyA, 5c)**



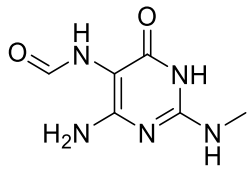
The nitrosopyrimidine **4b** (500 mg, 2.70 mmol) was reacted for 2 h. **5b** was the only product. Isolated yield of **5b**: 300 mg (1.50 mmol, 56%). The isolated **5b** (250 mg, 1.35 mmol, 1 eq.) was directly reacted under the same conditions for 7 d. H_2 was bubbled through the mixture before the start of reaction. **5c** was the only product. Isolated yield of **5c**: 106 mg (0.69 mmol, 51%). The NMR shows cis/trans rotamers (~5:1).

The overall yield without **5b** as isolated intermediate was determined by using synthetic nitrosopyrimidine **4b** (250 mg, 1.35 mmol) that was reacted for 7 d in the presence of Ni. H_2 is additionally bubbled through the solution for 2-3 minutes before the start of reaction. Isolated yield: 95 mg (0.62 mmol, 46%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 9.01 (s, cis, 1H; NH), 8.40 (d, trans, *J* = 11.5 Hz, 1H; NH), 8.09 (s, cis, 1H; CHO), 7.77 (d, trans, *J* = 11.5 Hz, 1H; CHO), 7.75 (s, trans, 1H; C2H), 7.74

(s, cis, 1H; C2H), 6.17 (s, trans, 4H; C4NH_2 and C6NH_2), 5.99 (s, cis, 4H; C4NH2 and C6NH2). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.02 (trans, CHO), 161.28 (cis, CHO; trans C4 and C6), 159.86 (cis, C4 and C6), 156.46 (trans, C2), 156.08 (cis, C2), 94.54 (trans, C5), 94.37 (cis, C5). **HRMS** (ESI+): calc. for: $[C5H8N5O]^+$ 154.0723, found: 154.0725 $[M+H]^+$

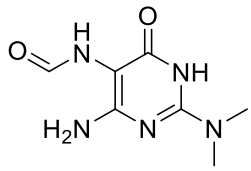
***N*-(4-amino-2-(methylamino)-6-oxo-1,6-dihydropyrimidin-5-yl)formamide (FaPym^{2}G, 5d)**



The nitrosopyrimidine **4g** (250 mg, 1.48 mmol) was reacted for 1 h. Isolated yield: 134 mg (0.74 mmol, 50%). The NMR shows cis/trans rotamers (~1.4:1)

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.11 (s, cis/trans, 2H; $NH_{arom.}$), 8.49 (d, cis, *J* = 1.6 Hz, 1H; NH), 8.00 (d, cis *J* = 1.6 Hz, 1H; CHO), 7.86 (d, trans, *J* = 11.7 Hz, 1H; NH), 7.73 (d, trans, *J* = 11.7 Hz, 1H; CHO), 6.19 (br s, 1H; N<u>H</u>CH_3), 6.14 (q, cis, *J* = 4.7 Hz, 1H, N<u>H</u>CH_3), 6.08 (s, trans, 1H; NH_2), 5.87 (s, cis, 2H; NH_2), 2.73 (d, cis/trans, *J* = 4.7 Hz, 6H; 2 CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 167.33 (trans, CHO), 161.60 (cis, C6), 161.24 (trans, C6), 161.10 (cis, CHO), 160.12 (cis, C4), 160.09 (trans, C4), 153.31 (trans, C2), 153.21 (cis, C2), 89.03 (trans, C5), 88.91 (cis, C5), 27.83 (cis/trans, 2 CH_3). **HRMS** (ESI+): calc. for $[C_6H_{10}N_5O_2]^+$ 184.0829, found: 184.0830 $[M+H]^+$.

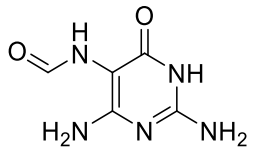
***N*-(4-amino-2-(dimethylamino)-6-hydroxypyrimidin-5-yl)formamide (FaPym$^2_2$G, 5e)**



The nitrosopyrimidine **4h** (250 mg, 1.37 mmol) was reacted for 1 h. Isolated yield: 108 mg (0.55 mmol, 40%). The NMR shows cis/trans rotamers (~1.7:1).

^{1}H-NMR (800 MHz, DMSO-d_6) δ = 10.23 (br, cis/trans, 2H; $NH_{arom.}$), 8.54 (d, cis, *J* = 1.7 Hz, 1H; NH), 8.00 (d, cis, *J* = 1.7 Hz, 1H; CHO), 7.87 (d, trans, *J* = 11.7 Hz, 1H; NH), 7.74 (d, trans, *J* = 11.7 Hz, 1H; CHO), 6.05 (s, trans, 2H; NH_2), 5.83 (s, cis, 2H; NH_2), 2.98 (s, cis/trans, 12H; 4 CH_3). **^{13}C-NMR** (201 MHz, DMSO-d_6) δ 167.29 (trans, CHO), 161.96 (trans, C6), 161.02 (cis, C6), 161.00 (cis, CHO), 160.81 (trans, C4), 159.50 (cis, C4), 153.03 (cis/trans, C2), 88.70 (trans, C5), 88.67 (cis, C5), 37.52 (cis/trans, 4 CH_3). **HRMS** (ESI-): calc. for $[C_7H_{10}N_5O_2]^-$ 196.0840, found: 196.0840 $[M-H]^-$.

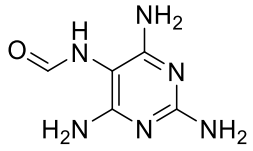
***N*-(2,4-diamino-6-oxo-1,6-dihydropyrimidin-5-yl)formamide (FaPyG, 5f)**



The nitrosopyrimidine **4i** (250 mg, 1.61 mmol) was reacted for 1.5 h. Isolated yield: 153 mg (0.91 mmol, 57%). The NMR shows cis/trans rotamers of the formamide group (~1.4:1).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 10.52 (s, cis/trans, 2H; $NH_{arom.}$), 8.50 (d, cis, *J* = 1.6 Hz, 1H; NH), 7.99 (d, cis, *J* = 1.6 Hz, 1H; CHO), 7.85 (d, trans, *J* = 11.7 Hz, 1H; NH), 7.72 (d, trans, *J* = 11.7 Hz, 1H; CHO), 6.47 (s, trans, 2H; $C2NH_2$), 6.36 (s, cis, 2H; $C2NH_2$), 5.98 (s, trans, 2H; $C4NH_2$), 5.75 (s, cis, 2H; $C4NH_2$). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.96 (trans, CHO), 161.50 (trans, C6), 161.30 (cis, CHO), 160.60 (cis, C6), 160.18 (trans, C4), 159.90 (cis, C4), 154.00 (trans, C2), 153.88 (cis, C2) , 88.72 (cis/trans, C5). **HRMS** (ESI-): calc. for $[C_5H_6N_5O_2]^-$ 168.0527, found: 168.0527 $[M-H]^-$.

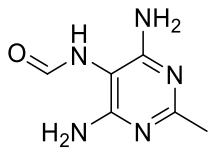
***N*-(2,4,6-triaminopyrimidin-5-yl)formamide (FaPyDA, 5g)**



The nitrosopyrimidine **4c** (250 mg, 1.62 mmol) was reacted overnight. Isolated yield: 144 mg (0.86 mmol, 53%, over 2 steps from guanidine (**2b**) salt of (hydroxyimino)malononitrile (**3**)). The NMR shows cis/trans rotamers of the formamide group (~1.6:1).

^{1}H-NMR (400 MHz, DMSO-d6) δ = 8.49 (s, cis, 1H; NH), 8.06 (d, trans J=11.7, 1H; NH), 8.05 (s, cis, 1H; CHO), 7.69 (d, trans J=11.7, 1H; CHO), 5.68 (s, trans, 2H; $C2NH_2$), 5.53 (s, trans, 4H; $C4NH_2$ and $C6NH_2$), 5.51 (s, cis, 4H; $C4NH_2$ and $C6NH_2$), 5.43 (s, cis, 2H; $C2NH_2$) **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.57 (trans, CHO), 161.67 (trans, C4 and C6), 161.23 (cis, CHO), 161.14 (trans, C2), 161.09 (cis, C2), 160.13 (cis, C4 and C6), 86.79 (trans, C5), 86.54 (cis, C5). **HRMS** (ESI+): calc. for: $[C5H9N6O]^+$ 169.0832, found: 169.0832 $[M+H]^+$.

***N*-(4,6-diamino-2-methylpyrimidin-5-yl)formamide (FaPym^{2}A, 5h)**



The nitrosopyrimidine (250 mg, 1.63 mmol) was reacted for 30 min. Isolated yield: 173 mg (1.04 mmol, 64%, over 2 steps from acetamidine (**2b**) salt of (hydroxyimino)malononitrile (**3**)). The NMR shows cis/trans rotamers (~5:1).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.68 (s, cis, 1H; NH), 8.24 (d, trans, J = 11.6 Hz, 1H; NH), 8.07 (d, cis, J = 1.2 Hz, 1H; CHO), 7.73 (d, trans, J = 11.6 Hz, 1H; CHO), 6.04 (s, trans, 4H, 2 NH_2), 5.86 (s, cis, 4H; 2 NH_2), 2.12 (s, cis/trans, 6H; 2 CH_3). **^{13}C-NMR** (101 MHz, DMSO-d_6) δ = 166.17 (trans, CHO), 164.70 (trans, C2), 164.23 (cis, C2), 161.35 (cis, CHO), 161.29 (trans, C4 and C6), 159.89 (cis, C4 and C6), 92.44 (trans, C5), 92.16 (cis, C5), 25.38 (cis/trans, 2 CH_3). **HRMS** (ESI+): calc. for $[C_6H_9N_5O]^+$ 168.0880, found: 168.0879 $[M+H]^+$.

Prebiotic nucleoside formation procedure (6a-h)

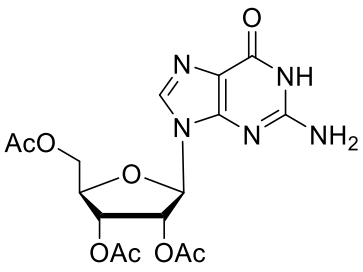
Nucleoside formation using ribose and FaPy (**5a-h**)

Ribose (56.5 mg, 0.375 mmol, 15 eq.) was thoroughly ground up with the corresponding FaPy compound **5a-h** (0.025 mmol, 1 eq.) and heated to 100 °C for 8 h in an oven. The remaining reaction mixture was taken up in basic solution (3 mL, 0.5 M Et_3N) and heated in a sealed tube (ACE 15 mL pressure tube) at 100 °C for several days (see below). 100 µL aliquots were removed and diluted to 1 mL. The aliquots were used for LC-MS analysis.

5a: reacted for 7d. **5b:** reacted for 14d**. 5c:** reacted for 21d. **5d:** reacted for 4d. **5e:** reacted for 4d. **5f:** reacted for 2d. **5g:** reacted for 21d. **5h:** reacted for 28d.

Synthesis of synthetic nucleoside standards

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-oxo-1,6-dihydro-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (7)

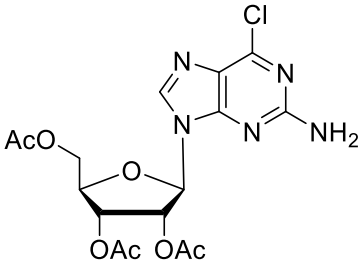


The reaction was carried out slightly modified to Robins *et al.*.[3,4]

Acetic anhydride (20 mL, 211 mmol, 12 eq.) was added to a mixture of guanosine (5.00 g, 18.6 mmol, 1 eq.) in DMF (12 mL) and pyridine (10 mL). After stirring for 1.5 h at 75 °C ethanol was added to quench the reaction. The mixture was filtered, the solvent removed *in vacuo* and the residue was recrystallized from isopropanol. The crystals were washed with cold isopropanol in order to obtain the desired product as a white solid (6.37 g, 15.6 mmol, 88%).

^{1}H-NMR (400 MHz, d6-DMSO) δ = 10.73 (s, 1H, NH), 7.93 (s, 1H, HC_{Ar}), 6.54 (s, 2H,NH_2), 5.98 (d, *J* =6.1 Hz, 1H, HC1'), 5.78 (t, *J*=6.1 Hz, 1H, HC2'), 5.49 (dd, *J* =5.9 Hz, *J* =4.0 Hz, 1H, HC3'), 4.72–4.08 (m, 3H, HC4', HC5'), 2.11 (s, 3H, $COCH_3$), 2.04 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$). **^{13}C-NMR** (101 MHz, d6-DMSO) δ = 170.13 (CH3CO), 169.48 (CH3CO), 169.31 (CH3CO), 156.64 (C6), 153.91 (C2), 151.13 (C4), 135.65 (C8), 116.82 (C5), 84.36 (C1'), 79.55 (C4'), 72.04 (C2'), 70.32 (C2'), 63.10 (C5'), 20.57 (CH3CO), 20.42 (CH3CO), 20.22 (CH3CO). **HRMS** (ESI$^+$): calc. for $[C_{16}H_{20}N_5O_8]^+$ 410.1306, found: 410.1307 $[M+H]^+$.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-chloro-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (8)



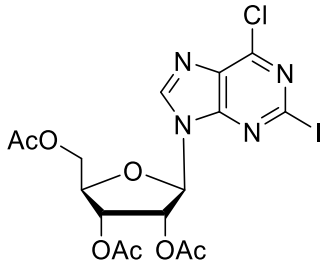
The reaction was carried out slightly modified to Robins *et al.*.[3,4]

Compound **7** (1.00 g, 2.44 mmol, 1 eq.) was dissolved in 80 mL dry acetonitrile. Triethylamine hydrochloride (0.81 g, 4.89 mmol, 2 eq), *N,N*-dimethylamine (0.31 mL, 2.44 mmol, 1 eq.) and phosphoryl chloride (1.37 mL, 14.66 mmol, 6 eq.) were added to the solution and the reaction mixture was heated to 100°C for 15 min. The solvent was removed *in vacuo* and the residue was dissolved in 30 mL water. The aqueous phase was extracted with chloroform. The combined organic layers were washed with $NaHCO_3$ and dried over $MgSO_4$ before the solvent was removed in vacuo. The crude product was purified by column chromatography (DCM : MeOH, 20:1) to afford the product as a white solid (0.56 g, 1.32 mmol, yield: 54%).

^{1}H-NMR (400 MHz, d_6-DMSO) *δ* = 8.36 (s, 1H, HC8), 7.08 (s, 2H, NH_2), 6.10 (d, ^3J=5.9 Hz, 1H, HC1'), 5.87 (t, *J*=5.9 Hz, 1H, HC2'), 5.53 (dd, *J*= 5.9 Hz, *J*=4.2 Hz, 1H, HC3'), 4.48 – 4.18 (m, 3H, HC4', H2C5'), 2.12 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$). **^{13}C-NMR** (101 MHz, d_6-DMSO) δ = 170.52 ($COCH_3$), 169.84 ($COCH_3$), 169.71 ($COCH_3$), 160.04 (C2), 153.87(C6), 150.03 (C4), 141.28 (C8), 123.42 (C5), 84.85 (C1'), 79.70 (C4'), 71.89 (C2'),

70.25 (C3'), 62.97(C5'), 20.55 ($COCH_3$), 20.41 ($COCH_3$), 20.22 ($COCH_3$). **HRMS** (ESI+): calc. for $[C_{16}H_{19}ClN_5O_7]^+$ 428.0968, found: 428.0969 $[M+H]^+$.

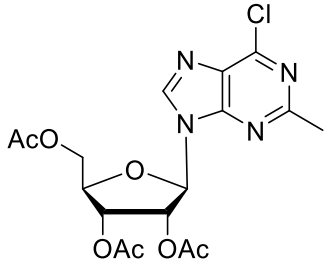
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(6-chloro-2-iodo-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (9) [5]



Nucleoside **8** (390 mg, 0.91 mmol, 1 eq), CuI (190 mg, 0.99 mmol, 1.1 eq.) and I_2 (230 mg, 0.91 mmol, 1 eq), was dissolved in dry THF (4.5 mL). CH_2I_2 (0.73 mL, 9.07 mmol, 10 eq.), and isoamyl nitrite (318 mg, 2.72 mmol, 3 eq.) were added and the solution was heated to 70°C. After 3 h the solvent was removed and the crude product was purified by column chromatography (DCM : MeOH, 200:1) to afford the product as a white solid (0.56 g, 1.32 mmol, yield: 54%).

^{1}H-NMR (400 MHz, d_6-DMSO) δ = 8.82 (s, 1H, HC8), 6.30 (d, ^3J = 4.9 Hz, 1H, HC1'), 5.89 (dd, ^3J = 5.9, 4.9 Hz, 1H, HC2'), 5.63 (t, J = 5.4 Hz, 1H, HC3'), 4.51 – 4.36 (m, 3H, H5', HC4'), 2.12 (s, 3H, $COCH_3$), 2.06 (s, 3H, $COCH_3$), 2.01 (s, 3H, $COCH_3$). **^{13}C-NMR** (101 MHz, d_6-DMSO) δ = 169.81 (COCH3), 169.66 (COCH3), 169.14 (COCH3), 151.98(C2), 148.79 (C6), 145.79 (C4), 131.45 (C8), 118.28 (C5), 84.85 (C1'), 86.00 (C4'), 69.59 (C2'), 69.42 (C3'), 62.42 (C5'), 20.38 ($COCH_3$), 20.18 ($COCH_3$), 20.05 ($COCH_3$). **HRMS** (ESI+): calc. for $[C_{16}H_{17}ClIN_4O_7]^+$ 538.9825, found: 538.9839 $[M+H]^+$.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(6-chloro-2-methyl-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (10)

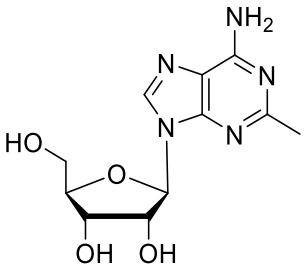


The reaction was carried out slightly modified to Yamazaki *et al.*.[6]

MeMgCl (2.56 M in THF, 0.22 mL, 573 µmol, 3 eq.) was added to $ZnCl_2$ (1.0 M in Et_2O, 0.29 mL, 286 µmol, 1.5 eq.) and) at 0°C and stirred for 3 h. This solution was diluted with THF (1mL) and added to a solution of **9** (103 mg, 192 µmol, 1 eq.) and $Pd(PPh_3)_4$ (22.1 mg, 19.1 µmol, 0.1 eq.) in dry THF (1.5 mL). The mixture was stirred for 10 min at 0°C and then at rt for additional 2 h. H_2O (10 mL) and DCM (5 mL) were added to the mixture. The phases were separated and the aqueous phase was extracted with DCM. The combined organic layers were washed with brine and dried over $MgSO_4$. The solvent was removed and the crude product was purified by column chromatography (DCM : MeOH, 20:1 → 10:1) to afford the product as a colorless foam (43.0 mg, 101 µmol, yield: 53%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ = 8.12 (s, 1H, HC8), 6.12 (d, J = 4.9 Hz, 1H, HC1'), 5.89 (dd, ^3J = 5.9, 4.9 Hz, 1H, HC2'), 5.65 (t, J = 5.1 Hz, 1H, HC3'), 4.47 – 4.25 (m, 3H, HC4', H2C5'), 2.74 (s, 3H, $C_{Ar}CH_3$), 2.09 (s, 3H, $COCH_3$), 2.03 (s, 6H, $COCH_3$). **^{13}C-NMR** (101 MHz, $CDCl_3$) δ = 170.43 ($COCH_3$), 169.70 ($COCH_3$), 169.49 ($COCH_3$), 163.25 (C2), 151.85 (C4), 151.16 (C6), 143.03 (C8), 130.24 (C5), 86.90 (C1'), 80.49 (C4'), 73.20 (C2'), 70.61 (C3'), 63.11 (C5'), 25.86 ($CArCH_3$), 20.87 ($COCH_3$), 20.69 ($COCH_3$), 20.55 ($COCH_3$). **HRMS** (ESI+): calc. for $[C_{17}H_{20}ClN_4O_7]^+$ 427.1015, found: 427.1024 $[M+H]^+$.

(2R,3R,4S,5R)-2-(6-amino-2-methyl-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol (m^{2}A, 11)



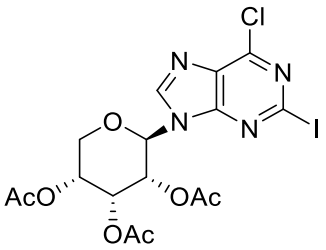
The reaction was carried out slightly modified to Yamazaki *et al.*.[7,8]

Nucleoside **10** (10.0 mg, 23.4 µmol, 1 eq.) was weighed into a pressure tube, into which ammonia was condensed. Subsequently the reaction mixture was stirred for 24 h at rt. Volatile compounds were removed in vacuo and the crude product was purified by HPLC to afford the product as a white solid (3.00 mg, 10.7 µmol, yield: 46%).

HPLC: Gradient: 100% HPLC buffer C, 0% HPLC buffer D → 80% HPLC buffer C, 20% HPLC buffer D in 45 min, retention time: 26.21 min, flow: 5 mL/min, column: VP 250/10 Nucleodur 100-5 C18ec.

^{1}H-NMR (400 MHz, CD_3OD) δ = 8.20 (s, 1H, HC8), 5.91 (d, *J* = 6.9 Hz, 1H, HC1'), 4.78 (dd, *J* = 6.8, 5.1 Hz, 1H HC2'), 4.31 (dd, *J* = 5.1, 1.0 Hz, HC3'), 4.18 (q, *J* = 2.1 Hz, 1H, HC4'), 3.90 (dd, ^{3}J = 12.6, 2.3 Hz, 1H, HaC5'), 3.74 (dd, *J* = 12.6, 2.3 Hz, 1H, HbC5'), 2.49 (s, 3H, $C_{Ar}CH_3$). **^{13}C-NMR** (101 MHz, CD_3OD) δ = 163.19 (C2), 150.36 (C6), 149.12 (C4), 141.71 (C8), 119.13 (C5), 91.37(C1'), 88.35 (C4'), 74.95 (C2'), 72.80 (C3'), 63.56 (C5'), 24.82 ($C_{Ar}CH_3$). **HRMS** (ESI+): calc. for $[C_{11}H_{16}N_5O_4]^+$ 282.1197, found: 282.1198 $[M+H]^+$.

(2S,3R,4R,5R)-2-(6-chloro-2-iodo-9H-purin-9-yl)tetrahydro-2H-pyran-3,4,5-triyl triacetate (12)

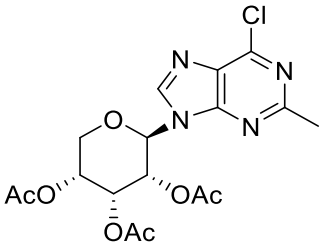


The reaction was carried out slightly modified to Vorbrüggen *et al.*.[9]

6-chloro-2-iodo-9H-purine (0.15 g, 0.53 mmol, 1.0 eq.) and acetylribose (323 mg, 1.02 mmol, 1.9 eq.) were suspended in dry acetonitrile (5.0 mL). *N*,*O*-Bis-(trimethylsilyl)-acetamide (195 mg, 0.96 mmol, 1.8 eq.) was added and the solution was heated to 60°C. After 30 min trimethylsilyltrifluoromethansolfonate (220 mg, 0.11 mmol, 2.0 eq.) was added and the solution turned from colorless to yellow-brown. The reaction mixture was cooled to rt, diluted with EtOAc and the organic layer was washed with $NaHCO_3$. The aqueous phase was washed with EtOAc and the combined organic layers were dried over $MgSO_4$. The solvent was removed and the crude product was purified by column chromatography (Hex : EtOAc, 1:1 → 1:4) to afford the product as a colorless foam (0.13 g, 0.23 mmol, yield: 44%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ = 8.31 (s, 1H, HC8), 5.98 (d, *J* = 9.7 Hz, 1H, HC1'), 5.92 – 5.63 (m, 1H, HC2'), 5.50 (dd, *J* = 9.7, 2.9 Hz, 1H, HC3'), 5.18 (ddd, *J* = 10.5, 6.0, 2.7 Hz, 1H H2C5'), 4.31 – 3.71 (m, 1H H2C4'), 2.26 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$), 1.80 (s, 3H, $COCH_3$). **^{13}C-NMR** (101 MHz, $CDCl_3$) δ = 170.40 ($COCH_3$), 169.82 ($COCH_3$), 169.42 ($COCH_3$), 149.88 (C2), 149.33 (C4), 149.22 (C6), 143.01 (C8), 123.58 (C5), 78.86 (C1'), 68.10 (C3'), 68.00 (C2'), 65.83 (C4'), 63.92 (C5'), 21.03 ($COCH_3$), 20.72 ($COCH_3$), 20.43 ($COCH_3$). **HRMS** (ESI+): calc. for $[C_{16}H_{17}ClIN_4O_7]^+$ 538.9825, found: 538.9835 $[M+H]^+$.

(2S,3R,4R,5R)-2-(6-chloro-2-methyl-9H-purin-9-yl)tetrahydro-2H-pyran-3,4,5-triyl triacetate (13)

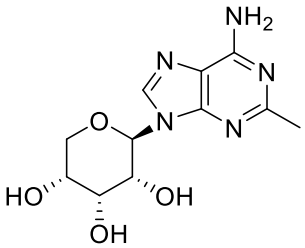


The reaction was carried out slightly modified to Yamazaki *et al.*.[6]

MeMgCl (2.56 M in THF, 0.22 mL, 557 µmol, 3 eq) was added to $ZnCl_2$ (1.0 M in Et_2O, 0.28 mL, 278 µmol, 1.5 eq.) and) at 0°C and stirred for 3 h. The solution was diluted with THF (1mL) and added to a solution of **12** (100 mg, 186 µmol, 1 eq.) and $Pd(PPh_3)_4$ (18.6 mg, 18.6 µmol, 0.1 eq.) in dry THF (1.5 mL). The mixture was stirred for 10 min at 0°C and then at rt for additional 2 h. H_2O (10 mL) and DCM (5 mL) were added to the mixture. The phases were separated and the aqueous phase was extracted with DCM. The combined organic layers were washed with brine and dried over $MgSO_4$. The solvent was removed and the crude product was purified by column chromatography (DCM : MeOH, 100:1) to afford the product as a colorless foam (36.0 mg, 84.4 µmol, yield: 45%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ = 8.15 (s, 1H, HC8), 6.00 (d, *J* = 9.6 Hz, 1H, HC1'), 5.86 (td, *J* = 2.9, 0.9 Hz, 1H, HC2'), 5.73 (dd, *J* = 9.7, 2.8 Hz, 1H, HC3'), 5.25 (ddd, *J* = 10.5, 6.0, 2.8 Hz, 1H, C4'), 4.14 – 3.96 (m, 2H, H2C5'), 2.28 (s, 3H, $COCH_3$), 2.06 (s, 3H, $COCH_3$), 1.80 (s, 3H, $COCH_3$). **^{13}C-NMR** (101 MHz, $CDCl_3$) δ = 170.40 ($COCH_3$), 169.82 ($COCH_3$), 169.42 ($COCH_3$), 149.88 (C2), 149.33 (C4), 149.22 (C6), 143.01 (C8), 123.58 (C5), 78.86 (C1'), 68.10 (C3'), 68.00 (C2'), 65.83 (C4'), 63.92 (C5'), 21.03 ($COCH_3$), 20.72 ($COCH_3$), 20.43 ($COCH_3$). **HRMS** (ESI+): calc. for $[C_{17}H_{20}ClIN_4O_7]^+$ 427.1015, found: 427.1020 $[M+H]^+$.

(2R,3R,4R,5R)-2-(6-amino-2-methyl-9H-purin-9-yl)tetrahydro-2H-pyran-3,4,5-triol (*p*-m^{2}A, 14)



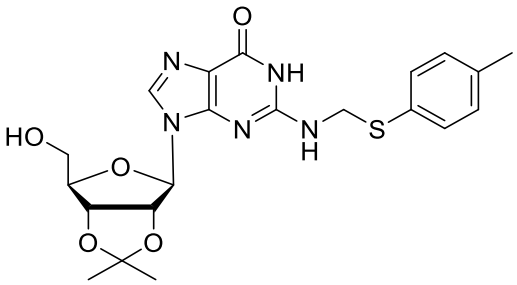
The reaction was carried out slightly modified to Yamazaki *et al.*.[7,8]

Nucleoside **13** (16.0 mg, 37.5 µmol, 1 eq.) was weighed into a pressure tube into which ammonia was condensed. Subsequently the reaction mixture was stirred for 24h at rt. Volatile compounds were removed *in vacuo* and the crude product was purified by HPLC to afford the product as a white solid (6.0 mg, 21.3 µmol, yield: 57%).

HPCL: Gradient: 100% HPLC buffer C, 0% HPLC buffer D → 80% HPLC buffer C, 20% HPLC buffer D in 45 min, retention time: 20.60 min, flow: 5 mL/min, column: VP 250/10 Nucleodur 100-5 C18ec.

^{1}H-NMR (400 MHz, CD_3OD) δ = 8.22 (s, 1H, HC8), 5.78 (d, ^3J = 8.7 Hz, 1H, HC1'), 4.26-4.22 (m, 2H, HC2'), 3.92-3.88 (m, 2H, H2C5'), 3.80-3.75 (m, 1H, C4'), 2.51 (s, 3H, $C_{Ar}CH_3$). **^{13}C-NMR** (101 MHz, CD_3OD) δ = 163.77 (C2), 156.95 (C6), 151.98 (C4), 140.90 (C8), 118.11 (C5), 81.67 (C1'), 72.81 (C4'), 70.38 (C2'), 68.18 (C3'), 66.60 (C5'), 25.27 ($C_{Ar}CH_3$).**HRMS** (ESI+): calc. for $[C_{11}H_{16}N_5O_4]^+$ 282.1197, found: 282.1198 $[M+H]^+$.

9-((3aR,4R,6R,6aR)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-2-(((p-tolylthio)methyl)amino)-1,9-dihydro-6H-purin-6-one (15)



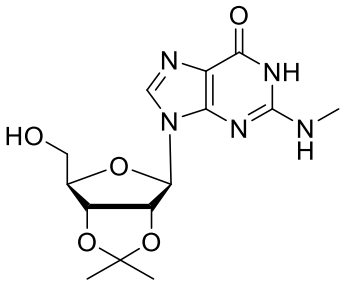
The reaction was carried out slightly modified to Bridson *et al*..[10]

2',3'-O-isopropylideneguanosine (2.50 g, 7.73 mmol, 1 eq.) and p-thiocresole (1.25 g, 10.1 mmol, 1.3 eq.) were suspended in EtOH (60.0 mL). Acetic acid (1.9 mL) and formaldehyde (2.0 mL) were added to the suspension and the mixture was heated to 90°C. After 14 h the solvent was removed *in vacuo* and the crude product was washed with EtOH to afford the product as a white solid (2.75 g, 5.88 mmol, yield: 76%).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.79 (s, 1H, $N_{Ar}H$), 7.95 (s, 1H, HC8), 7.41 – 7.31 (m, 2H, *J* = 8.0 Hz, HC_{Ar}), 7.17 (d, *J* = 8.0 Hz, 2H, HC_{Ar}), 6.01 (d, *J* = 2.6 Hz, 1H, HC1'), 5.36 (dd, *J* = 6.3, 2.7 Hz, 1H, HC2'), 5.02 (t, *J* = 5.5 Hz, 1H, OH), 4.95 (dd, *J* = 6.3, 3.0 Hz, 1H, HC3'), 4.84 (h, *J* = 6.7, 6.3 Hz, 2H, CH_2C_{Ar}), 4.13 (td, *J* = 5.5, 3.0 Hz, 1H, HC4'), 3.54 (hept, *J* = 5.7 Hz, 2H, HC5'), 2.28 (s, 3H, $C_{Ar}CH_3$), 1.54 (s, 3H, CH_3), 1.35 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, DMSO-d_6) δ = 154.14 (C6), 151.76 (C2), 149.98 (C4), 137.49 (C8), 136.99 ($C_{Ar}CH_3$), 136.31

($C_{Ar}S$), 131.43 ($C_{Ar}S$), 130.24 (2C, C_{Ar}), 118.02 (C5), 113.49 ($C(CH_3)_2$), 89.17 (C1'), 87.10 (C4'), 83.75 (C2'), 81.74 (C3'), 61.99 (C5'), 46.71 (CH_2C_{Ar}), 27.41 (CH_3), 25.52(CH_3), 21.04 ($C_{Ar}CH_3$). **HRMS** (ESI+): calc. for $[C_{21}H_{26}N_5O_5S]^+$ 460.1649, found: 460.1650 $[M+H]^+$.

9-((3aR,4R,6R,6aR)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-2-(methylamino)-1,9-dihydro-6H-purin-6-one (16)

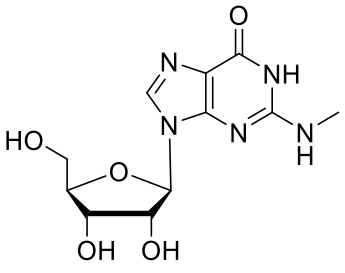


The reaction was carried out slightly modified to Bridson *et al*..[10]

Compound **15** (0.65 g, 1.41 mmol, 1 eq.) was dissolved in DMSO (9.0 mL). $NaBH_4$ (0.11 g, 2.81 mmol, 2 eq.) was added and the mixture was heated to 100°C. After cooling to rt the mixture was purified by column chromatography (DCM : MeOH, 9:1) to afford the product as a white solid (0.41 g, 1.22 mmol, yield: 86%).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.80 (s, 1H, $N_{Ar}H$), 7.90 (s, 1H, HC8), 6.46 (m, 1H, $NHCH_3$), 6.00 (d, *J* = 2.6 Hz, 1H, HC1'), 5.31 (dd, *J* = 6.2, 2.6 Hz, 1H, HC2'), 5.00 (t, *J* = 5.5 Hz, 1H, OH), 4.95 (dd, *J* = 6.2, 3.0 Hz, 1H, HC3'), 4.12 (td, *J* = 5.5, 3.0 Hz, 1H, HC4'), 3.53 (m, 2H, HC5'), 2.82 (d, *J* = 4.6 Hz, 3H, NCH_3), 1.53 (s, 3H, CH_3), 1.32 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, DMF-d_7) δ = 157.49 (C2), 154.08 (C6), 150.98 (C4), 136.86 (C8), 117.84 (C5), 113.59 ($C(CH_3)_2$), 89.79 (C1'), 87.52 (C4'), 84.22 (C2'), 82.17 (C3'), 62.50 (C5'), 27.88 (NCH_3), 27.03 (CH_3), 25.06 (CH_3). **HRMS** (ESI+): calc. for $[C_{14}H_{20}N_5O_5]^+$ 338.1457, found: 338.1459 $[M+H]^+$.

9-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-(methyl-amino)-1,9-dihydro-6H-purin-6-one (m^{2}G, 17)

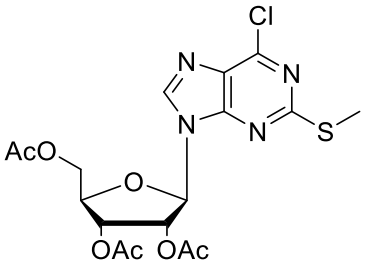


The reaction was carried out slightly modified to Ubiali *et al.*.[11]

The protected nucleoside **16** (0.28 g, 0.83 mmol, 1 eq.) was dissolved in H_2O (25.0 mL) and TFA (25.0 mL). After stirring for 1.5 h the solvent was removed in vacuo. The residue was suspended in Aceton and H_2O. The solvent was removed in vacuo and the raw product was recrystalized in EtOH/ H_2O to yield the product as a white solid (0.16 mg, 0.54 mmol, yield: 65%).

^{1}H-NMR (400 MHz, CD_3OD) δ = 10.77 (s, 1H, $N_{Ar}H$), 7.95 (s, 1H, HC8), 6.28 (d, ^3J = 5.0 Hz, 1H, NH), 5.73 (d, ^3J = 6.0 Hz, 1H, HC1'), 4.52 (t, J = 5.5 Hz, 1H, HC2'), 4.12 (dd, J = 5.1, 3.4 Hz, 1H, HC3'), 3.88 (d, J = 3.9 Hz, 1H, HC4'), 3.75 – 3.45 (m, 2H, HC5'), 2.81 (d, ^3J = 4.7 Hz, 3H, NCH_3). **^{13}C-NMR** (101 MHz, CD_3OD) δ = 156.77 (C6), 153.29 (C2), 150.97 (C4), 136.15 (C8), 116.65 (C5), 86.79 (C1'), 85.31 (C4'), 73.39 (C2'), 70.55 (C3'), 61.60 (C5'), 27.74 (NCH_3). **HRMS** (ESI+): calc. for $[C_{11}H_{16}N_5O_4]^+$ 298.1146, found: 298.1142 $[M+H]^+$.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(6-chloro-2-(methylthio)-9H-purin-9-yl)tetrahydro-furan-3,4-diyl diacetate (18)

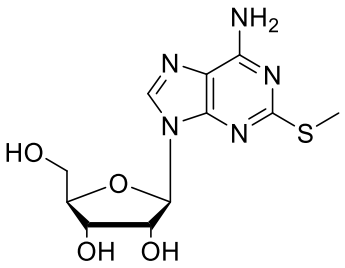


The reaction was carried out slightly modified to Kierzek *et al*..[12]

Nucleoside **8** (2.00 g, 4.68 mmol, 1 eq.) was dissolved in ACN (3.5 mL). Dimethylsulfide (4.40 g, 46.8 mmol, 10 eq.) and isopentenylnitrite (0.95 g, 9.35 mmol, 2 eq.) were added to the solution and the mixture was heated to 65°C. After 1 h, 2 more equivlents of isopentenylnitrite (0.95 g, 9.35 mmol, 2 eq.) were added. After 30 min the solvent was removed *in vacuo* and the crude product was purified by column chromatography (DCM : MeOH, 100:1) to afford the product as a yellow foam (1.59 g, 3.47 mmol, yield: 74%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.70 (s, 1H, HC8), 6.30 (d, J = 4.1 Hz, 1H, HC1'), 6.04 (dd, J = 6.0, 4.2 Hz, 1H), 5.80 – 5.62 (m, 1H, HC2'), 4.48 – 4.17 (m, 4H, HC3', HC4', HC5'), 2.63 (s, 3H, SCH_3), 2.11 (s, 3H, $COCH_3$), 2.08 (s, 3H, $COCH_3$), 1.96 (s, 3H, $COCH_3$). **^{13}C-NMR** (151 MHz, $CDCl_3$) δ = 170.2 (CH_3CO), 169.4 (CH3CO), 169.2 (CH3CO), 167.4 (C2), 151.9 (C6), 151.3 (C4), 142.1 (C8), 129.1 (C5), 87.0 (C1´), 80.0 (C4´), 72.9 (C2´), 70.1 (C3´), 62.7 (C5´), 20.7 (CH_3CO), 20.5 (CH_3CO), 20.4 (CH_3CO), 14.8 (SCH_3). **HRMS** (ESI^+): calc. for $[C_{17}H_{20}ClN_4O_7S]^+$ 459.0738, found: 459.0736 $[M+H]^+$.

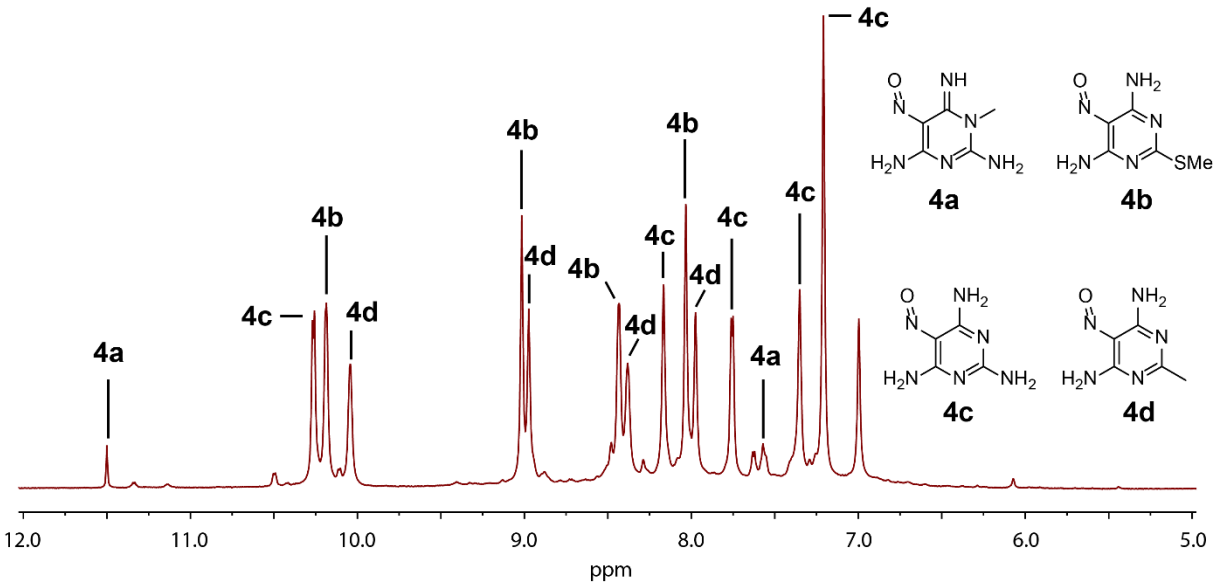
(2R,3R,4S,5R)-2-(6-amino-2-(methylthio)-9H-purin-9-yl)-5-(hydroxymethyl)tetra-hydrofuran-3,4-diol (ms^{2}A, 19)



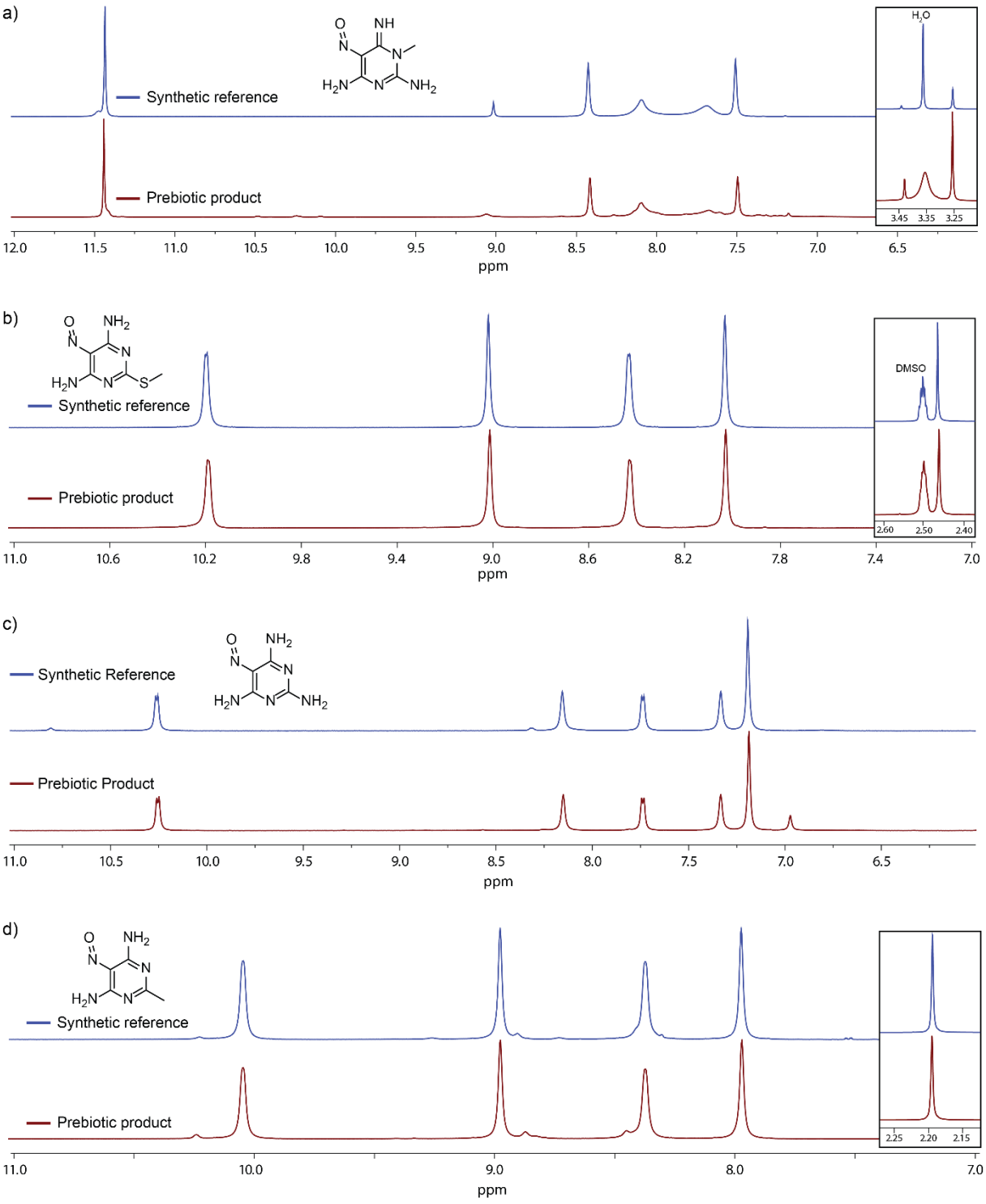
The reaction was carried out slightly modified to Yamazaki *et al.*.[7,8]

Nucleoside **18** (50.0 mg, 110 µmol, 1 eq.) was condensed into a pressure tube with ammonia and stirrerd for 24 h at rt. The solvent was removed and the crude product was recrystallized in EtOH to afford the product as a white solid (26.0 mg, 79.4 µmol, yield: 73%).

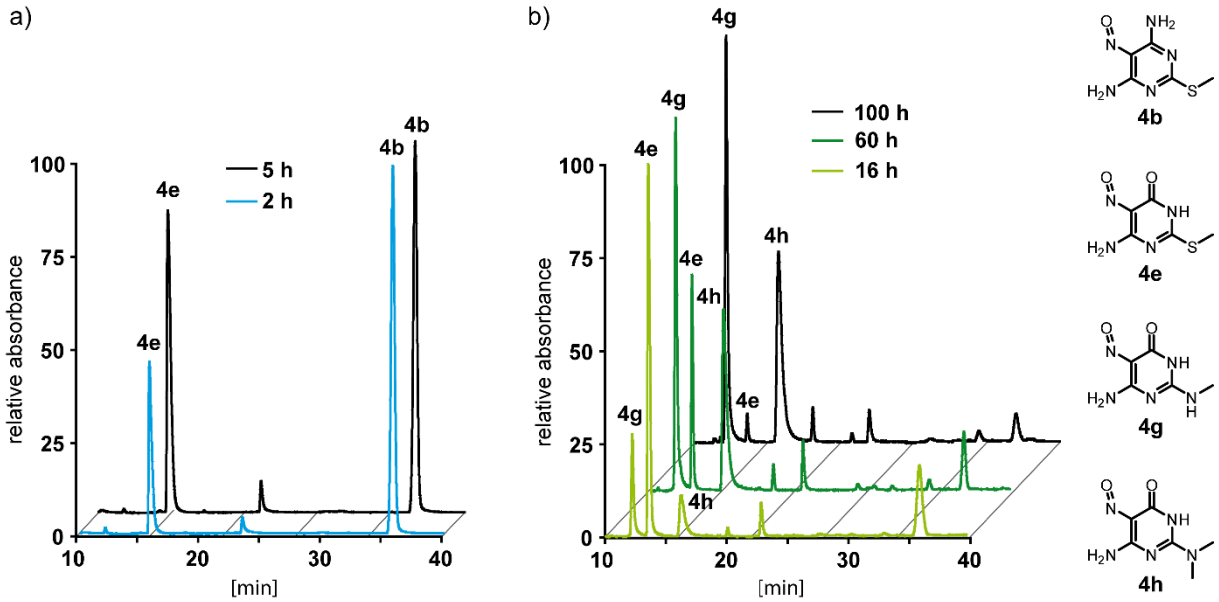
^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.23 (s, 1H, HC8), 7.38 (s, 2H, NH_2), 5.83 (d, *J* = 6.0 Hz, 1H, HC1‘), 5.45 (d, *J* = 6.2 Hz, 1H, OH), 5.21 (d, *J* = 4.8 Hz, 1H, OH), 5.03 (t, *J* = 5.6 Hz, 1H, OH), 4.62 (q, *J* = 5.8 Hz, 1H, HC2‘), 4.15 (td, *J* = 4.9, 3.3 Hz, 1H, HC3‘), 3.91 (q, *J* = 4.1 Hz, 1H, HC4‘), 3.70 – 3.49 (m, 2H, HC5‘), 2.47 (s, 3H, SCH_3). **^{13}C-NMR** (101 MHz, DMSO) δ = 164.62 (C2), 155.90 (C6), 150.61 (C4), 139.24 (C8), 117.32 (C5´), 87.68 (C1´), 85.89 (C4´), 73.67 (C2´), 70.96 (C3´), 62.04 (C5´), 14.13 (SCH_3). **HRMS** (ESI$^+$): calc. for $[C_{11}H_{16}N_5O_4S]^+$ 314.0915, found: 314.0918 $[M+H]^+$.



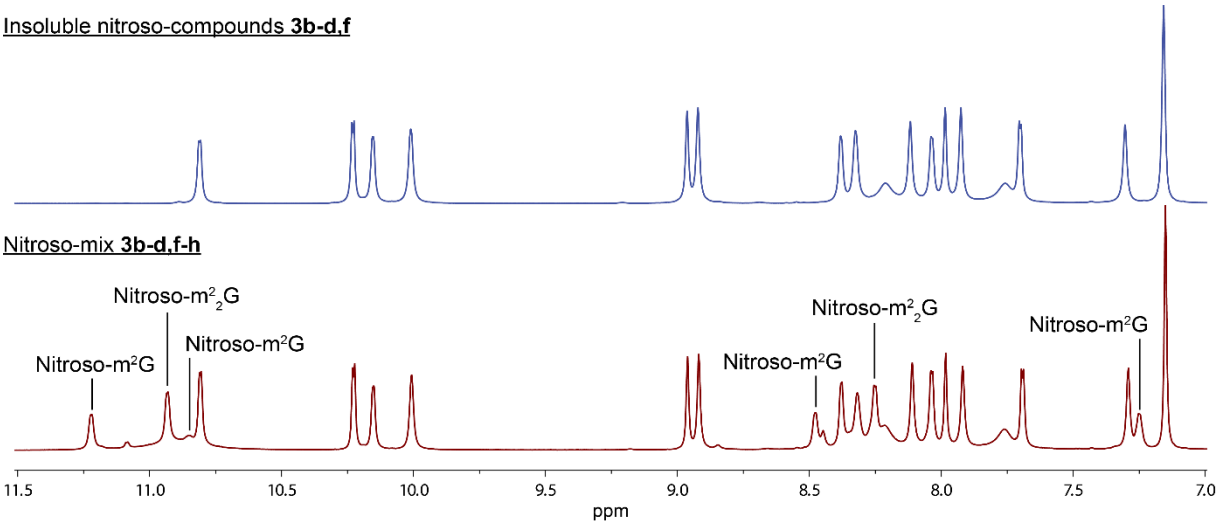
Supplementary Figure 1: Simultaneous formation of nitroso-pyrimidines **4a-d**. ^{1}H-NMR after salts containing **2a-d** and **3** were mixed and heated with a temperature gradient (1 °C/5 min, from 100-160 °C).



Supplementary Figure 2: ^{1}H-NMR comparism of prebiotic and synthetic product **4a-d**. The spectroscopic data for the prebiotic formation of nitroso-pyrimidines **4a-d** after physical enrichment shows very clean products, when compared with synthetic standards.



Supplementary Figure 3: One-pot conversion of **4b** to **4g** and **4h** by pH change. The figure shows the UV-chromatogram (at 325 nm) from LC/MS measurements of the hydrolysis of **4b** to **4e** and subsequent aminolysis to **4g** and **4h**. a) hydrolysis of **4b** to **4e** after 2 h and 5 h in 3 M HCl at room temperature containing methylamine hydrochloride (300 mM) and dimethylamine hydrochloride (100 mM). b) the pH of the reaction mixture was carefully adjusted after 5 h with Na_2CO_3 to pH ~9-10. Under basic conditions **4b** mainly precipitates whereas **4e** stays in solution. The reaction mixture is shown after 16 h, 60 h, and 100 h. **4e** is almost completely converted due to nucleophilic substitution into **4g** and **4h** after ~4 days.



Supplementary Figure 4: Separation of nitroso-pyrimidines from a mixture (**4b-d,f-h**). Even though **4i** is soluble under basic conditions, this compound was excluded due to it being insufficiently soluble for NMR measurements. Procedure: a mixture containing 12 mg each of **4b-d,f-h** was treated with 15% NH_4OH (4 ml). The insoluble compounds were filtered off and dried. NMR of the insoluble compounds shows that compound **4g** (Nitroso-m^2G) and **4h** (Nitroso-$m^2{}_2G$) are absent from the mixture.

Supplementary Table 1: crystal data of 1-methylguanidine (**2a**) salt of (hydroxyimino)-malononitrile (**3**)

CCDC number	1574226
net formula	$C_5H_8N_6O$
M_r/g mol^{-1}	168.17
crystal size/mm	0.100 × 0.030 × 0.030
T/K	100.(2)
radiation	MoKα
diffractometer	'Bruker D8 Venture TXS'
crystal system	orthorhombic
space group	'P n a 21'
a/Å	16.4092(14)
b/Å	13.0181(10)
c/Å	3.7470(3)
α/°	90
β/°	90
γ/°	90
V/Å^3	800.42(11)
Z	4
calc. density/g cm^{-3}	1.396
μ/mm^{-1}	0.106
absorption correction	Multi-Scan
transmission factor range	0.8400–0.9705
refls. measured	6118
R_{int}	0.0560
mean σ(I)/I	0.0514
θ range	3.367–26.345
observed refls.	1447
x, y (weighting scheme)	0.0441, 0.3047
hydrogen refinement	H(C) constr, H(N) refall
Flack parameter	0.5
refls in refinement	1597
parameters	130
restraints	1
$R(F_{obs})$	0.0417
$R_w(F^2)$	0.1057
S	1.083
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.187
min electron density/e Å^{-3}	−0.240

Supplementary Table 2: crystal data of methylthioamidine (**2b**) salt of (hydroxyimino)-malononitrile (**3**)

CCDC number	1574223
net formula	$C_5H_7N_5OS$
M_r/g mol^{-1}	185.22
crystal size/mm	0.100 × 0.090 × 0.050
T/K	100.(2)
radiation	MoKα
diffractometer	'Bruker D8 Venture TXS'
crystal system	triclinic
space group	'P -1'
a/Å	7.5580(4)
b/Å	9.9840(5)
c/Å	12.0318(7)
α/°	73.888(2)
β/°	86.151(2)
γ/°	80.423(2)
V/Å^3	859.88(8)
Z	4
calc. density/g cm^{-3}	1.431
μ/mm^{-1}	0.337
absorption correction	Multi-Scan
transmission factor range	0.8536–0.9705
refls. measured	8943
R_{int}	0.0247
mean σ(I)/I	0.0344
θ range	3.207–27.485
observed refls.	3268
x, y (weighting scheme)	0.0269, 0.3770
hydrogen refinement	H(C) constr, H(N) refall
refls in refinement	3886
parameters	251
restraints	1
$R(F_{obs})$	0.0311
$R_w(F^2)$	0.0787
S	1.070
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.265
min electron density/e Å^{-3}	−0.255

Supplementary Table 3: crystal data of guanidine (**2c**) salt of (hydroxyimino)malononitrile (**3**)

CCDC number	1574225
net formula	$C_4H_6N_6O$
M_r/g mol^{-1}	154.15
crystal size/mm	0.100 × 0.060 × 0.040
T/K	100.(2)
radiation	MoKα
diffractometer	'Bruker D8 Venture TXS'
crystal system	orthorhombic
space group	'F d d 2'
a/Å	21.2099(12)
b/Å	33.2970(18)
c/Å	3.8794(2)
α/°	90
β/°	90
γ/°	90
V/Å^3	2739.7(3)
Z	16
calc. density/g cm^{-3}	1.495
μ/mm^{-1}	0.117
absorption correction	Multi-Scan
transmission factor range	0.9133–0.9705
refls. measured	10489
R_{int}	0.0381
mean σ(I)/I	0.0222
θ range	3.843–26.355
observed refls.	1322
x, y (weighting scheme)	0.0310, 7.2954
hydrogen refinement	refall
Flack parameter	0.5
refls in refinement	1384
parameters	124
restraints	2
$R(F_{obs})$	0.0360
$R_w(F^2)$	0.0926
S	1.158
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.324
min electron density/e Å^{-3}	−0.189

Supplementary Table 4: crystal data of acetamidine (**2d**) salt of (hydroxyimino)-malononitrile (**3**)

CCDC number	1574224
net formula	$C_5H_7N_5O$
M_r/g mol^{-1}	153.16
crystal size/mm	0.080 × 0.050 × 0.040
T/K	100.(2)
radiation	MoKα
diffractometer	'Bruker D8 Venture TXS'
crystal system	triclinic
space group	'P -1'
a/Å	6.4976(3)
b/Å	6.9460(4)
c/Å	9.2031(5)
α/°	100.345(2)
β/°	94.349(2)
γ/°	113.040(2)
V/Å^3	371.08(3)
Z	2
calc. density/g cm^{-3}	1.371
μ/mm^{-1}	0.104
absorption correction	Multi-Scan
transmission factor range	0.6396–0.9705
refls. measured	3842
R_{int}	0.0569
mean σ(I)/I	0.0765
θ range	3.273–27.481
observed refls.	1389
x, y (weighting scheme)	0.0499, 0.1265
hydrogen refinement	H(C) constr, H(N) refall
refls in refinement	1673
parameters	117
restraints	0
$R(F_{obs})$	0.0520
$R_w(F^2)$	0.1361
S	1.079
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.373
min electron density/e Å^{-3}	−0.289

Supplementary References

1 Brown, D. J. & Jacobsen, N. W. 612. Pyrimidine reactions. Part IV. The methylation of 2,4- and 4,5-diaminopyrimidine and related compounds. *J. Chem. Soc.*, 3172-3179 (1962).

2 Chang, L. *et al.* Imidazopyridine- and Purine-Thioacetamide Derivatives: Potent Inhibitors of Nucleotide Pyrophosphatase/Phosphodiesterase 1 (NPP1). *J. Med. Chem.* **57**, 10080-10100 (2014).

3 Robins, M. J. & Uznański, B. Nucleic acid related compounds. 33. Conversions of adenosine and guanosine to 2, 6-dichloro, 2-amino-6-chloro, and derived purine nucleosides. *Can. J. Chem.* **59**, 2601-2607 (1981).

4 Reiter, V. *Synthese von eukaryotischen RNA-Modifikationenund Quantifizierung nicht kanonischer Nukleoside sowie Untersuchungen zu deren Biosynthese*, Thesis, LMU München, (2013).

5 Matsuda, A. *et al.* Nucleosides and nucleotides. 103. 2-Alkynyladenosines: a novel class of selective adenosine A2 receptor agonists with potent antihypertensive effects. *J. Med. Chem.* **35**, 241-252 (1992).

6 Yamazaki, A. *et al.* Synthesis of Thioinosine and Thio-AICA-riboside Analogs. *Chem. Pharm. Bull.* **21**, 692-696 (1973).

7 Davoll, J. & Lowy, B. A. Some Synthetic Analogs of the Natural Purine Nucleosides1. *J. Am. Chem. Soc.* **74**, 1563-1566 (1952).

8 Yamazaki, A., Kumashiro, I. & Takenishi, T. Synthesis of 2-methyladenosine and its 5'-phosphate. *J. Org. Chem.* **33**, 2583-2586 (1968).

9 Vorbrüggen, H., Krolikiewicz, K. & Bennua, B. Nucleoside syntheses, XXII1) Nucleoside synthesis with trimethylsilyl triflate and perchlorate as catalysts. *Chem. Ber.* **114**, 1234-1255 (1981).

10 Bridson, P. K. & Reese, C. B. A novel method for the methylation of heterocyclic amino groups. Conversion of guanosine into its 2-N-methyl-and 2-N,2-N-dimethyl derivatives. *Bioorg. Chem.* **8**, 339-349 (1979).

11 Ubiali, D. *et al.* Production, characterization and synthetic application of a purine nucleoside phosphorylase from Aeromonas hydrophila. *Adv. Synth. Catal.* **354**, 96-104 (2012).

12 Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res.* **31**, 4472-4480 (2003).

5.3 Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth — Generation by Prebiotic Methylations and Carbamoylations

Christina Schneider, Sidney Becker, Dr. Hidenori Okamura, Antony Crisp, Dr. Tynchtyk Amatov, Michael Stadlmeier, Prof. Dr. Thomas Carell

Prolog

Die RNA Welt Hypothese besagt, dass das Leben auf der Erde aus RNA entstand. RNA ist nicht nur in der Lage Informationen zu speichern, sondern besitzt auch katalytische Aktivität. Heutige RNA besteht jedoch nicht nur aus den vier Nukleotiden A, G, C, U, sondern auch aus mehr als 120 modifizierten RNA-Bausteinen. Modifikationen sind von enormer biologischer Bedeutung, da sie eine Reihe von fundamentalen Prozessen steuern. Es ist weiterhin fraglich, ab wann modifizierte RNA-Bausteine während der Entstehung des Lebens eine Rolle spielten. Bisher ist nicht bekannt, ob die heutigen modifizierten Nukleoside auf einer frühen Erde überhaupt verfügbar waren. In dieser Arbeit sollte deshalb evaluiert werden, inwiefern RNA-Modifikationen auf der frühen Erde aus den kanonischen Nukleosiden entstehen konnten. Präbiotische Modifikationen waren vermutlich sehr simpel und entstanden vor allem durch einfache Methylierungen. Es konnte gezeigt werden, dass eine Vielzahl an nicht-kanonischen Nukleosiden durch präbiotisch plausible Methylierung entstehen konnten. Ebenfalls wurden auch Aminosäure-modifizierte Nukleoside durch Carbamoylierung nachgewiesen. Es scheint also, dass modifizierte RNA-Bausteine durchaus die Chance hatten Teil der chemischen Evolution zu sein, aus der zuerst die RNA-Welt und später dann komplexes Leben entstand. Damit könnten RNA-Modifikationen in der Tat lebende chemische Fossilien einer frühen Erde sein.

Autorenbeitrag

Es wurden präbiotische Methylierungen an Adenosin etabliert sowie die Synthese für die *N*-Nitroso-*N*-Aminosäure Harnstoff Verbindungen entwickelt. Außerdem wurden LC-MS Analysen durchgeführt.

Lizenz

Prebiotic Chemistry

International Edition: DOI: 10.1002/anie.201801919
German Edition: DOI: 10.1002/ange.201801919

Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth—Generation by Prebiotic Methylations and Carbamoylations

*Christina Schneider, Sidney Becker, Hidenori Okamura, Antony Crisp, Tynchtyk Amatov, Michael Stadlmeier, and Thomas Carell**

Abstract: *The RNA-world hypothesis assumes that life on Earth started with small RNA molecules that catalyzed their own formation. Vital to this hypothesis is the need for prebiotic routes towards RNA. Contemporary RNA, however, is not only constructed from the four canonical nucleobases (A, C, G, and U), it also contains many chemically modified (non-canonical) bases. A still open question is whether these noncanonical bases were formed in parallel to the canonical bases (chemical origin) or later, when life demanded higher functional diversity (biological origin). Here we show that isocyanates in combination with sodium nitrite establish methylating and carbamoylating reactivity compatible with early Earth conditions. These reactions lead to the formation of methylated and amino acid modified nucleosides that are still extant. Our data provide a plausible scenario for the chemical origin of certain noncanonical bases, which suggests that they are fossils of an early Earth.*

More than 120 modified bases have been identified in RNA that are important for correct folding into complex three-dimensional structures and for fine-tuning RNA/RNA and RNA/protein interactions.[1–3] Modified nucleosides are, for example, found in proximity to the anticodon stem loop in tRNA, where they are involved in translation of the genetic code.[4,5] Methylated nucleosides such as m^6A are involved in regulating mRNA stability,[6] splicing,[7,8] translation,[9–11] and X-chromosome inactivation.[12] Another methylated nucleoside, m^7G, is part of the 5′-cap structure of eukaryotic mRNA.[13]

The RNA-world hypothesis postulates that life started with self-replicating RNA molecules that were amenable to the processing of chemical evolution through replication, randomization, and selection.[14] Since RNA is able to store genetic information and perform catalytic processes, the hypothesis further posits that an early replicating cell could proliferate and maintain a primitive metabolism in the absence of coded proteins. Noncoded polypeptides[15,16] and simple anabolic pathways[17,18] may have supported an early RNA-based metabolism.

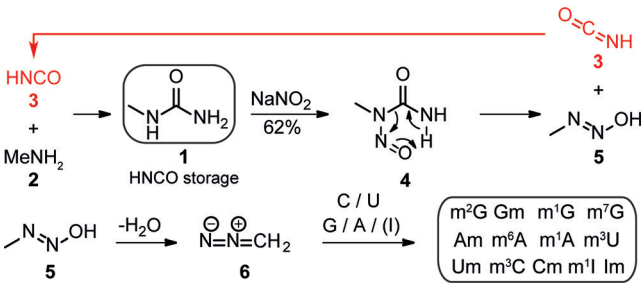
[*] C. Schneider, S. Becker, Dr. H. Okamura, A. Crisp, Dr. T. Amatov, M. Stadlmeier, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, LMU München
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de
Supporting information and the ORCID identification numbers for the authors of this article can be found under: https://doi.org/10.1002/anie.201801919.

This hypothesis requires the presence of the key building blocks of life, such as nucleosides and amino acids, or of primitive anabolic processes that led to their formation.[19] This raises the question of whether life began with only the four canonical bases (A, C, G, and U)[20,21] or if an early pre-RNA was chemically more diverse[22] and contained non-canonical nucleosides.[22] Those noncanonical bases that are still found in RNA might be considered fossils of this early phase of chemical evolution.[23–25] Finding evidence for this idea requires simple chemical synthetic routes compatible with geochemical models of early Earth that generate these noncanonical bases. Here we show that the majority of methylated nucleosides, which play important roles in RNAs of all three domains of life, can be prebiotically generated by the reaction of canonical nucleosides with nitrosylated *N*-methylurea (**1**; Scheme 1).[26]

NO_2^- was potentially available on the early Earth from NO and NO_2,[27] which are formed during lightning in an N_2 atmosphere.[28] Alternatively, NO can form by the reaction of N_2 with CO_2 in hot impact plumes.[29]

Besides the methylated RNA bases, we also find nucleosides modified with amino acids among the many contemporary noncanonical RNA bases.[30,31] They are directly involved in decoding the genetic information.[32,33] We show that our NO_2^--based reactions also provide these modified bases, which suggests an early intimate contact between nucleobases and amino acids that might have formed the basis for the co-evolution of RNA and proteins and the establishment of primitive protometabolic pathways.

The synthetic route starts with methylurea (**1**), which is one of the molecules that was likely present on the early Earth.[34] Methylurea (**1**) is, for example, formed by the reaction of ammonia with methyl isocyanate, which was detected on comet 67P/Churyumov-Gerasimenko.[35,36] Methylurea (**1**) was also shown to form directly in the



Scheme 1. Reactions that lead to the formation of methylated derivatives of canonical nucleobases that are today found in RNA in all three domains of life. Methylurea functions as a storage molecule for reactive isocyanic acid.

Urey–Miller experiment[37] and it is available in high yields from the reaction of methylamine (**2**) with HNCO (**3**) in water (90%).[38] HNCO in turn was detected in interstellar gases[39] and, likewise, on comet 67P/Churyumov-Gerasimenko.[35,40] Urea itself is also known to decompose into ammonium isocyanate.[41,42] Despite the potential formation of **3**, however, it is difficult to conceive the accumulation of HNCO (**3**) because of its high reactivity. If, however, a small amount of methylurea (**1**) is present, it can readily react with NO^+ (Scheme 1). Methylurea (**3**) is easily nitrosylated, which gives *N*-methyl-*N*-nitrosourea (**4**)[26] in a yield of 62%. This compound physically separates as a foam from the aqueous phase, which could potentially allow **4** to accumulate, so that it may have been locally available at high concentrations. Under slightly basic conditions, for example in the presence of borax (reported to be important for ribose-forming reactions),[43] **4** quickly decomposes to furnish 1-hydroxy-2-methyldiazene (**5**) under liberation of HNCO (**3**). As such, only small amounts of HNCO are required to help convert $MeNH_2$ and $NaNO_2$ into 1-hydroxy-2-methyldiazene (**5**). 1-Hydroxy-2-methyldiazene (**5**) in turn eliminates water and decomposes to diazomethane (**6**), which is a common methylating agent.[44] Since all starting materials are likely components of the organic matter on the early Earth, it is, therefore, plausible that diazomethane was an accessible component. The controlled release of **6** from the stable precursor methylurea (**1**) could have made it available for chemical transformations, despite its high reactivity and consequently short half-life time on the early Earth.

When we performed this base-catalyzed formation of diazomethane (**6**) in the presence of the canonical nucleobases, we obtained a large set of methylated compounds (Figure 1). For the experiment we dissolved the nucleosides in a 1:1 mixture of borate buffer and formamide. Formamide is accessible under early Earth conditions through the reaction of HCN with H_2O.[45] *N*-Methyl-*N*-nitrosourea (**4**) was then added to the nucleoside mixture in one portion. After one hour at 70°C, samples were taken and analyzed by LC-MS and tandem mass spectrometry (Figure 1). To correctly assign the resulting methylated nucleosides, co-injections with synthetic reference compounds were performed (see the Supporting Information). The products were further elucidated by analysis of the fragmentation patterns in $LC\text{-}MS^2$ experiments. When we performed the reaction in the presence of adenosine, we obtained m1A, Am, and m6A, together with the 3′- and 5′-methylated derivatives (marked as mxA; see the Supporting Information). When guanosine was methylated under the same conditions, we detected m7G (7%) as well as Gm, m1G, and m2G, all of which are known noncanonical bases. In the presence of cytidine, the bases m3C and Cm were generated. Furthermore, the reaction of uridine furnished the methylated compounds Um and m3U. m3U was formed in a high yield of 11%. We also investigated the methylation reaction with inosine (I), the hydrolysis product of A.[46,47] When inosine was subjected to the same conditions, we detected the formation of Im and m1I (see the Supporting Information). Importantly, nearly all the methylated nucleosides that we observed are today found in RNAs of all three domains of life.[2,48]

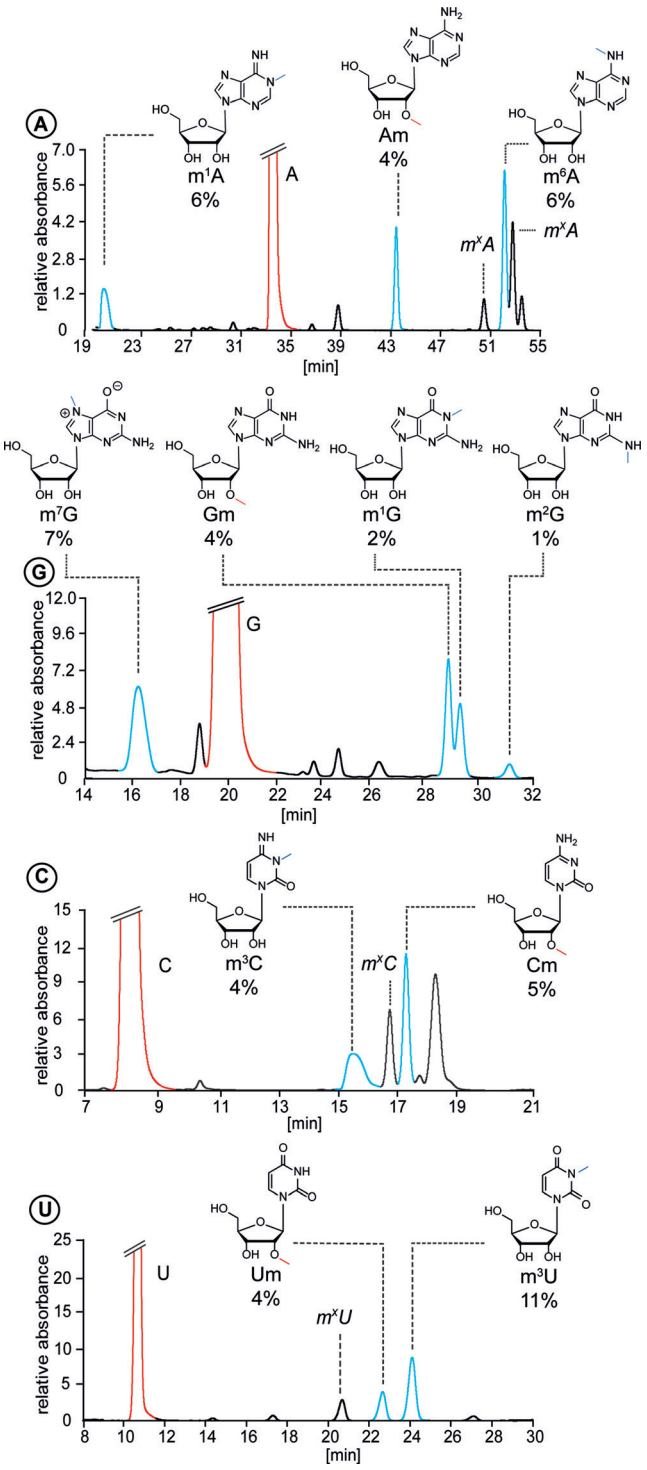


Figure 1. HPLC traces of the reaction mixtures obtained in the reaction of *N*-methyl-*N*-nitrosourea (**4**) in the presence of the canonical nucleobases A, G, C, and U. The modified nucleosides are shown in blue, the canonical ones in red. Peaks labeled with "mˣ" were identified as sugar-modified nucleosides based on data from fragmentation studies (see the Supporting Information).

We next asked the question of whether the simple reactions could be used to enable the attachment of larger chemical moieties, such as amino acids, to the canonical nucleobases to give RNA modifications such as t^6A and g^6A.

This was indeed possible (Figure 2) when we replaced HNCO by methyl isocyanate (CH_3NCO, **7**). CH_3NCO (**7**) can be generated under prebiotic conditions by UV irradiation of CH_4 and HNCO. In an aqueous environment, we observed that **7** reacts rapidly with amino acids such as glycine (**8a**) and threonine (**8b**) to give the corresponding methylurea derivatives **9** (Figure 2) in nearly quantitative yields. Compounds **9a** and **b** can be nitrosylated[49] under the same conditions as **1** to form the nitroso compounds **10a** and **10b** in high yields of 82–95%. A pH change to slightly basic conditions through the use of either phosphate or borate buffer converts the intermediate nitroso compounds **10a**,**b** into the isocyanates of the corresponding amino acids **11a**,**b**. Upon treatment with adenosine, these intermediates react to give the corresponding N^6-derivatives g^6A and t^6A. Since the reaction takes place under basic conditions, not only N^6 but also the 2′-, 3′-, and 5′-hydroxy groups can react with the isocyanate derivative of the amino acids (Figure 2). Interestingly, the selectivity of the reaction can be controlled to favor the N^6 position by the addition of a Ni^{2+} salt, which is generated during prebiotic formation of nucleoside.[22] At the same time, CH_2N_2 (**6**) is formed, which can facilitate subsequent methylations (see the Supporting Information). Interestingly, the amino acid modified nucleosides that are formed as described here, are present today in all three domains of life.[2,48]

Recently, a comparative phylogenetic analysis[50] has suggested that noncanonical bases were likely already present in the ancient parent of all life on Earth, known conventionally as LUCA (the last universal common ancestor). An overlay of the nucleosides accessed in this study with those derived from the genetic analysis shows surprising consensus (Figure 3). Most of the simple modifications that were present in LUCA could also be formed by the reactions presented here.

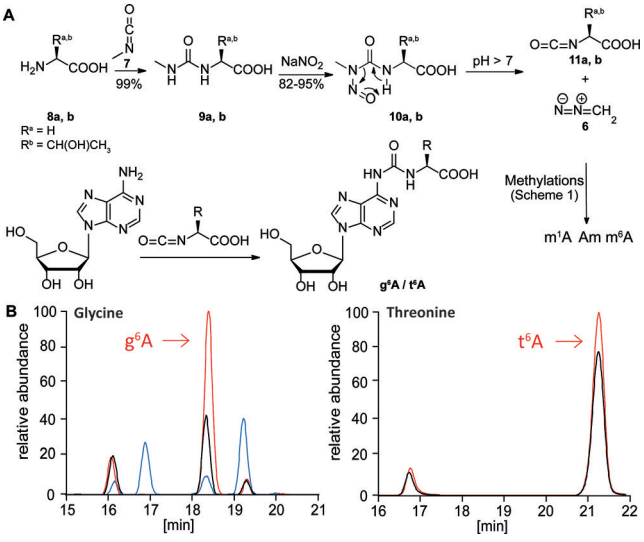


Figure 2. A) Plausible reaction scheme for prebiotic access to t^6A and g^6A. B) MS chromatograms of the reactions of *N*-methylurea derivatives **9a**,**b** with the canonical nucleoside A under formation of isocyanates of the corresponding amino acids. The chromatogram in blue shows the reaction without the addition of Ni^{2+} salts and the ones in black represents the reaction in the presence of Ni^{2+} salts. Co-injections with synthetic standards are shown in red. The additional peaks arise from the reaction of the sugars with the amino acid isocyanate. The selectivity can be increased by the addition of $[Ni(ClO_4)_2]$.

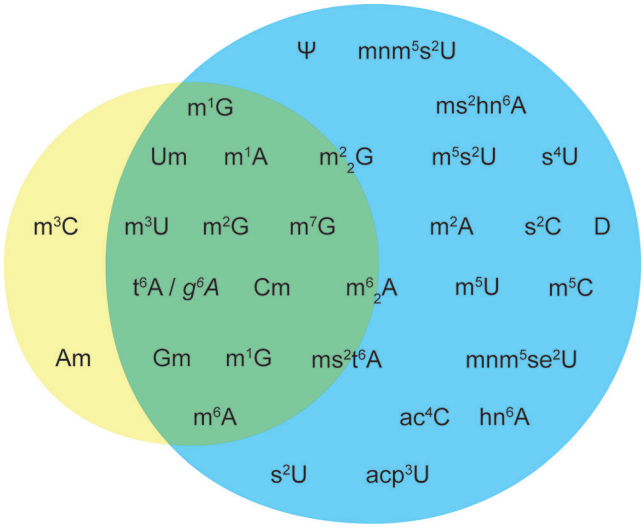


Figure 3. Noncanonical nucleosides generated in this study (yellow area) and noncanonical nucleosides that were found based on phylogenetic analysis to be likely early nucleobases bases in the biosphere (blue area).[50] Modified nucleosides that were found in both studies are shown in the green area. ms^2t^6A, $m^2{}_2G$, and $m^6{}_2A$ were placed at the border since they could also be generated using the here-described reactions starting from ms^2A, m^2G, and m^6A.

In summary, we report a simple cascade reaction that starts from isocyanic acid, methylisocyanate, methylamine, ammonia, and sodium nitrite. In this cascade the unstable molecule isocyanic acid (**3**) is captured by methylamine and stored in the form of methylurea. It can be released under basic conditions from *N*-methyl-*N*-nitrosourea (**4**), which is produced by nitrosylation of methylurea (**1**). These reactions allow us to convert the canonical pyrimidine and purine bases, for which prebiotically plausible formation processes were recently described,[20,21,51,52] into noncanonical nucleosides. As such, our results provide chemical evidence that the canonical and many noncanonical ribonucleosides can form spontaneously under plausible prebiotic conditions. The here described reactions can be linked to the nitrosylation reactions that were recently reported to enable the parallel formation of canonical and noncanonical bases.[22] The noncanonical bases, particularly the amino acid modified purines, potentially increase the chemical diversity of RNA to broaden its folding and catalytic capabilities. This complements ideas that non-canonical base pairs might have existed in pre-RNA.[22,53]

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support through SFB1032 (TP-A5), SPP1784, CA275/11-1, and the Excellence Cluster CiPS^M. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation

programme (grant agreement no. EPiR 741912). We thank Dr. Markus Müller for comments on this manuscript.

Conflict of interest

The authors declare no conflict of interest.

Keywords: methylation · nucleosides · nucleoside modification · origin of life · prebiotic chemistry

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 5943 – 5946
Angew. Chem. **2018**, *130*, 6050 – 6054

[1] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110 – 7131; *Angew. Chem.* **2012**, *124*, 7220 – 7242.
[2] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, *Nucleic Acids Res.* **2012**, *41*, D262 – D267.
[3] A. Czerwoniec, S. Dunin-Horkawicz, E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118 – D121.
[4] P. C. Thiaville, B. El Yacoubi, C. Köhrer, J. J. Thiaville, C. Deutsch, D. Iwata-Reuyl, J. M. Bacusmo, J. Armengaud, Y. Bessho, C. Wetzel, *Mol. Microbiol.* **2015**, *98*, 1199 – 1221.
[5] J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, *33*, 7869 – 7876.
[6] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, *Nature* **2014**, *505*, 117.
[7] X. Zhao, Y. Yang, B.-F. Sun, Y. Shi, X. Yang, W. Xiao, Y.-J. Hao, X.-L. Ping, Y.-S. Chen, W.-J. Wang, *Cell Res.* **2014**, *24*, 1403.
[8] W. Xiao, S. Adhikari, U. Dahal, Y.-S. Chen, Y.-J. Hao, B.-F. Sun, H.-Y. Sun, A. Li, X.-L. Ping, W.-Y. Lai, *Mol. Cell* **2016**, *61*, 507 – 519.
[9] K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S.-B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999 – 1010.
[10] J. Zhou, J. Wan, X. Gao, X. Zhang, S. R. Jaffrey, S.-B. Qian, *Nature* **2015**, *526*, 591.
[11] X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388 – 1399.
[12] D. P. Patil, C.-K. Chen, B. F. Pickering, A. Chow, C. Jackson, M. Guttman, S. R. Jaffrey, *Nature* **2016**, *537*, 369.
[13] N. Sonenberg, M. A. Morgan, W. C. Merrick, A. J. Shatkin, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 4843 – 4847.
[14] A. C. Ferretti, G. F. Joyce, *Biochemistry* **2013**, *52*, 1227 – 1235.
[15] S. Tagami, J. Attwater, P. Holliger, *Nat. Chem.* **2017**, *9*, 325.
[16] J. G. Forsythe, S. S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernández, N. V. Hud, *Angew. Chem. Int. Ed.* **2015**, *54*, 9871 – 9875; *Angew. Chem.* **2015**, *127*, 10009 – 10013.
[17] K. B. Muchowska, S. J. Varma, E. Chevallot-Beroux, L. Lethuillier-Karl, G. Li, J. Moran, *Nat. Ecol. Evol.* **2017**, *1*, 1716.
[18] A. J. Coggins, M. W. Powner, *Nat. Chem.* **2017**, *9*, 310 – 317.
[19] E. O. Leslie, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99 – 123.
[20] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833 – 836.
[21] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239 – 242.
[22] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.
[23] M. P. Robertson, G. F. Joyce, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608.
[24] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat. Commun.* **2016**, *7*, 11328.
[25] A. C. Rios, Y. Tor, *Isr. J. Chem.* **2013**, *53*, 469 – 483.
[26] A. Tsubura, Y.-C. Lai, H. Miki, T. Sasaki, N. Uehara, T. Yuri, K. Yoshizawa, *in vivo* **2011**, *25*, 11 – 22.
[27] Y. L. Yung, M. B. Mcelroy, *Science* **1979**, *203*, 1002 – 1004.
[28] W. L. Chameides, J. C. G. Walker, *Origins life* **1981**, *11*, 291 – 302.
[29] J. F. Kasting, *Origins Life Evol. Biospheres* **1990**, *20*, 199 – 231.
[30] G. B. Chheda, *Life Sci.* **1969**, *8*, 979 – 987.
[31] C. Deutsch, B. El Yacoubi, V. de Crécy-Lagard, D. Iwata-Reuyl, *J. Biol. Chem.* **2012**, *287*, 13666 – 13673.
[32] S. E. Kolitz, J. R. Lorsch, *FEBS Lett.* **2010**, *584*, 396 – 404.
[33] F. V. Murphy IV, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186.
[34] T. G. Waddell, L. L. Eilders, B. P. Patel, M. Sims, *Origins Life Evol. Biospheres* **2000**, *30*, 539 – 548.
[35] F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Krüger, L. Le Roy, A. J. MacDermott, *Science* **2015**, *349*, aab0689.
[36] N. F. W. Ligterink, A. Coutens, V. Kofman, H. S. P. Müller, R. T. Garrod, H. Calcutt, S. F. Wampfler, J. K. Jørgensen, H. Linnartz, E. F. van Dishoeck, *Mon. Not. R. Astron. Soc.* **2017**, *469*, 2219 – 2229.
[37] S. L. Miller, H. C. Urey, *Science* **1959**, *130*, 245 – 251.
[38] B. T. Golding, C. Bleasdale, J. McGinnis, S. Müller, H. T. Rees, N. H. Rees, P. B. Farmer, W. P. Watson, *Tetrahedron* **1997**, *53*, 4063 – 4082.
[39] K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin, *Sci. Adv.* **2016**, *2*, e1600285.
[40] P. Ehrenfreund, S. B. Charnley, *Annu. Rev. Astron. Astrophys.* **2000**, *38*, 427 – 483.
[41] W. H. Shaw, J. J. Bordeaux, *J. Am. Chem. Soc.* **1955**, *77*, 4729 – 4733.
[42] B. P. Callahan, Y. Yuan, R. Wolfenden, *J. Am. Chem. Soc.* **2005**, *127*, 10828 – 10829.
[43] H.-J. Kim, A. Ricardo, H. I. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye, S. A. Benner, *J. Am. Chem. Soc.* **2011**, *133*, 9457 – 9468.
[44] K. K. Park, J. S. Wishnok, M. C. Archer, *Chem.-Biol. Interact.* **1977**, *18*, 349 – 354.
[45] R. Saladino, C. Crestini, S. Pino, G. Costanzo, E. Di Mauro, *Phys. Life Rev.* **2012**, *9*, 84 – 104.
[46] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531 – 543.
[47] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25 – 33.
[48] M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262 – D267.
[49] T. Machinami, T. Suami, *Bull. Chem. Soc. Jpn.* **1975**, *48*, 1333 – 1334.
[50] M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 16116.
[51] H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11315 – 11320.
[52] R. Saladino, E. Carota, G. Botta, M. Kapralov, G. N. Timoshenko, A. Y. Rozanov, E. Krasavin, E. Di Mauro, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2746 – E2755.
[53] C. B. Winiger, M.-J. Kim, S. Hoshika, R. W. Shaw, J. D. Moses, M. F. Matsuura, D. L. Gerloff, S. A. Benner, *Biochemistry* **2016**, *55*, 3847 – 3850.

Manuscript received: February 12, 2018
Accepted manuscript online: March 13, 2018
Version of record online: April 17, 2018

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth—Generation by Prebiotic Methylations and Carbamoylations

*Christina Schneider, Sidney Becker, Hidenori Okamura, Antony Crisp, Tynchtyk Amatov, Michael Stadlmeier, and Thomas Carell**

anie_201801919_sm_miscellaneous_information.pdf

Materials and Methods

General Information

Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR, Carbosynth or Acros Organics and used without further purification. Solutions were concentrated *in vacuo* on a Heidolph rotary evaporator. The solvents were of reagent grade or purified by distillation. Chromatographic purification of products was accomplished using flash column chromatography on Merck Geduran Si 60 (40-63 µm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on Merck 60 (silica gel F254) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or standard staining solutions. ^{1}H- and ^{13}C-NMR spectra were recorded in deuterated solvents on Varian Oxford 200, Bruker ARX 300, Varian VXR400S, Varian Inova 400, Bruker AMX 600 and Bruker AVIIIHD 400 spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. High-resolution ESI spectra were obtained on the mass spectrometers Thermo Finnigan LTQ FT-ICR. IR measurements were performed on Perkin Elmer Spectrum BX FT-IR spectrometer with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were measured on a Büchi B-540. For preparative HPLC purification a Waters 1525 binary HPLC Pump in combination with a Waters 2487 Dual Absorbance Detector was used, with Macherey-Nagel VP 250/10 Nucleosil 100-7 C18 reversed phase column. The analysis of the prebiotic reactions were analyzed by LC-ESI-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Interchim Uptisphere120-3HDO C18 column. The column temperature was maintained at 30 °C. Eluting buffers were buffer A (2 mM $HCOONH_4$ in H2O (pH 5.5)) and buffer B (2 mM $HCOONH_4$ in H_2O/MeCN 20/80 (pH 5.5)). For samples containing methylated and aminoacylated nucleosides the following gradient was used: 0 → 55 min, 0% → 8% HPLC/MS Puffer B; 55 → 70 min, 8% → 60% HPLC/MS Puffer B; 70 → 72 min, 60% → 100% HPLC/MS Puffer B; 72 → 78 min, 100% HPLC/MS Puffer B; 78 → 81 min, 100 → 0% HPLC/MS Puffer B; 81 → 90 min, 0% HPLC/MS Puffer B. The elution was monitored at 260 nm (Dionex Ultimate 3000 Diode Array Detector). The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 120-1000 with a resolution of 30000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5 µM). The parameters used in this section were sheath gas flow rate, 16 arb; auxiliary gas flow rate, 11 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200 °C; capillary voltage, 20 V, tube lens 65 V.

Synthethis of the methylated nucleosides under prebiotic conditions:

Methylated Adenosine

Adenosine (6.00 mg, 22.5 µmol, 1 eq.) was dissolved in a suspension of borax (105 mg, 275 µmol, 12 eq) in H_2O (250 µL) and formamide (100 µL).1-methyl-1-nitrosourea (80.0 mg, 559 µmol, 25 eq.) was dissolved in formamide (150 µL). The suspension was stirred for 18 h at 50°C before 25 µL of the mixture were diluted into 975 µL H_2O and analyzed with LC-MS. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards (see Fig. S4). Yields of the prebiotic reactions were determined with calibration curves (see Fig.S6 and S7).

Methylated Guanosine

Guanosine (6.40 mg, 22.5 µmol, 1 eq.) was dissolved in a suspension of borax (105 mg, 275 µmol, 12 eq) in H_2O (250 µL) and formamide (100 µL).1-methyl-1-nitrosourea (80.0 mg, 559 µmol, 25 eq.) was dissolved in formamide (150 µL). The suspension was stirred for 18 h at 50°C before 25 µL of the mixture were diluted into 975 µL H_2O and analyzed with LC-MS. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards (see Fig. S4). Yields of the prebiotic reactions were determined with calibration curves (see Fig.S6 and S7).

Methylated Cytidine

Cytidine (5.50 mg, 22.5 µmol, 1 eq.) was dissolved in a suspension of borax (105 mg, 275 µmol, 12 eq) in H_2O (250 µL) and formamide (100 µL).1-methyl-1-nitrosourea (80.0 mg, 559 µmol, 25 eq.) was dissolved in formamide (150 µL). The suspension was stirred for 18 h at 50°C before 25 µL of the mixture were diluted into 975 µL H_2O and analyzed with LC-MS. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards (see Fig. S4). Yields of the prebiotic reactions were determined with calibration curves (see Fig.S6 and S7).

Methylated Uridine

Uridine (5.50 mg, 22.5 µmol, 1 eq.) was dissolved in a suspension of borax (105 mg, 275 µmol, 12 eq) in H_2O (250 µL) and formamide (100 µL).1-methyl-1-nitrosourea (80.0 mg, 559 µmol, 25 eq.) was dissolved in formamide (150 µL). The suspension was stirred for 18 h at 50°C before 25 µL of the mixture were diluted into 975 µL H_2O and analyzed with LC-MS. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards (see Fig. S4). Yields of the prebiotic reactions were determined with calibration curves (see Fig.S6 and S7).

Methylated Inosine

Inosine (6.00 mg, 22.5 µmol, 1 eq.) was dissolved in a suspension of borax (105 mg, 275 µmol, 12 eq) in H_2O (250 µL) and formamide (100 µL).1-methyl-1-nitrosourea (80.0 mg, 559 µmol, 25 eq.) was dissolved in formamide (150 µL). The suspension was stirred for 18 h at 50°C before 25 µL of the mixture were diluted into 975 µL H_2O and analyzed with LC-MS. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards (see Fig. S4). Yields of the prebiotic reactions were determined with calibration curves (see Fig.S6 and S7).

Synthesis of the amino acid modified nucleosides under prebiotic conditions:

Without Nickel:

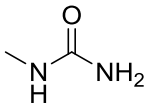
Adenosine (2.67 mg, 10.0 µmol, 1 eq.) was dissolved in phosphate buffer (370 µL, 30 mM, pH 8) and H_2O. (40 µL). The (methyl(nitroso)carbamoyl)amino acid (Glycine: 3.32 mg, 20.9 µmol, 2 eq; Threonine: 4.10 mg, 20.9 µmol, 2 eq) was suspended in water (40 µL) and was slowly added to the reaction mixture. After 17 h at 70°C 50 µL samples were removed and diluted to 1 mL. The samples were used for LC-MS analysis. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards

With Nickle:

Adenosine (2.67 mg, 10.0 µmol, 1 eq.) was dissolved in phosphate buffer (370 µL, 30 mM, pH 8). $[Ni(ClO_4)_2]$ x 6 H_2O (91.4 mg, 250 µmol, 25 eq) was added in H_2O (40 µL). The (methyl(nitroso)carbamoyl)amino acid (Glycine: 3.32 mg, 20.9 µmol, 2 eq; Threonine: 4.10 mg, 20.9 µmol, 2 eq) was suspended in water (40 µL) and was slowly added to the reaction mixture. After 17 h at 70°C 50 µl samples were removed and diluted to 1 mL. The samples were used for LC-MS analysis. The modifications were confirmed via coinjection of the prebiotic samples with synthetic standards

Prebiotic access to Urea und Urea-Nitroso-Compounds

1-methylurea (1)

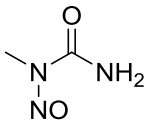


An aqueous solution of ammonia (100 µL, 25%) was added to H_2O (150 µL) and cooled to 0°C. Methylisocyanate (20.0 µL, 326 µmol, 1.0 eq) was added and the mixture was stirred at

20°C. After 3.5 h the solvent was evaporated and the product could be obtained as a white solid (23.92 mg, 322 µmol, 99%).

^{1}H-NMR (600 MHz, CD_3CN) δ = 5.04 (bs, 1H, NH), 4.71 (s, 2H, NH_2), 2.60 (d, 3H, NCH_3).
^{13}C-NMR (600 MHz, CD_3CN) δ = 160.32 (1C, CO), 26.94 (s, 1C, NCH_3).
HRMS (EI^+): calc. for $[C_2H_5N_3O_2]$ 103.0383, found: 103.0374 [M].
IR (cm^{-1}): ṽ = 3412 (w), 3320 (m), 3203 (w), 1650 (m), 1625 (m), 1565 (s), 1420 (m), 1348 (m), 1167 (m), 1024 (s), 1004 (m), 657 (m).

1-methyl-1-nitrosourea (4)

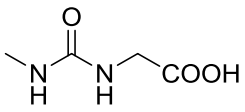


The reaction was carried out slightly modified to Arndt *et al.*[1].

Methylurea (1.00 g, 13.5 mmol, 1 eq) was dissolved in H_2O (8 mL). $NaNO_2$ (1.02 g, 14.9 mmol, 1.1 eq) was added at 0°C and aq. HCl (1.9 mL, 18.5 mmol, 1.4 eq) was added dropwise to the reaction mixture. The precipitated solid was filtered off, washed with H_2O and dried *in vacuo*. The product was obtained as a white solid (0.86 g, 8.31 mmol, 62%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.15 (s, 1H, NH), 7.82 (s, 1H, NH), 3.06 (s, 3H, NCH_3).
^{13}C-NMR (400 MHz, DMSO-d_6) δ = 150.08 (1C, CO), 48.64 (s, 1C, NCH_3).
HRMS (EI^+): calc. for $[C_2H_5N_3O_2]$ 103.0383, found: 103.0374 [M].
mp: 108°C
IR (cm-1): ṽ = 3377 (w), 3245 (w), 1722 (m), 1604 (m), 1457 (m), 1417 (s), 1369 (m), 1206 (m), 1086 (m), 974 (s), 841 (m), 775 (m), 641 (s).

(Methylcarbamoyl)glycine (9a)



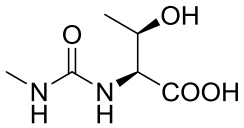
The reaction was carried out slightly modified to Machinami *et al.* [2].

Glycine (150 mg, 2.00 mmol) was dissolved in 10% NaOH (1mL) and methylisocyanate (150 µL, 2.42 mmol, 1.2 eq; prepared by thermal cleavage of 3-methyl-1,1-diphenylurea[3])

was added to the solution at 0°C. The reaction mixture was stirred for 3 hours at room temperature before the pH was adjusted to 2-3 with Amberlite IR-120 (H^+-type). The reaction mixture was freeze-dried to obtain a white powder (261 mg, 1.98 mmol, 99%). The crude product could be used without further purification.

^{1}H-NMR (400 MHz, DMSO-d_6) δ 6.10 (t, *J* = 5.8 Hz, 1H, $NHCH_2$), 6.02 (q, *J* = 4.8 Hz, 1H, $NHCH_3$), 5.75 (s, 1H), 3.65 (d, *J* = 5.5 Hz, 2H, CH_2), 2.54 (d, *J* = 4.6 Hz, 3H, NCH_3).
^{13}C NMR (101 MHz, DMSO) δ 173.10 (1C, COOH), 158.99 (1C, CON_2), 42.14 (1C, CH_2), 26.77 (1C, NCH_3).
HRMS (ESI^-): calc. for $[C_4H_7N_2O_3]^-$: 131.0462, found: 131.0462 $[M-H]^-$.
IR (cm^{-1}): ṽ = 3354 (w), 2976 (w), 2936 (w), 2359 (br., w), 1708 (m), 1623 (s), 1552 (s), 1413 (m), 1223 (m), 1185 (s), 1185 (m), 1122 (s) 1077 (w), 1014 (w) 949 (w), 904 (w), 862 (w), 761 (w).

3-hydroxy-2-(3-methylureido)butanoic acid (9b)

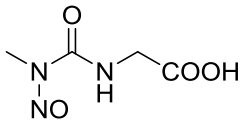


The reaction was carried out slightly modified to Machinami *et al.* [2].

Threonine (238 mg, 2.00 mmol) was dissolved in 10% NaOH (1 mL) and methylisocyanate (150 µl, 2.42 mmol, 1.2 eq; prepared by thermal cleavage of 3-methyl-1,1-diphenylurea[3]) was added to the solution at 0°C. The reaction mixture was stirred for 3 hours at room temperature before the pH was adjusted to 2-3 with Amberlite IR-120 (H^+-type). The reaction mixture was freeze-dried to obtain a white powder (348 mg, 1.98 mmol, 99%). The crude product could be used without further purification.

^{1}H-NMR (400 MHz, D_2O) δ = 4.32 (m, 1H, CHOH), 4.23 (m, 1H, CHNHCOOH), 2.69 (s, 3H, NCH_3), 1.19 (d, *J*=6.4 Hz, 3H, $CHCH_3$).
^{13}C-NMR (400 MHz, DMSO-d_6) δ = 173.58 (s, 1C, COOH), 158.66 (s, 1C, N_2CO), 65.31 (s, 1C, CHCOOH), 56.67 (s, 1C, CHOH), 24.01 (s, 1C, NCH_3), 16.58 (s, 1C, $CHCH_3$).
HRMS (ESI^-): calc. for $[C_6H_{11}N_2O_4]^-$ 175.0724, found: 175.0723 $[M-H]^-$.
IR (cm^{-1}): ṽ = 3368 (w), 2976 (w), 2938 (w), 1701 (s), 1528 (m), 1483 (m), 1405 (m), 1224 (m), 1163 (s), 1127 (m), 1079 (m) 981 (s), 923 (m), 841 (w), 781 (s), 663 (m).

(Methyl(nitroso)carbamoyl)glycine (10a)

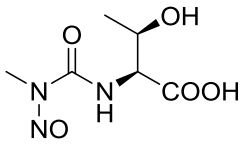


The reaction was carried out slightly modified to Machinami *et al.* [2].

(Methylcarbamoyl)glycine (754 mg, 5.71 mmol, 1 eq) was dissolved in water (10 mL) and acetic acid (1.0 mL). $NaNO_2$ (393 mg, 5.71 mmol, 1 eq) was added at 0°C. The reaction mixture was stirred for 4 h before the pH was adjusted to 3 using Dowex 50WX8-100. The reaction mixture was freeze-dried to obtain a light-yellow solid (735 mg, 4.68 mmol, 82%). The crude product could be used without further purification.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 9.00 (t, *J* = 6.0 Hz, 1H, NH), 3.94 (d, *J* = 6.0 Hz, 2H, CH_2), 3.11 (s, 2H, NCH_3).
^{13}C-NMR (101 MHz, DMSO-d_6) δ = 170.85 (s, 1C, COOH), 153.47 (s, 1C, $CONHCH_3$), 42.01 (s, 1C, CH_2), 27.04 (s, 1C, NCH_3).
HRMS (ESI^-): calc. for $[C_4H_6N_3O_4]^-$ 160.0364, found: 160.0362 $[M-H]^-$.
mp: 115°C
IR (cm^{-1}): ṽ = 3310 (w), 2943 (w), 2578 (w), 1723 (m), 1689 (s), 1537 (m), 1489 (s), 1434 (m), 1404 (m), 1348 (w), 1258 (m), 1174 (s) 1109 (w), 1067 (m), 1025 (m), 922 (s), 988 (m), 849 (m) 786 (m), 766 (m), 703 (m).

3-hydroxy-2-(3-methyl-3nitroureido)butanoic acid (10b)



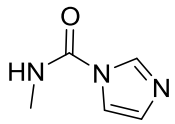
The reaction was carried out slightly modified to Machinami *et al.* [2].

3-hydroxy-2-(3-methylureido)butanoic acid (99.0 mg, 0.56 mmol, 1 eq) was dissolved in water (5 mL) and acetic acid (0.5 mL). $NaNO_2$ (58.2 mg, 0.84 mmol, 1.5 eq) was added at 0°C. The reaction mixture was stirred for 4 h before the pH was adjusted to 3 using Dowex 50WX8-100. The reaction mixture was freeze-dried to obtain a yellow oil (110 mg, 0.54 mmol, 95%). The crude product could be used without further purification.

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 8.05 (d, *J*=1.20, 1H, NH), 4.36 (dd, *J* = 8.7, 3.1 Hz, 1H, CHOH), 4.25 (qd, *J* = 6.4, 3.1 Hz, 1H CHNHCOOH), 3.31 (s, 3H, NCH_3), 1.18 (d, *J*=6.4, 3H, $CHCH_3$).
^{13}C-NMR (400 MHz, DMSO-d_6) δ = 172.08 (s, 1C, COOH), 153.72 (s, 1C, N_2CO), 66.69 (s, 1C, CHCOOH), 59.90 (s, 1C, CHOH), 27.51 (s, 1C, NCH_3), 21.04 (s, 1C, $CHCH_3$).
HRMS (ESI^-): calc. for $[C_6H_{10}N_3O_5]^-$ 204.0626, found: 204.0624 $[M-H]^-$.
IR (cm^{-1}): ῡ = 3394 (w), 2979 (w), 2938 (w), 1701 (m), 1528 (s), 1483 (m), 1405 (s), 1350 (m), 1291 (m), 1224 (w), 1163 (m), 1127 (s) 1079 (w), 981 (m), 923 (m), 874 (s), 841 (m), 781 (m) 753 (m), 663.

Synthetic access to amino acid-urea-compounds

N-methyl-1H-imidazole-1-carboxamide (12)

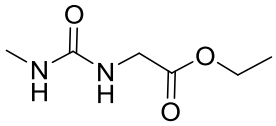


The reaction was carried out slightly modified to Duspara *et al.*[4].

CDI (15.0 g, 92.5 mmol, 1.1 eq) was dissolved dry DMF (14 mL) and dry ACN (41 mL). $MeNH_2Cl$ (5.68 mg, 84.1 mmol, 1.5 eq) was added. After stirring the mixture at rt for 3 h the solvent was removed *in vacuo*. The crude product was purified by column chromatography (DCM : MeOH, 19:1 → 17:3) to afford the product as a white solid (9.87 g, 78.9 mmol, yield: 94%).

^{1}H-NMR (400 MHz, $CDCl_3$) δ = 8.20 (bs, 1H), 7.73 (d, *J* = 3.6 Hz, 1H, NH), 7.48 (t, *J* = 1.5 Hz, 1H), 7.01 (dd, *J* = 1.5, 0.9 Hz, 1H), 2.99 (d, *J* = 4.6 Hz, 3H, CH_3).
^{13}C NMR (101 MHz, $CDCl_3$) δ 149.81 (s, 1C, CO), 135.87 (s, 1H, C_{Ar}), 129.74 (s, 1H, C_{Ar}), 116.73 (s, 1H, C_{Ar}), 27.49 (s, 1C, NCH_3).
HRMS (ESI^+): calc. for $[C_5H_8N_3O]^+$: 126.0662, found: 126.0664 $[M+H]^+$.

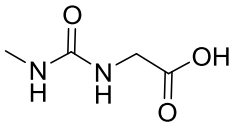
Ethyl (methylcarbamoyl)glycinate (13)



N-methyl-1H-imidazole-1-carboxamide **12** (3.00 g, 23.9 mmol, 2.0 eq) was dissolved in DCM (60 mL)). Ethyl glycinate hydrochloride (1.67 g, 11.9 mmol, 1.5 eq) and NEt_3 (3.32 mL, 23.9 mmol, 2.0 eq) were added. After stirring the mixture at rt for 18 h the solvent was removed *in vacuo*. The crude product was purified by column chromatography (DCM : MeOH, 9:1, 0.1 % NEt_3) to afford the product as a white solid (9.87 g, 78.9 mmol, yield: 94%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 6.21 (t, *J* = 6.0 Hz, 1H, NH), 6.01 (q, *J* = 4.8 Hz, 1H, $NHCH_3$), 4.07 (q, *J* = 7.1 Hz, 2H, CH_2), 3.73 (d, *J* = 6.0 Hz, 2H, CH_2NH), 2.54 (d, *J* = 4.6 Hz, 3H, $NHCH_3$), 1.18 (t, *J* = 7.1 Hz, 3H, CH_3).
^{13}C-NMR (101 MHz, DMSO-d_6) δ = 171.31 (s, 1C, COOEt), 158.53 (s, 1C, N_2CO), 60.17 (s, 1C, OCH_2), 41.56 (s, 1C, $NHCH_2$), 26.36 (s, 1C, NCH_3), 14.16 (s, 1C, CH_3).
HRMS (ESI$^+$): calc. for $[C_6H_{13}N_2O_3]^+$: 161.0924, found: 161.0921 $[M+H]^+$.
mp: 83 °C
IR (cm^{-1}): ṽ = 3341 (m), 2936 (w), 1754 (m), 1635 (m), 1575 (s), 1479 (m), 1456 (w), 1424 (w), 1407 (m), 1370 (m), 1283 (m), 1183 (s), 1115 (m) 1065 (m), 1019 (s), 910 (w), 870 (w), 779 (w).

(Methylcarbamoyl)glycine (9a)



Ethyl (methylcarbamoyl)glycinate **13** (0.98 g, 6.12 mmol, 1.0 eq) was dissolved in 10% KOH (5.0 mL). After stirring the mixture at rt 1 h it was treated with Amberlite IR 120 (H^+ type) and the solvent was removed in vacuo. The product could be afforded as a white solid (0.53 g, 3.99 mmol, yield: 65%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 6.12 (t, *J* = 5.9 Hz, 1H, $NHCH_2$), 6.01 (q, *J* = 4.9 Hz, 1H, $NHCH_3$), 3.67 (q, *J* = 5.9 Hz, 2H, CH_2), 2.54 (d, *J* = 4.6 Hz, 3H, NCH_3).

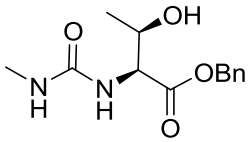
^{13}C-NMR (101 MHz, DMSO-d_6) δ = 172.73 (s, 1C, COOH), 158.64 (s, 1C, N_2CO), 41.60 (s, 1C, CH_2), 26.39 (s, 1C, NCH_3).

HRMS (ESI^-): calc. for $[C_4H_7N_2O_3]^-$: 131.0462, found: 131.0462 $[M-H]^-$.

mp: 144°C

IR (cm^{-1}): ṽ = 3390 (m), 3367 (m), 2363 (br., w), 1959 (br., w), 1707 (m), 1600 (s), 1550 (s), 1496 (m), 1442 (s), 1414 (m), 1354 (m), 1234 (s) 1087 (m), 1040 (m), 1005 (m), 912 (m), 772 (w), 740 (w).

Benzyl (methylcarbamoyl)-*L*-threoninate (15)



N-methyl-1H-imidazole-1-carboxamide **12** (1.50 g, 11.9 mmol, 2.0 eq) was dissolved in DCM (30 mL). Threonine benzyl ester hydrochloride (1.48 g, 5.98 mmol, 1 eq) and NEt_3 (1.80 mL, 11.9 mmol, 2.0 eq) were added. After stirring the mixture at rt for 18 h the solvent was removed *in vacuo*. The crude product was purified by column chromatography (MeOH : Hex : EtOAc, 2:4.5:4.5) to afford the product as a white solid (0.78 mg, 4.45 mmol, yield: 74%).

^{1}H-NMR (400 MHz, DMSO-d_6) δ = 7.41 – 7.30 (m, 5H, H_{Ar}), 6.23 (q, *J* = 4.6 Hz, 1H, $NHCH_3$), 6.08 (d, *J* = 9.0 Hz, 1H, NH), 5.12 (s, 2H, CH_2C_{Ph}), 5.12 (d, *J* = 4.6 Hz, 1H, OH), 4.13-4.19 (m, 2H, CHOH, CHNH), 2.56 (d, *J* = 4.6 Hz, 3H, NCH_3), 1.07 (d, *J* = 6.1 Hz, 3H, CH_3).

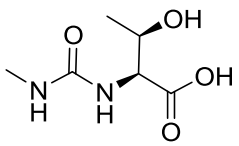
^{13}C-NMR (101 MHz, DMSO-d_6) δ = 172.02 (s, 1C, COOH), 158.70 (s, 1C, N_2CO), 136.20 (s, 1C, C_{Ph}), 128.41 (s, 2C, C_{Ph}), 127.93 (s, 1C, C_{Ph}), 127.65 (s, 2C, C_{Ph}), 66.57 (s, 1C, CHCOOH), 65.68 (s, 1C, CHOH), 26.21 (s, 1C, NCH_3), 20.48 (s, 1C, $CHCH_3$).

HRMS (ESI^+): calc. for $[C_{13}H_{19}N_2O_4]^+$: 267.1339, found: 267.1340 $[M+H]^+$.

mp: 136°C

IR (cm^{-1}): ṽ = 3467 (w), 3349 (w), 1707 (m), 1628 (s), 1588 (m), 1536 (w), 1500 (w), 1448 (w), 1377 (w), 1290 (s), 1127 (w), 1082 (m), 1021 (s) 870 (w), 727 (s), 692 (m).

(Methylcarbamoyl)-*L*-threonine (9b)



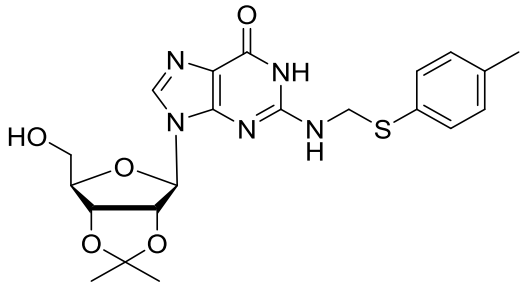
Benzyl (methylcarbamoyl)-L-threoninate **15** (0.10 g, 0.38 mmol, 1.0 eq) was dissolved in MeOH (5.0 mL). Pd/C (10.0 mg, 10 wt%) was added and the mixture was stirred in an H_2-

atmosphere. After 1 h the catalysator was filtered off and the solvent was removed *in vacuo* at 25°C. The product could be afforded as a colorless oil (58.0 mg, 0.33 mmol, yield: 88%).

^{1}H-NMR (400 MHz, D_2O) δ = 4.32 (m, 1H, CHOH), 4.23 (m, 1H, CHNHCOOH), 2.69 (s, 3H, NCH_3), 1.19 (d, *J* = 6.4 Hz, 3H, $CHCH_3$).
^{13}C-NMR (400 MHz, D_2O) δ = 173.58 (s, 1C, COOH), 158.66 (s, 1C, N_2CO), 65.31 (s, 1C, CHCOOH), 56.67 (s, 1C, CHOH), 24.01 (s, 1C, NCH_3), 16.58 (s, 1C, $CHCH_3$).
HRMS (ESI^-): calc. for $[C_6H_{11}N_2O_4]^-$ 175.0724, found: 175.0723 $[M-H]^-$.
IR (cm^{-1}): ṽ = 3354 (w), 2976 (w), 2936 (w), 1709 (m), 1622 (s), 1500 (m), 1550 (m), 1414 (m), 1277 (m), 1223 (m), 1185 (s), 1122 (m) 1077 (s), 1014 (m) 949 (w), 904 (w), 862 (m), 761 (w).

Synthesis of synthetic nucleoside standards

9-((3aR,4R,6R,6aR)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-2-(((p-tolylthio)methyl)amino)-1,9-dihydro-6H-purin-6-one (16)



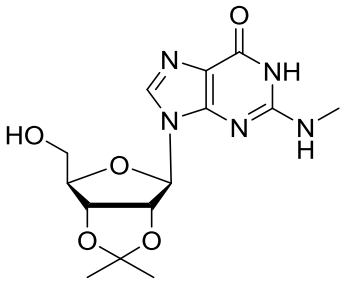
The reaction was carried out slightly modified to Bridson *et al*..[5]

2‘,3‘-O-isopropylideneguanosine (2.50 g, 7.73 mmol, 1 eq.) and p-thiocresole (1.25 g, 10.1 mmol, 1.3 eq.) were suspended in EtOH (60.0 mL). Acetic acid (1.9 mL) and formaldehyde (2.0 mL) were added to the suspension and the mixture was heated to 90°C. After 14 h the solvent was removed *in vacuo* and the crude product was washed with EtOH to afford the product as a white solid (2.75 g, 5.88 mmol, yield: 76%).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.79 (s, 1H, $N_{Ar}H$), 7.95 (s, 1H, HC8), 7.41 – 7.31 (m, 2H, *J* = 8.0 Hz, HC_{Ar}), 7.17 (d, *J* = 8.0 Hz, 2H, HC_{Ar}), 6.01 (d, *J* = 2.6 Hz, 1H, HC1’), 5.36 (dd, *J* = 6.3, 2.7 Hz, 1H, HC2’), 5.02 (t, *J* = 5.5 Hz, 1H, OH), 4.95 (dd, *J* = 6.3, 3.0 Hz, 1H, HC3’), 4.84 (h, *J* = 6.7, 6.3 Hz, 2H, CH_2C_{Ar}), 4.13 (td, *J* = 5.5, 3.0 Hz, 1H, HC4’), 3.54 (hept, *J* = 5.7 Hz, 2H, HC5’), 2.28 (s, 3H, $C_{Ar}CH_3$), 1.54 (s, 3H, CH_3), 1.35 (s, 3H, CH_3).
^{13}C NMR (101 MHz, DMSO-d_6) δ = 154.14 (C6), 151.76 (C2), 149.98 (C4), 137.49 (C8), 136.99 ($C_{Ar}CH_3$), 136.31 ($C_{Ar}S$), 131.43 ($C_{Ar}S$), 130.24 (2C, C_{Ar}), 118.02 (C5), 113.49

($C(CH_3)_2$), 89.17 (C1'), 87.10 (C4'), 83.75 (C2'), 81.74 (C3'), 61.99 (C5'), 46.71 (CH_2C_{Ar}), 27.41 (CH_3), 25.52(CH_3), 21.04 ($C_{Ar}CH_3$).
HRMS (ESI^+): calc. for $[C_{21}H_{26}N_5O_5S]^+$ 460.1649, found: 460.1650 $[M+H]^+$.

9-((3aR,4R,6R,6aR)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-2-(methylamino)-1,9-dihydro-6H-purin-6-one (17)

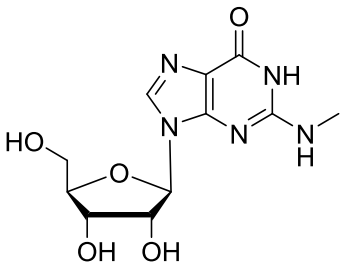


The reaction was carried out slightly modified to Bridson *et al*..[5]

Compound **16** (0.65 g, 1.41 mmol, 1 eq.) was dissolved in DMSO (9.0 mL). $NaBH_4$ (0.11 g, 2.81 mmol, 2 eq.) was added and the mixture was heated to 100°C. After cooling to rt the mixture was purified by column chromatography (DCM : MeOH, 9:1) to afford the product as a white solid (0.41 g, 1.22 mmol, yield: 86%).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.80 (s, 1H, $N_{Ar}H$), 7.90 (s, 1H, HC8), 6.46 (m, 1H, $NHCH_3$), 6.00 (d, *J* = 2.6 Hz, 1H, HC1'), 5.31 (dd, *J* = 6.2, 2.6 Hz, 1H, HC2'), 5.00 (t, *J* = 5.5 Hz, 1H, OH), 4.95 (dd, *J* = 6.2, 3.0 Hz, 1H, HC3'), 4.12 (td, *J* = 5.5, 3.0 Hz, 1H, HC4'), 3.53 (m, 2H, HC5'), 2.82 (d, *J* = 4.6 Hz, 3H, NCH_3), 1.53 (s, 3H, CH_3), 1.32 (s, 3H, CH_3).
^{13}C NMR (101 MHz, DMF-d_7) δ = 157.49 (C2), 154.08 (C6), 150.98 (C4), 136.86 (C8), 117.84 (C5), 113.59 ($C(CH_3)_2$), 89.79 (C1'), 87.52 (C4'), 84.22 (C2'), 82.17 (C3'), 62.50 (C5'), 27.88 (NCH_3), 27.03 (CH_3), 25.06 (CH_3).
HRMS (ESI^+): calc. for $[C_{14}H_{20}N_5O_5]^+$ 338.1457, found: 338.1459 $[M+H]^+$.

9-((2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-(methyl-amino)-1,9-dihydro-6H-purin-6-one (m^{2}G, 18)

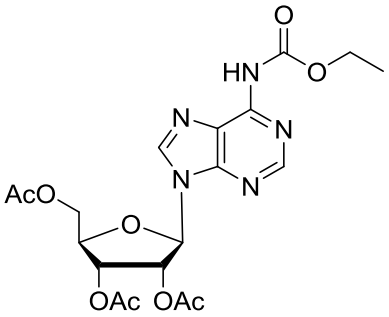


The reaction was carried out slightly modified to Ubiali *et al.*.[6]

The protected nucleoside **17** (0.28 g, 0.83 mmol, 1 eq.) was dissolved in H_2O (25.0 mL) and TFA (25.0 mL). After stirring for 1.5 h the solvent was removed *in vacuo*. The residue was suspended in acetone and H_2O. The solvent was removed *in vacuo* and the crude product was recrystallized in EtOH/ H_2O to yield the product as a white solid (0.16 mg, 0.54 mmol, yield: 65%).

^{1}H-NMR (400 MHz, CD_3OD) δ = 10.77 (s, 1H, $N_{Ar}H$), 7.95 (s, 1H, HC8), 6.28 (d, ^3J = 5.0 Hz, 1H, NH), 5.73 (d, ^3J = 6.0 Hz, 1H, HC1'), 4.52 (t, J = 5.5 Hz, 1H, HC2'), 4.12 (dd, J = 5.1, 3.4 Hz, 1H, HC3'), 3.88 (d, J = 3.9 Hz, 1H, HC4'), 3.75 – 3.45 (m, 2H, HC5'), 2.81 (d, ^3J = 4.7 Hz, 3H, NCH_3).
^{13}C-NMR (101 MHz, CD_3OD) δ = 156.77 (C6), 153.29 (C2), 150.97 (C4), 136.15 (C8), 116.65 (C5), 86.79 (C1'), 85.31 (C4'), 73.39 (C2'), 70.55 (C3'), 61.60 (C5'), 27.74 (NCH_3).
HRMS (ESI^+): calc. for $[C_{11}H_{16}N_5O_4]^+$ 298.1146, found: 298.1142 $[M+H]^+$.

Ethyl 9-(2',3',5'-tri-*O*-acetyl-*β*-D-ribofuranosyl)-9*H*-purine-6-carbamate (19)[7]



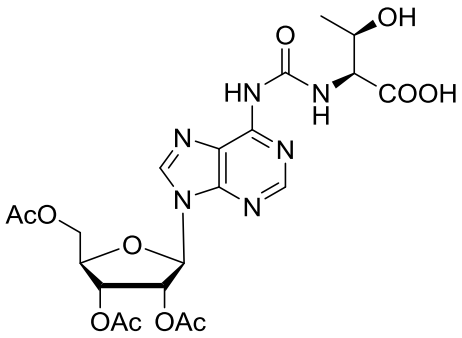
Acetyl protected adenosine (2.00mg, 5.08 mmol, 1 eq) was dissolved in absolute pyridine (37 mL) and cooled to 0°C. Ethyl chloroformate (33.8 µL, 0.35 mmol, 3 eq) was then added dropwise with stirring during 30 min. The mixture was allowed to warm to rt and stirred overnight. The solvent was removed *in vacuo* by co-evaporation with toluene and the crude product was purified by column chromatography (DCM : MeOH, 100:1) to afford the product as a colorless foam (1.05 g, 2.26 mmol, yield: 44%).

^{1}H NMR: (400 MHz, DMSO-d_6) δ: 10.56 (s, 1H, NH), 8.66 (s, 1H, HC8), 8.65 (s, 1H, HC2), 6.31 (d, J=5.3, HC1'), 6.06 (t, 1H,J = 5.6 Hz, HC2'), 5.65 (t, 1H,J = 5.6 Hz, HC3'), 4.46-4.33 (m, 2H, HC4', HC5' [a]), 4.32-4.22 (m, 1H, HC5'[b]), 4.18 (q, 1H, J = 7.1 Hz, CH_2), 2.12 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.26 (7, 3H, J=7.1, CH_3).

^{13}C NMR: (101 MHz, DMSO-d_6) δ: 170.03 ($COCH_3$), 169.45 ($COCH_3$), 169.28 ($COCH_3$), 152.21 (C(NH)(OCH_2CH_3)) 151.92 (C6), 151.41(C2), 150.11 (C4), 143.16 (C8), 123.93 (C5), 85.76 (C1'), 79.53 (C4'), 71.90 (C2'), 69.99 (C3'), 62.72 (C5'), 61.02 (CH_2), 20.49 ($COCH_3$), 20.38 ($COCH_3$), 20.21 ($COCH_3$), 14.37 (CH_3).

HRMS (ESI^+): calc. for $[C_{19}H_{24}N_5O_9]^+$ 466.1569, found: 466.1569 $[M+H]^+$.

2-(3-(9-((2R,3R,4R,5R)-3,4-diacetoxy-5-(acetoxymethyl)tetrahydrofuran-2-yl)-9H-purin-6-yl)ureido)-3-hydroxybutanoic acid (20)



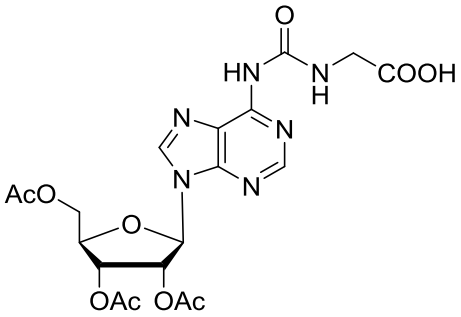
Carbamate **19** (0.30 g, 0.64 mmol, 1 eq) was dissolved in absolute pyridine (30 mL) and L-threonine (0.15 g, 1.29 mmol, 2 eq) was added. The mixture was stirred under reflux for 6 h and cooled down to room temperature. Excess of L-threonine was removed by filtration and the. The solvent was removed *in vacuo* by co-evaporation with toluene and the crude product was recrystallized in EtOH to afford the product as a white solid (0.29 g, 0.54 mmol, yield: 84%).

^{1}H NMR: (400 MHz, DMSO-d_6) δ: 9.92 (s, 1H, COOH), 9.68 (d, *J* = 8.4 Hz, 1H, NH), 8.66 (s, 1H, HC8), 8.58 (s, 1H, HC2), 6.31 (d, *J* = 5.6 Hz, 1H, HC1'), 6.04 (t, *J* = 5.7 Hz, 1H HC2'), 5.62 (dd, *J* = 5.9, 4.4 Hz, 1H, HC3'), 4.48 – 4.36 (m, 2H, CHOH, CHNH), 4.36 – 4.13 (m, 3H, HC5',HC4'), 2.14 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc), 1.15 (d, *J* = 6.2 Hz, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6) δ = 172.85 (COOH), 170.53 ($COCH_3$), 169.94 ($COCH_3$), 169.74 ($COCH_3$), 154.12 (CON_2), 151.61 (C2), 150.94 (C6), 150.56 (C4), 143.09 (C8), 121.00 (C5), 86.07 (C1´), 80.11 (COH), 72.42 (C4´), 70.50 (C2´), 66.65 (C3´), 63.26 (CHCOOH), 59.09 (C5´), 21.26 ($COCH_3$), 20.99 ($COCH_3$), 20.86 ($COCH_3$), 20.67 (CH_3) ppm.

HRMS (ESI^+): calc. for $[C_{21}H_{27}N_6O_{11}]^+$ 539.1732, found: 539.1736 $[M+H]^+$.

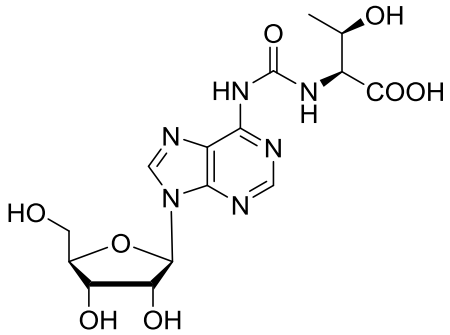
((9-((2R,3R,4R,5R)-3,4-diacetoxy-5-(acetoxymethyl)tetrahydrofuran-2-yl)-9H-purin-6-yl)carbamoyl)glycine (21)



Carbamate **19** (0.75 g, 1.61 mmol, 1 eq) was dissolved in absolute pyridine (60 mL) and L-glycine (0.24 g, 3.22 mmol, 2 eq) was added. The mixture was stirred under reflux for 7 h and cooled down to room temperature. Excess of glycine was removed by filtration and the. The solvent was removed *in vacuo* by co-evaporation with toluene and the crude product was recrystallized in EtOH to afford the product as a white solid (0.24 g, 0.50 mmol, yield: 31%).

^{1}H NMR: (400 MHz, **DMSO-d_6**) δ: 12.69 (s, 1H, COOH), 9.99 (s, 1H, NH), 9.61 (bs, 1H, NH), 8.65 (s, 1H, HC8), 8.59 (s, 1H, HC2), 6.30 (d, *J* = 5.3 Hz, 1H, HC1‘), 6.03 (t, *J* = 5.6 Hz, 1H HC2‘), 5.63 (t, *J* = 5.3, 1H, HC3‘), 4.45 – 4.38 (m, 2H, HC5‘), 4.31 – 4.20 (m, 1H, HC4‘), 3.99 (d, *J* = 5.3 Hz, 1H, CH_2), 2.12 (s, 3H, $COCH_3$), 2.04 (s, 3H, COCH), 2.01 (s, 3H, COCH).
HRMS (ESI+): calc. for $[C_{19}H_{23}N_6O_{10}]^+$ 495.1470, found: 495.1466 $[M+H]^+$.

t^{6}A (22)

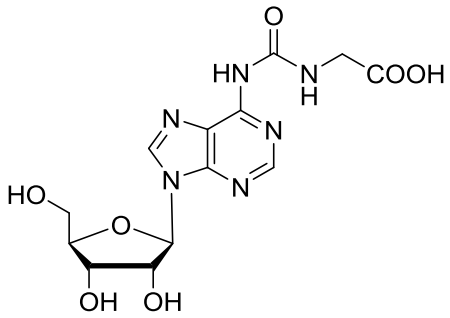


Protected nucleoside **20** (0.17 g, 0.32 mmol, 1 eq) was dissolved in 7 N Ammonia in MeOH (8 mL). The mixture was heated to 40°C for 1 h and then stirred overnight at rt. The solvent was removed *in vacuo* and the crude product was recrystallized in EtOH to afford the product as a white solid (0.97 g, 0.24 mmol, yield: 75%).

^{1}H NMR: (400 MHz, **DMSO-d_6**) δ: 9.61 (d, *J* = 6.4 Hz, 1H, NH), 8.64 (s, 1H, HC8), 8.52 (s, 1H, HC2), 5.98 (d, *J* = 5.6 Hz, 1H, HC1‘), 4.59 (t, *J* = 5.3 Hz, 1H HC2‘), 4.18 (t, *J* = 4.1 Hz, 1H, HC3‘), 4.06 – 4.01 (m, 2H, CHOH, CHNH), 3.96 (q, *J* = 3.8 Hz, 1H, HC4‘), 3.63 (ddd, *J* = 48.7, 12.0, 4.0 Hz, HC5‘), 1.00 (d, *J* = 6.1 Hz, 3H, CH_3).

^{13}C-NMR (101 MHz, DMSO-d_6) δ = 172.61 (COOH), 153.21 (CON_2), 151.42 (C2), 150.83 (C6), 150.62 (C4), 142.49 (C8), 120.80 (C5), 88.13 (C1´), 86.08 (COH), 74.22 (C4´), 70.69 (C2´), 66.48 (C3´), 61.70 (C5´), 58.86 (CHCOOH), 19.89 (CH_3) ppm.
HRMS (ESI^+): calc. for $[C_{15}H_{21}N_6O_8]^+$ 413.1415, found: 413.1415 $[M+H]^+$.

g^{6}A (21)



Protected nucleoside **21** (202 mg, 0.41 mmol, 1 eq) was dissolved in 7 N Ammonia in MeOH (10 mL). The mixture was heated to 40°C for 1 h and then stirred overnight at rt. The solvent was removed *in vacuo* and the crude product was recrystallized in EtOH to afford the product as a white solid (74.0 mg, 0.20 mmol, yield: 49%).

^{1}H NMR: (400 MHz, **DMSO-d_6**) δ: 9.55 (d, *J* = 6.4 Hz, 1H, NH), 8.66 (s, 1H, HC8), 8.54 (s, 1H, HC2), 5.99 (d, *J* = 5.6 Hz, 1H, HC1‘), 4.18 (dd, *J* = 4.9, 3.7 Hz, 1H HC2‘), 3.97 (t, *J* = 4.1 Hz, 1H, HC3‘), 3.78 (d, *J* = 4.9 Hz, 1H, HC4‘), 3.64 (ddd, *J* = 48.7, 12.0, 4.0 Hz, HC5‘).

^{13}C-NMR (101 MHz, DMSO-d_6) δ = 171.67 (COOH), 153.51 (C2), 151.34 (C6), 150.74 (C4), 142.56 (C8), 120.73 (C5), 88.14 (C1´), 86.14 (C4´), 74.24 (C2´), 70.73 (C3´), 61.73 (C5´), 43.85 (CH_2) ppm.
HRMS (ESI^+): calc. for $[C_{13}H_{17}N_6O_7]^+$ 369.1153, found: 369.1152 $[M+H]^+$.

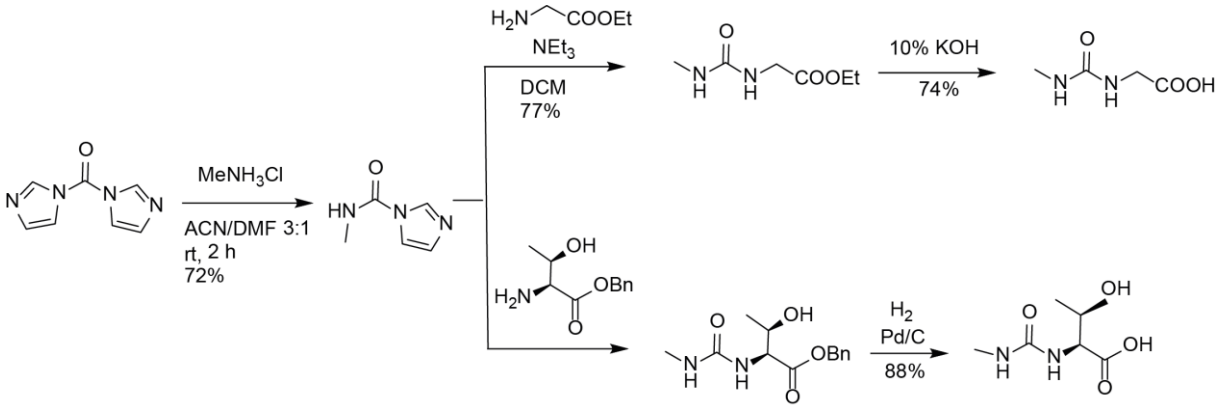


Figure S 1: Reaction scheme of the organic synthesis of the amino acid urea compounds.

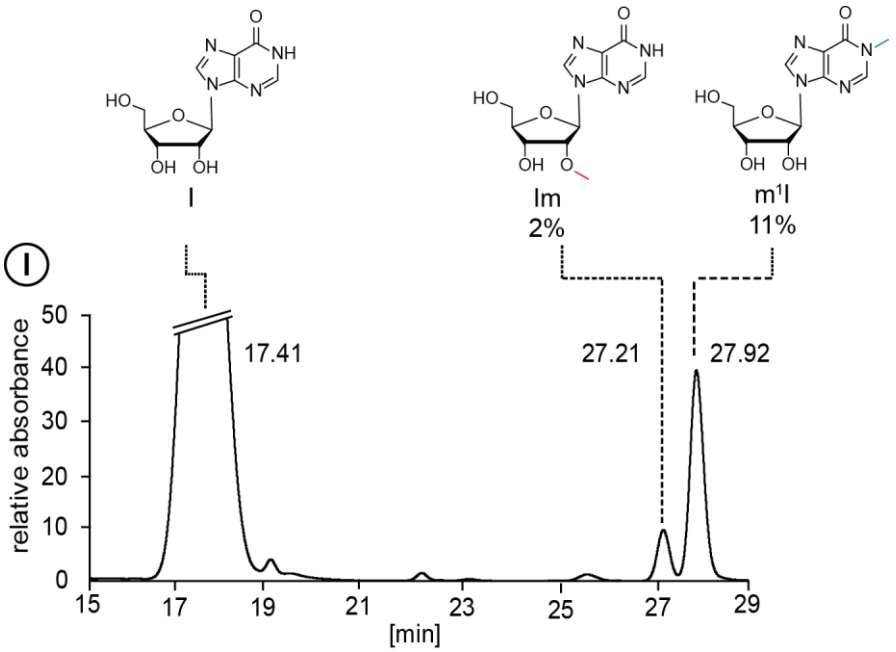


Figure S 2: UV-chromatogram (at 260 nm) of the reaction of Inosine with *N*-methly-*N*-nitroso-urea under prebiotic conditions.

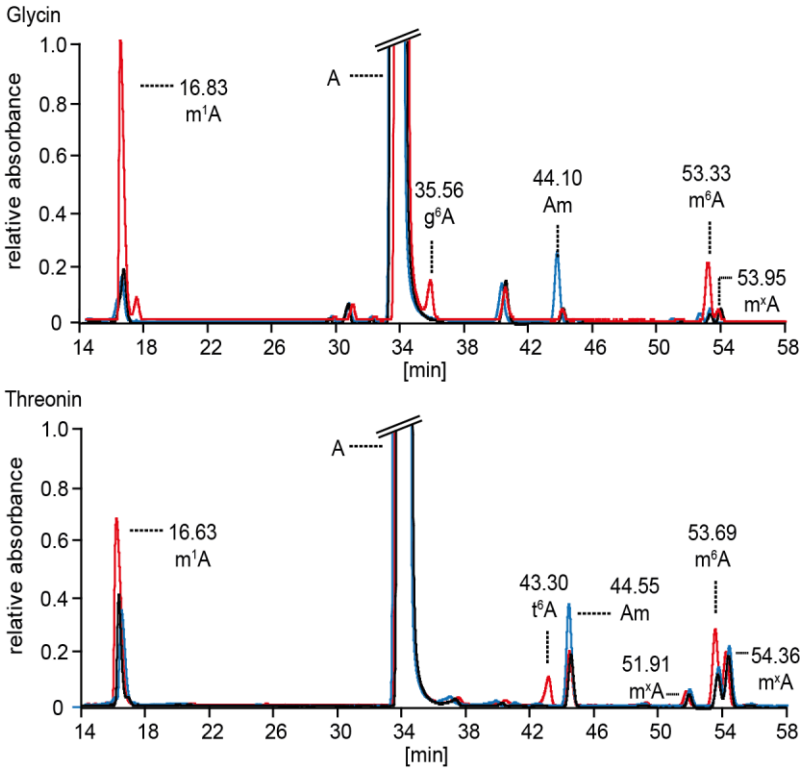


Figure S 3: UV-chromatogram (at 260 nm) of the reaction of Adenosine with amino acid-nitroso compounds under prebiotic conditions. Chromatogram of the reaction itself is shown in black, coinjections of the different modifications are shown in blue and red.

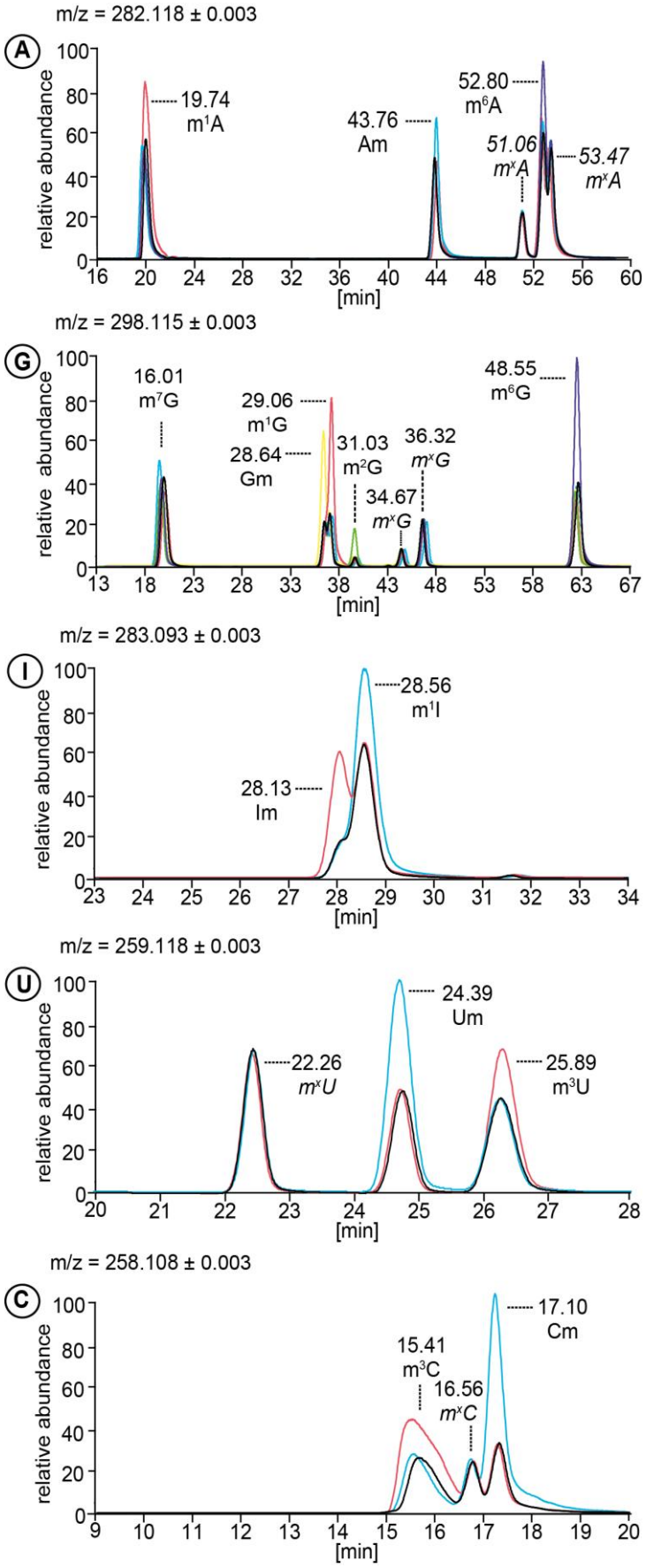


Figure S 4: MS-chromatogram of the reaction of the different canonical nucleosides and Inosine with *N*-methly-*N*-nitroso-urea (black) and the coinjections with synthetic standards (red, blue, yellow).

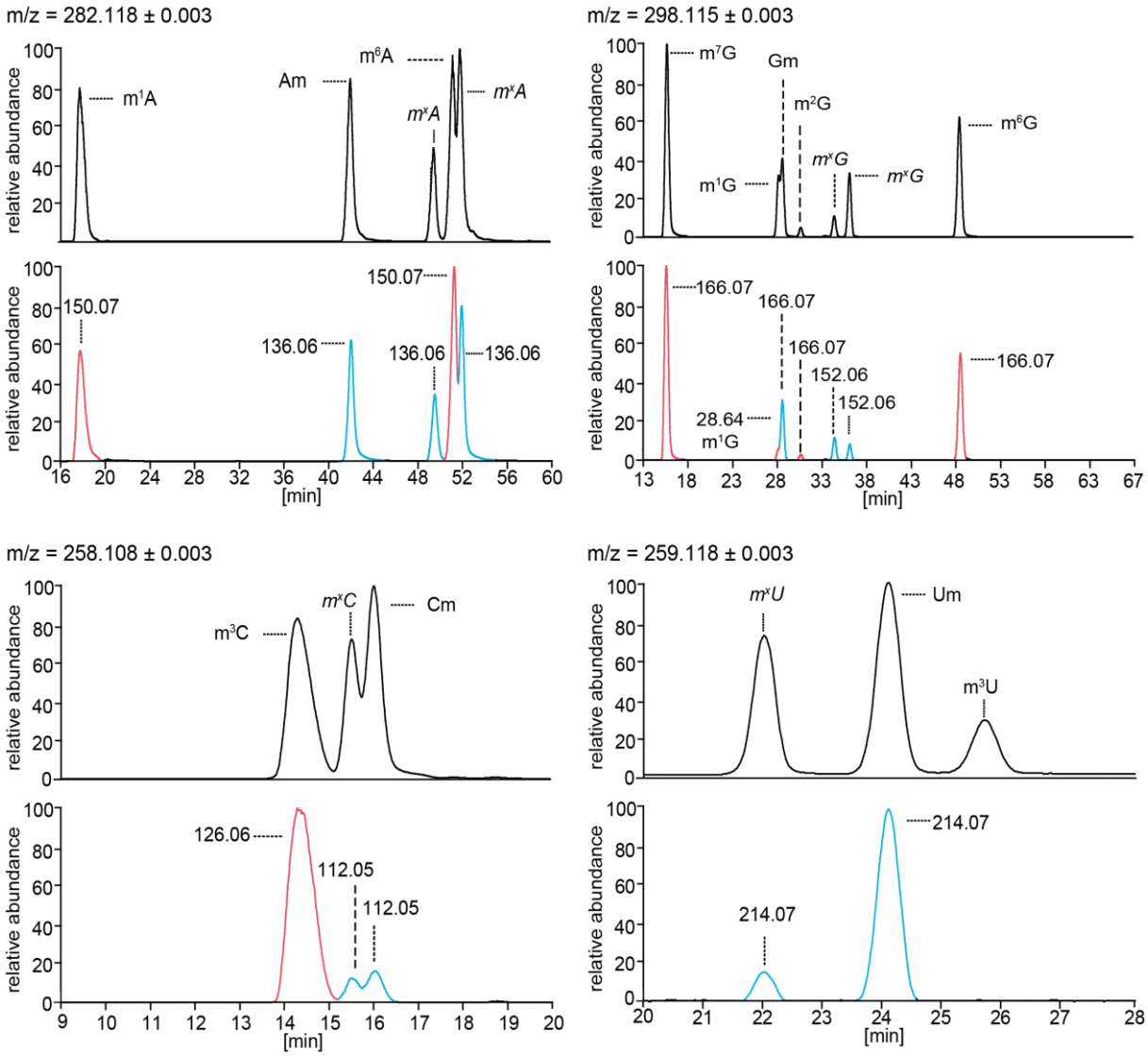


Figure S 5: Fragmentation studies of the methylation products of the prebiotic reaction of *N*-methly-*N*-nitroso-urea with the canonical nucleosides. Sugar modified nucleosides are shown in blue, and the nucleosides with a methylation at the base in red. The numbers represent the base peak of the modification.

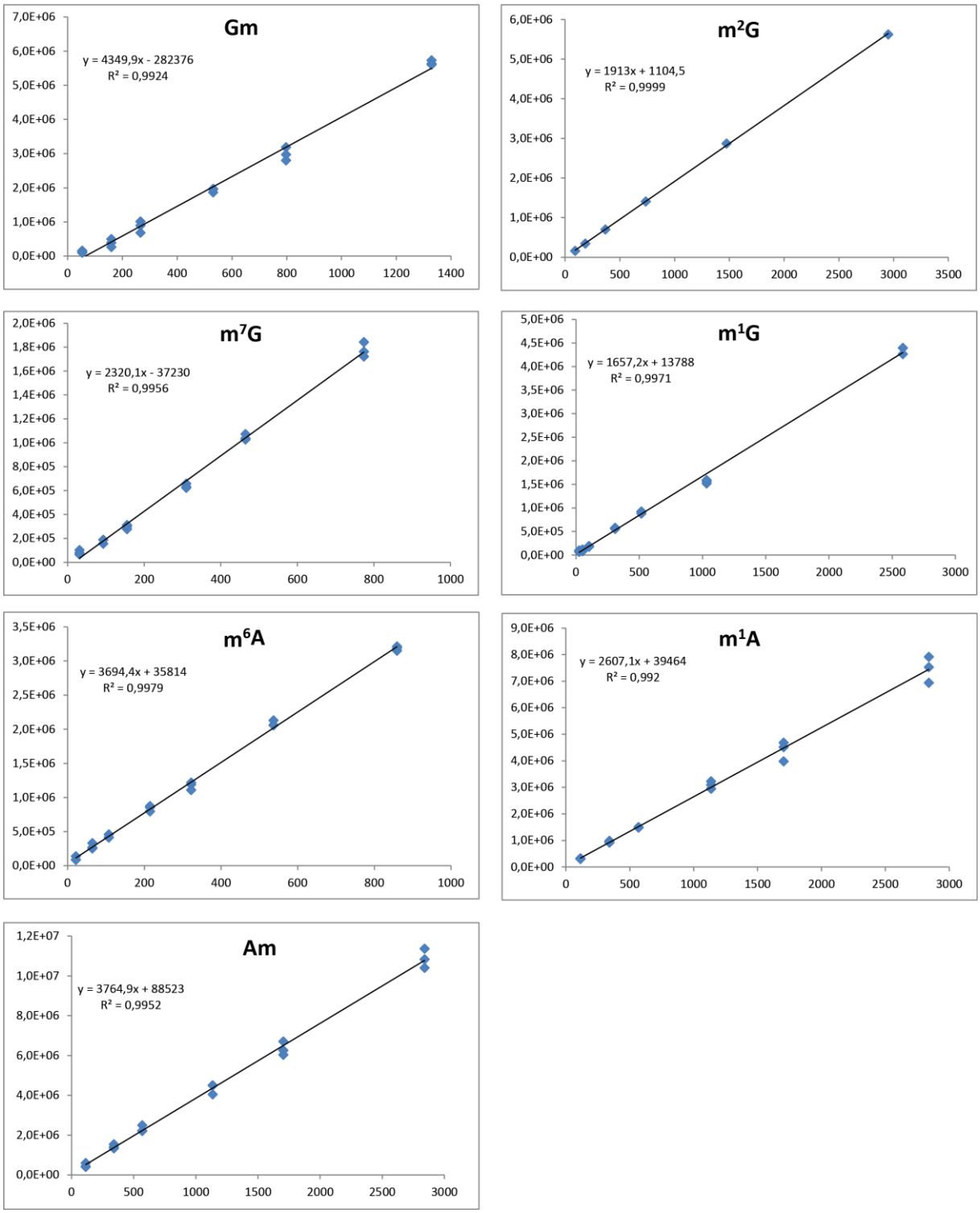


Figure S 6: Calibration curves at 260 nm for quantification of modified A and G nucleosides from UV-signals.

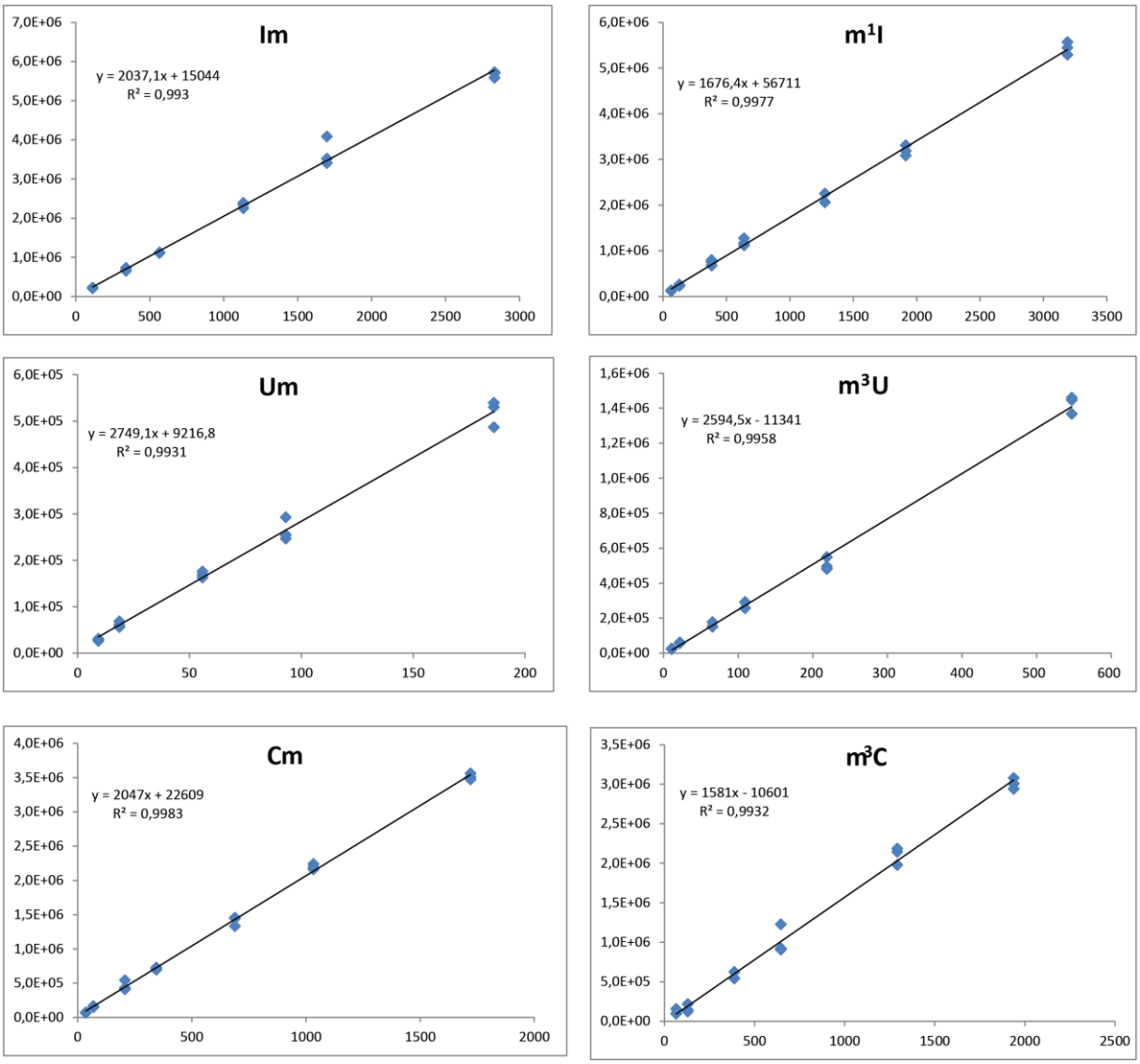


Figure S 7: Calibration curves at 260 nm for quantification of modified C, U and I nucleosides from UV-signals.

Literature:

[1] F. Arndt, *Org. Synth.* **1943**, 3-3.
[2] T. Machinami, T. Suami, *Bull. Chem. Soc. Jpn.* **1975**, *48*, 1333-1334.
[3] F. U. Richter, *Chem. Eur. J.* **2009**, *15*, 5200-5202.
[4] P. A. Duspara, M. S. Islam, A. J. Lough, R. A. Batey, *J. Org. Chem.* **2012**, *77*, 10362-10368.
[5] P. K. Bridson, C. B. Reese, *Bioorg. Chem.* **1979**, *8*, 339-349.
[6] D. Ubiali, C. D. Serra, I. Serra, C. F. Morelli, M. Terreni, A. M. Albertini, P. Manitto, G. Speranza, *Adv. Synth. Catal.* **2012**, *354*, 96-104.
[7] M. Matuszewski, E. Sochacka, *Bioorganic Med. Chem. Lett.* **2014**, *24*, 2703-2706.

6 Resultate: nicht-publizierte Arbeiten

6.1 Simultaneous formation of all four Watson-Crick RNA building blocks under plausible prebiotic conditions

Sidney Becker[#], Hidenori Okamura[#], Stefan Wiedemann, Martin Rossa, Antony Crisp, Katharina Iwan, Tynchtyk Amatov, Christina Schneider, and Thomas Carell

[#] geteilte Erstautorenschaft

Prolog

Als Voraussetzung für das RNA-Welt-Konzept ist die Synthese von RNA unter präbiotischen Bedingungen essentiell. Dafür bedarf es der Verfügbarkeit aller Purin- und Pyrimidin-RNA-Bausteine. Dies ist notwendig, um so Watson-Crick Basenpaarung (A:U und G:C) zu ermöglichen. Bisher gibt es jedoch keinen plausiblen Reaktionspfad, welcher alle vier RNA Bausteine unter den gleichen Bedingungen verfügbar machen kann. Alle bekannten präbiotisch plausiblen Synthesen können nur einen Teil der vier kanonischen RNA Bausteine darstellen. Im folgendem Manuskript wurde ein neuer Reaktionspfad ausgearbeitet, welcher Pyrimidinnukleoside und 5‘-Nukleotide in nur vier Schritten zugänglich machen kann. Die Bedingungen sind gleichzeitig kompatibel mit der Synthese von Purinnukleosiden über unseren entwickelten Formamidopyrimidin-Weg (siehe Abschnitt 5.2). Ebenfalls konnte nachgewiesen werden, dass eine nachfolgende Phosphorylierung sehr selektiv die natürlich vorkommenden 5'-mono- und 5'-di-phosphorylierten Ribofuranoside liefert. Dies könnte eine mögliche Lösung für die Frage sein, warum die thermodynamisch ungünstigen Furanside nicht aber die Pyranoside ausschließlich im Rückgrat von RNA und DNA zu finden sind.

Autorenbeitrag

Entwicklung des Syntheseweges, sowie von LC-MS Methoden zur Trennung und Detektion verschiedener Nukleosid bzw. Nukleotid Isomere. Nachweis von 5‘-Nukleotiden. Auswertung und Interpretation von Daten.

Simultaneous formation of all four Watson-Crick RNA building blocks under plausible prebiotic conditions

Sidney Becker*, Hidenori Okamura*, Stefan Wiedemann, Martin Rossa, Antony Crisp, Katharina Iwan, Tynchtyk Amatov, Christina Schneider, and Thomas Carell#

Center for Integrated Protein Science (CiPS^{M}) at the Department of Chemistry, LMU München
Butenandtstr. 5-13, 81377 München
Fax: (+) 49 2180 77756
E-mail: Thomas.Carell@lmu.de
Homepage: www.carellgroup.de

* These authors contributed equally

corresponding author: thomas.carell@lmu.de

Abstract

The RNA world hypothesis suggests that life first emerged from (self)-replicating RNA. From this genetic polymer a chemical evolution process generated complex integrated chemical systems, followed by the emergence of life. Replication of RNA requires formation of complementary pyrimidine-purine Watson-Crick base pairs, which are a prerequisite for accurate genetic information transfer. Although separate prebiotic pathways to either purine or pyrimidine RNA building blocks have been reported, the question of how all four nucleosides could have formed in parallel remains unanswered. Here, we report a prebiotically plausible pathway that provides all four constituents of RNA simultaneously, and in high yields. The chemistry also selectively affords the naturally occurring 5'-mono- and 5'-di-phosphorylated ribofuranosides, providing a rationale for why these thermodynamically disfavoured constitutional isomers are exclusively present in the backbone of RNA and DNA. The chemistry reported here demonstrates, that all constituents of RNA, needed for Watson-Crick base pairing, could have arisen simultaneously on early Earth.

Main Text

The discovery of catalytic RNA[1,2] and the development of replicating RNA systems[3-5] has lent strong support to the concept of an RNA world. The RNA world hypothesis predicts that life started with RNAs that were able to (self)-recognize and replicate. Through a process of chemical evolution a complex RNA and later RNA-protein world evolved, from which life ultimately emerged.[6] RNA is today constructed from four nucleotide building blocks, which can be categorized into two different classes, the purines (A and G) and the pyrimidines (C, U). These nucleotides form the two complementary H-bonded pyrimidine-purine Watson-Crick base pairs A:U and G:C, needed for faithful information transfer. The strict complementarity of the Watson-crick base pairs allows for template directed replication of RNA. Watson-Crick base pairing and the fact that RNA has the ability to catalyse reactions, allow us today to conceptually compass a (self)-replicating and evolving RNA world, despite the unreached goal to experimentally simulate such a system.

The prerequisite for the RNA world is the ability to create RNA under prebiotic conditions. This requires as the first elementary step, the concomitant formation of the necessary pyrimidine and purine constituents. The question of how the four RNA building blocks could have emerged simultaneously in the same environment is currently intensively investigated, but yet unsolved.[7-11] The main obstacle to form RNA nucleosides under early Earth conditions is the formation of the glycosidic C-N bond. Reaction of the canonical pyrimidine and purine heterocycles with ribose is highly inefficient due to low nucleophilicity of the nucleobases. Whereas pyrimidines do not react at all, the purine nucleosides are formed in very low yields and only as a mixture of different regioisomers.[12] To date, even the best prebiotically plausible pathways generate only a subset of the four RNA constituents. The purine nucleosides can for example be generated from more nucleophilic nucleobase precursors that are first reacted with ribose followed by a subsequent conversion into the full nucleobase.[7,13] For the pyrimidines both the nucleobase and the sugar have been constructed in a stepwise process, bypassing the need for ribose.[8] Since these pathways give either purines or pyrimidines under vastly different conditions, the spontaneous evolution of a Watson Crick base pairing system is not possible.

Herein, we report the prebiotic formation of all four RNA building blocks in the same geochemical environment. We discovered a new chemical route that affords pyrimidine and purine nucleosides, as well as 5'-nucleotides. The chemistry was developed around the prebiotically well-established compound cyanoacetylene in addition to a small number of simple prebiotically plausible molecules, such as ammonia, hydroxylamine, urea and different amidines (Fig. 1). These can be assembled into isoxazolyl-urea as a pyrimidine precursor or into formamidopyrimidines (FaPys) to provide the purines (Fig. 1). Importantly, the chemistry allows for selective formation of 5'-mono- and 5'-di-phosphorylated ribofuranosides as the central constituents of DNA and RNA, providing an explanation why nature selected the 5-membered sugar ring isomer.[14,15] Our results show that all central constituents of RNA could have been part of the same prebiotic nucleoside/tide pool, from which RNA at some point evolved (Fig. 1).

We started with the development of a new plausible prebiotic pathway towards the pyrimidine nucleosides and 5'-nucleotides starting from cyanoacetylene **1**, as shown in Fig. 2a and b. Cyanoacetylene **1** was considered as the starting material because it is ubiquitously found throughout the universe[16-18] and it is, as well as formamide, one of the few molecules that can efficiently form the pyrimidine heterocycle.[19-21] The reaction of **1** with hydroxylamine in water under basic conditions was

shown to give 3-aminoisoxazole **2** with high yields (67%) and under full regiocontrol. The reaction yield was further improved, when we reacted **1** with hydroxylurea **3**, which can be obtained via the reaction of hydroxylamine with cyanate.[22,23] LC-MS analysis (Fig. 3a) shows that the formation of **2** is clean and that the yield is almost quantitative (95%). Hydroxylamine itself is prebiotically available from the reduction of nitrites with SO_2 or HSO_3 as exploited in the Raschig process.[24] Alternatively, hydroxylamine can form by reduction of nitric acid or by hydrogenation of NO.[25,26]

When the 3-aminoisoxazole **2** was subsequently reacted with urea **4** (Fig. 2a) a second spot-to-spot conversion to *N*-isoxazolyl-urea **5** (42%) was observed. Again, LC-MS analysis of the reaction confirms clean formation of compound **5** (Fig. 3b). In this reaction, however, we never observed complete conversion, possibly because the liberated ammonia reacts in this transamination reaction with **5** back to **2** and **4**. Importantly, the reaction is free of any side products allowing unreacted starting material to undergo further conversion. We noticed that the transamination reaction to *N*-isoxazolyl-urea **5** could be catalysed by boric acid, which therefore serves two purposes (*vide infra*). Boric acid is a reagent that has also been shown to catalyse formation of ribose **6** and that helps to stabilize ribofuranosides, as discovered by Benner and co-workers.[27,28]

The efficiently formed *N*-isoxazolyl-urea **5** is in our pathway the central starting material for the subsequent ribosylation depicted in Fig. 2a. Compound **5** already possesses all the atoms with the correct connectivity needed to form the pyrimidine base cytosine. Subsequent elaboration to give cytosine requires only reductive cleavage of the labile N-O bond of the isoxazole and formation of a new C-N bond (Fig. 2a and b). The isoxazole moiety consequently solves two chemical problems. First, it ensures that **5** contains only one nucleophilic atom for the subsequent boric acid-catalysed ribosylation, which fully eliminates potential regiochemistry problems. Second, it simultaneously masks an aldehyde and amine functionality (Fig. 2b). The aldehyde is subsequently needed for a cyclization reaction that gives directly the pyrimidine ring of cytidine with concomitant elimination of water.

Using again HPLC analysis, we found that **5** was indeed sufficiently nucleophilic to react efficiently with ribose **6**. When we mixed ribose **6** with *N*-isoxazolyl-urea **5** and heated the mixture to 95°C in the presence of boric acid, we noted a fast and high yielding reaction that provides the corresponding ribosylated products **7a-d** in total riboside yields of 95% (Fig. 3c). This high yield corresponds to almost quantitative, side product free, formation of the ribosides. Careful experimentation with minerals showed that similar high yields are also obtained when we replaced the boric acid by other borate minerals such as lüneburgite or borax. In all cases, the glycosylation yield was usually high with > 70% (Fig. S1).

Due to the slightly acidic conditions of the boric acid catalysed ribosylation, the dominant products of the reaction are the α/β-pyranosides (**7c** and **7d**), which form in a 6:1 ratio together with the α/β-furanosides (**7a** and **7b**, Fig. 3c). We next investigated the pyranoside/furanoside ratio in more detail. When we heated the mixture of **7a-d** under slightly basic conditions in the presence of borates and analysed the reaction mixture by LC-MS, we noted that the furanosides (**7a** and **7b**) became the dominating products (54%) with an α/β-ratio of 3:1 (Fig. 3d). This reaction also results to some extent in hydrolysis, affording once more the *N*-isoxazolyl-urea **5** (31%). However, because **5** is stable under these conditions it is not lost but available for another round of ribosylation. The enrichment of the furanosides can be explained by the stabilizing effect of the borate, which complexes cis-diols.[27] This complexation is more stable for the furanosides, establishing a slight advantage, which leads to faster hydrolysis of the pyranosides. This represents a physicochemical selection step towards the natural occurring furanoside sugar conformation. Importantly, the ribosylation of **5** with **6**, followed by this

isomerization provides the isoxazole ribofuranosides nucleosides **7a** and **b** in excellent yields (54%) in a clean reaction (Fig. 3d).

We next investigated, the reductive cleavage of the isoxazole N-O bond and the subsequent cyclization to cytidine under plausible prebiotic conditions (Fig. 2b). For these experiments we used the furanoside mixture obtained in the ribosylation reaction before, which had a typical α:β ratio of 3:1 (Fig. 3d). Gratifyingly, reductive ring-opening was possible with a variety of reducing agents. Investigation of this step, however, was performed with the concern that the reaction should be compatible with the envisioned parallel formation of the purine nucleosides along the FaPy-pathway, which requires slightly basic conditions.[7,13] To our delight, we found that a Fe^{2+}/thiol system met the criteria (Fig. 2b).[29] The reduction consequently furnishes disulfides as oxidation products. As a Fe^{2+} source we found that the mineral pyrite (FeS_2), which was extensively investigated in the context of prebiotic chemistry by G. Wächtershäuser, gives excellent results.[30-32] LC-MS analysis of the reaction (Fig. 4) showed that heating of a solution of the isoxazole-ribosides in basic carbonate or carbonate/borate buffer in the presence of pyrite powder and thiols provided the expected pyrimidine nucleosides **8a**-**d** in excellent yields. We also noted formation of the uridine nucleoside **9a**-**b** as hydrolysis products. If the reaction mixture was heated at neutral rather than basic pH, only small amounts of starting material converted into the pyrimidine nucleosides. This result suggests that mechanistically the rate-limiting step is the deprotonation of the thiols for complexation with the Fe^{2+} of the pyrite mineral (Fig. 2b). To show that the reduction is not just limited to pyrite, we also tested other Fe^{2+} sources such as water soluble Fe^{2+} salts or FeS. Indeed efficient pyrimidine nucleoside formation was in all cases observed. Remarkably, we found that even 0.0001 eq. of soluble Fe^{2+} is sufficient to achieve the reduction. We next examined different thiol sources and observed that all tested thiols such as DTT, propanedithiol, mercaptoethanol and even the amino acid cysteine gave similar results (Fig. S2). We found that dithiols lead to slightly faster reductions compared with monothiols, probably due to a more efficient complexation with Fe^{2+} (Fig. 2b). In summary, pyrimidine formation by reductive N-O cleavage of isoxazoles is a very robust method that has the potential to simplify even contemporary syntheses of pyrimidine nucleoside pharmaceuticals.

Addition of naturally occurring minerals such as hydroxyapatite, colemanite or lüneburgite to the reduction process influenced the distribution of the four expected cytidine isomers. In our experiments lüneburgite gave the best results with a combined yield of 85% (Fig. 4). Considering that we started with a 3:1 α/β mixture, the reduction step provided the canonical nucleoside cytidine in a remarkable yield (β-f-C, **8b**) of 27%. The α-anomer (α-f-C, **8a**) is formed in 33% yield. The total yield of the β-anomer could potentially be further increased by UV-anomerization that converts the α-anomer into the natural occurring β-anomer as recently reported.[33] Besides the furanoside products, we also detected cytosine (9%), α-pyranosyl-cytidine (a-p-C, **8c**, 5%) and the β-pyranosyl-cytidine (β-p-C, **8d**, 9%), as well as small amounts of uridine α- and β-anomers (**9a** and **b**) with 1.1% and 0.7%, respectively. The pathway we discovered therefore provides uridine as the hydrolysis product of cytidine as well (Fig. 4).

We next investigated the effect of increasing the number of molar equivalents of ribose (Fig. 4), since an excess is used in the ribosylation reaction. Here we discovered that adding more ribose influenced the α:β ratios positively in the direction of the β-furanoside (Fig 4), while the total yields dropped slightly to about 70-80%. The yield of the uridine nucleosides also increased due to hydrolysis. Interestingly, we found that ribose can reduce the N-O bond of the isoxazole moiety under basic conditions in the absence of Fe^{2+} and thiols. This might be the result of a Cannizzaro reaction.[34]

In the next step, we wanted to investigate if the pathway is efficient enough to allow for direct access of the phosphorylated nucleotides in one-pot (Fig. 5a). This seemed plausible because the mineral

lüneburgite in combination with urea, which is also part of our isoxazole pathway, was reported to enable selective phosphorylation of the 5'-position.[35] We therefore added urea to the reaction mixture after pyrimidine formation and allowed the mixture to evaporate to dryness at 85°C over a period of about 20 h. Indeed, LC-MS analysis of the reaction showed that nucleotides are obtained in about 20% yield relative to cytidine (Fig. 5b, Fig. S3). In addition to cytidine-5'-mono-phosphate **10a**/**b** we also found di-phosphorylated cytidine **11a**/**b** with a ratio of around 3:1. The 5'-mono- and 5'-di-phosphorylated uridine nucleotides **12a**/**b** and **13a**/**b** were also detected. To confirm nucleotide formation under these conditions, we isolated the corresponding peaks from the prebiotic reaction and enzymatically removed the phosphate groups. LC-MS analysis of the resulting sample confirmed the presence of the corresponding nucleosides, whereas the nucleotides were now absent (Fig. 5b). Closer analysis of the data showed that the phosphorylation reaction led to a dramatic decrease of the pyranosides **14** in the nucleotide pool. We observed almost exclusive formation of the furanosidic phosphorylation products, which dominated under these conditions with over 94% yield relative to the pyranosides. This corresponds to an increase of the furanoside/pyranoside ratio from initially 4:1 to 17:1. We found no discrimination between α- and β-anomers after phosphorylation (Fig 5b). The furanoside selectivity is certainly caused by the presence of a primary hydroxyl group, which is absent in the pyranosides. Selective phosphorylation of the 5'-OH group, shifts the equilibrium towards the furanosides. The *in situ* phosphorylation therefore represents an additional chemical selection step that leads to the selective emergence of the furanoside 5'-nucleotides, which are the sole isomeric constituents found in RNA and DNA today.

To elucidate the correct structures of the phosphorylated nucleotide products we used LC-MS analysis with the help of an enzymatic digest, in combination with co-injection studies (Fig. 5c). To confirm formation of 5'-pyrophosphates we used the enzyme apyrase, which selectively degrades 5'-tri- and 5'-di-phosphates to the 5'-mono-phosphates. Indeed, when we treated the reaction mixture with apyrase, the signals for the 5'-diphosphates disappeared almost completely, while the mono-phosphates stayed intact. This confirmed the formation of 5'-pyrophosphates. We observed that two peaks for the cytidine and uridine derivatives fully disappeared, which is caused by the α- and β-anomers of 5'-cytidine-di-phosphate (**11a** and **b,** α-/β-CDP) and 5'-uridine-di-phosphate (**13a** and **b**, α-/β-CDP). With the help of additional co-injection studies, we confirmed formation of β-CMP **10b**, β-CDP **11b**, β-UMP **12b** and β-UDP **13b** as shown in Fig. 5c and Fig. S4. For CMP and UMP the UV-signals represent a mixture of the α- and β-anomers **10a**/**b** and **12a**/**b**, since these signals were not separable (Fig. 5c).

The final goal was now to show that the new pyrimidine pathway was compatible with the chemistry needed to form the canonical purine RNA building blocks. Again, cyanoacetylene **1** can serve as a starting point towards the purine nucleosides. In contrast to pyrimidine formation, **1** can be reacted with ammonia instead of hydroxylamine (Fig. 6a). This gives 3-aminoacrylonitrile **15**. Since ammonia and hydroxylamine are both products of the Raschig process, the formation of pyrimidines and purines are directly linked. Heating 3-aminoacrylonitrile **15** releases HCN to form acetonitrile **16**, which reacts to malononitrile **17** with cyanogen **18**, as described by Eschenmoser and co-workers (Fig. 6a).[36] Malononitrile **17** is a well-established starting material for the prebiotic formation of purine nucleosides via formamidopyrimidine intermediates (FaPys, **19**).[13] We therefore ribosylated FaPyG **19a** and FaPyA **19b**, the precursors to guanosine and adenosine, under the exact same conditions needed for the pyrimidine pathway (5 eq. ribose, 0.25 eq. boric acid, Fig. 6a). The FaPy-ribosides **20a**/**b** are subsequently heated in the same buffer that also leads to the formation of the pyrimidine nucleosides (100 mM carbonate, 50 mM borate, pH 9.7, 0.0005 eq. Fe^{2+}, 1.5 eq. DTT). Analysing the reactions showed that all nucleosides indeed formed in yields between 10% and 90% depending on the nucleobase precursor (*N*-isoxazolyl-urea: 90%, FaPyA: 60%, FaPyG: 10%). The experiment provides

consequently all four canonical nucleosides at the same time under the same conditions and in excellent yields (Fig. 6b, in red and blue): cytidine (C, **8b**, 24%), uridine (U, **9b**, 1%), guanosine (G, **21**, 4%) and adenosine (A, **22**, 17%). Since uridine **9b** forms directly from cytidine by spontaneous hydrolysis in water, the yield increases with time. The data shows that all canonical pyrimidine and purine nucleosides (A, G, C, U) needed to form Watson-Crick base pairs can form in the same chemical environment (Fig. 6b).

Discussion

Today, it is believed that life began with the chemical formation of RNA that established an RNA world in which RNA strands replicated and evolved to give increasingly complex chemical systems. Whether RNA was directly assembled from the canonical nucleotides (A, C, G and U) or if it evolved from a simpler genetic system is unclear.[37] It is plausible that different (non) Watson-Crick base pairing systems might have played a role before contemporary RNA finally evolved.[38,39] The chemical pathway described here, however, suggest that all four RNA building blocks could have been directly part of the prebiotic nucleoside/tide pool from which RNA at some point evolved. This does not answer the question of whether RNA was directly assembled from its components or if it evolved from a different molecular system, but the simultaneous formation of all RNA building blocks provides a plausible pathway towards the evolution of RNA on early Earth.

In our prebiotic scenario, the molecule cyanoacetylene **1**, which is frequently detected throughout the universe, serves as the key starting material.[16-18] Cyanoacetylene **1** is a strong electrophile, which reacts quickly with nucleophiles such as water or amines. As such **1** would have been short lived in an aqueous environment. We believe its reactivity would need to be exploited early in any plausible reaction scheme to avoid potential availability problems.[8] Therefore, our pathways both use cyanoacetylene **1** as a starting point allowing it to react with small nucleophiles such as hydroxylamine. This gives 3-aminoisoxazole **2** in excellent yields as an intermediate towards pyrimidine nucleosides. In contrast, the reaction of **1** with ammonia, another excellent nucleophile, finally provides malononitrile **17**, which is an ideal starting material for purine nucleosides along the FaPy-pathway.[13] The two nucleophiles ammonia and hydroxylamine are chemically closely related, because both are produced simultaneously by the reduction of nitrite or nitrate.[40] The pyrimidine and purine reaction pathways are also in other aspects closely correlated. Formation of **5**, as the central precursor for the pyrimidines, for example, requires urea, which is the hydrolysis product of guanidine.[41] Guanidine belongs to the class of amidines, which are used to produce the central purine nucleobase precursor **19**.[13] Importantly, reaction of the central pyrimidine and purine precursors **5** and **19** with ribose proceeds under identical conditions. These conditions furnish the corresponding ribosides (**7a-d** and **20a**/**b**) in excellent yields with excellent regioselectivity (Fig. 6a). All the ribosides finally react in the same geochemical environment to give the purine and pyrimidine nucleosides under slightly basic conditions. We chose carbonate to guide the reactions, which may be the most likely base available on the early Earth if we consider that the atmosphere of the early Earth was mainly composed of N_2, CO_2 and H_2O.[42,43] Our reaction pathways avoid the need for other anions such as free phosphates, which are often used in high concentrations to establish general acid base catalysis.[8,44] Phosphates, however, generate low solubility ionic compounds, particularly in the presence of divalent cations such as the highly abundant Fe^{2+} or Ca^{2+}. This limits their availability for general acid base catalysis on early Earth.[45-48] The formation of nucleotides, however, was shown to be possible in the absence of free phosphates.[35,48,49] In our scenario we exploit the mineral lüneburgite for the phosphorylation reactions. Lüneburgite is a magnesium borophosphate, first investigated for its prebiotic potential by Benner and co-workers.[35] Indeed, we found that Lüneburgite is fully compatible with our chemical

scenario allowing for the formation of the naturally occurring 5'-nucleotides as mono- and di-phosphates with excellent specificity. We discovered that this phosphorylation reaction is highly selective for the furanoside isomers in comparison to the pyranosides. This provides an explanation to the question of why nature constructed DNA and RNA from the thermodynamically disfavoured 5-membered furanosides.[15] Importantly, 5'-di-phosphates are also formed as activated nucleotides. Their formation is desirable because 5'-phosphorylation is necessary for non-enzymatic copying of RNA.[50-52] All this together now provides a plausible prebiotic scenario of how the Watson-Crick base pairing nucleobases were created on the early earth as the central starting point for the emergence of life.

Acknowledgement

We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support via the programs SFB1032 (TP-A5), SFB749 (TP-A4), SPP-1784, CA275/11-1. Additional funding from the Excellence Cluster EXC117 CiPS^{M} is acknowledged. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912). H.O. received a Marie Sklodowska-Curie Individual Fellowship (Project; PRENUCRNA) from the European Commission. We further thank Jonathan Kampmann for XRD measurements.

Literature

1 Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147-157 (1982).
2 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
3 Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E. & Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife* **7**, e35255 (2018).
4 Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **113**, 9786-9791 (2016).
5 Lincoln, T. A. & Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science* **323**, 1229-1232 (2009).
6 Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618 (1986).
7 Becker, S. *et al.* A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833-836 (2016).
8 Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239 (2009).
9 Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. *J. Am. Chem. Soc.* **132**, 16677-16688 (2010).
10 Stairs, S. *et al.* Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Comm.* **8**, 15270 (2017).
11 Kim, H.-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci. USA* **114**, 11315-11320 (2017).
12 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).
13 Becker, S. *et al.* Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Comm.* **9**, 163 (2018).

14 Stefan, P., Sebastian, W., Bernhard, J. & Albert, E. Why Pentose- and Not Hexose-Nucleic Acids??. Part VII. Pyranosyl-RNA ('p-RNA'). Preliminary communication. *Helv. Chim. Acta* **76**, 2161-2183 (1993).
15 Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure. *Science* **284**, 2118-2124 (1999).
16 Niemann, H. B. *et al.* The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779 (2005).
17 Walmsley, C. M., Winnewisser, G. & Toelle, F. Cyanoacetylene and cyanodiacetylene in interstellar clouds. *Astron. Astrophys.* **81**, 245-250 (1980).
18 Thaddeus, P. The prebiotic molecules observed in the interstellar gas. *Philos. Trans. Royal Soc. B* **361**, 1681-1687 (2006).
19 Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Cyanoacetylene in Prebiotic Synthesis. *Science* **154**, 784-785 (1966).
20 Saladino, R., Botta, G., Pino, S., Costanzo, G. & Di Mauro, E. Genetics first or metabolism first? The formamide clue. *Chem. Soc. Rev.* **41**, 5526-5565 (2012).
21 Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: III. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol.* **33**, 693-704 (1968).
22 Dresler, W. F. & Stein, R. Ueber den Hydroxylharnstoff. *Liebigs Ann.* **150**, 242-252 (1869).
23 Kofod, H. On the Isomerism of Hydroxyurea. I. Kinetics of the Reaction between Hydroxylammonium Ion and Cyanate Ion. *Acta Chem. Scand.* **7**, 274-279 (1953).
24 Ritz, J. F., H.; Perryman, H. Hydroxylamine in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH Verlag, 2000).
25 Julius, T. Die elektrolytische Reduktion der Salpetersäure bei Gegenwart von Salzsäure oder Schwefelsäure. *Z. Anorg. Allg. Chem.* **31**, 289-325 (1902).
26 Fedoseev, G. *et al.* Efficient surface formation route of interstellar hydroxylamine through NO hydrogenation. II. The multilayer regime in interstellar relevant ices. *J. Chem. Phys.* **137**, 054714 (2012).
27 Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate Minerals Stabilize Ribose. *Science* **303**, 196-196 (2004).
28 Kim, H.-J. *et al.* Synthesis of Carbohydrates in Mineral-Guided Prebiotic Cycles. *J. Am. Chem. Soc.* **133**, 9457-9468 (2011).
29 Kijima, M., Nambu, Y. & Endo, T. Reduction of cyclic compounds having nitrogen-oxygen linkage by dihydrolipoamide-iron(II). *J. Org. Chem.* **50**, 1140-1142 (1985).
30 Wächtershäuser, G. Pyrite Formation, the First Energy Source for Life: a Hypothesis. *Syst. Appl. Microbiol.* **10**, 207-210 (1988).
31 Wächtershäuser, G. Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* **52**, 452-484 (1988).
32 Wächtershäuser, G. Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. USA* **87**, 200-204 (1990).
33 Fernandez-Garcia, C., Grefenstette, N. M. & Powner, M. W. Selective aqueous acetylation controls the photoanomerization of alpha-cytidine-5'-phosphate. *Chem. Comm.* **54**, 4850-4853 (2018).
34 Cannizzaro, S. Ueber den der Benzoësäure entsprechenden Alkohol. *Liebigs Ann.* **88**, 129-130 (1853).
35 Hyo-Joong, K. *et al.* Evaporite Borate-Containing Mineral Ensembles Make Phosphate Available and Regiospecifically Phosphorylate Ribonucleosides: Borate as a Multifaceted Problem Solver in Prebiotic Chemistry. *Angew. Chem. Int. Ed.* **55**, 15816-15820 (2016).
36 Trinks, U. & Eschenmoser, A. *Zur Chemie der Aminopyrimidine*, ETH Zurich, (1987).
37 Krishnamurthy, R. On the Emergence of RNA. *Isr. J. Chem.* **55**, 837-850 (2015).
38 Krishnamurthy, R. Role of pK(a) of Nucleobases in the Origins of Chemical Evolution. *Acc. Chem. Res.* **45**, 2035-2044 (2012).
39 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. Comm.* **7**, 11328 (2016).

40 Taniguchi, I., Nakashima, N., Matsushita, K. & Yasukouchi, K. Electrocatalytic reduction of nitrate and nitrite to hydroxylamine and ammonia using metal cyclams. *J. Electroanal. Chem.* **224**, 199-209 (1987).
41 Bell, J. CLVII.-The hydrolysis of guanidine. *J. Chem. Soc.* **129**, 1213-1219 (1926).
42 Trail, D., Watson, E. B. & Tailby, N. D. The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature* **480**, 79 (2011).
43 Kasting, J. Earth's early atmosphere. *Science* **259**, 920-926 (1993).
44 Fernández-García, C., Coggins, A. J. & Powner, M. W. A Chemist's Perspective on the Role of Phosphorus at the Origins of Life. *Life* **7**, 31 (2017).
45 Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc. Natl. Acad. Sci. USA* **110**, 10089-10094 (2013).
46 Keefe, A. D. & Miller, S. L. Are polyphosphates or phosphate esters prebiotic reagents? *J. Mol. Evol.* **41**, 693-702 (1995).
47 Oelkers, E. & Valsami-Jones, E. Phosphate Mineral Reactivity and Global Sustainability. *Elements* **4**, 83-87 (2008).
48 Burcar, B. *et al.* Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. *Angew. Chem. Int. Ed.* **55**, 13249-13253 (2016).
49 Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 212 (2017).
50 Li, L. *et al.* Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* **139**, 1810-1813 (2017).
51 Prywes, N., Blain, J. C., Del Frate, F. & Szostak, J. W. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife* **5**, e17756 (2016).
52 Zhang, W., Walton, T., Li, L. & Szostak, J. W. Crystallographic observation of nonenzymatic RNA primer extension. *eLife* **7**, e36422 (2018).

Figures

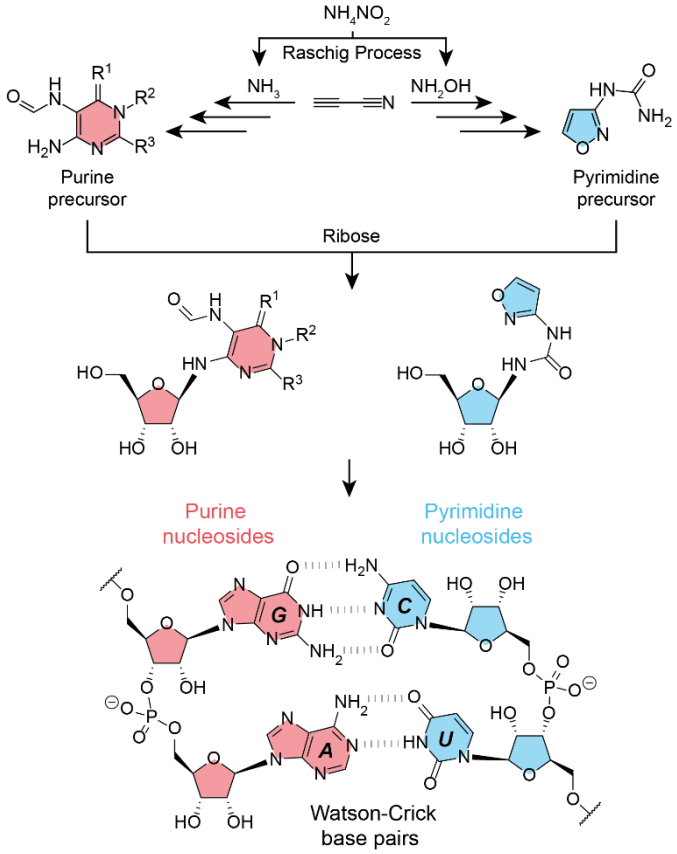


Fig. 1. Unifying scheme for the synthesis of all four RNA nucleosides A, C, G and U, required for the formation of the two Watson Crick base pairs G:C and A:U. Starting from cyanoacetylene, reaction with related nucleophiles, ammonia or hydroxylamine, is needed for the simultaneous formation of purines (A and G) or pyrimidines (U and C) respectively.

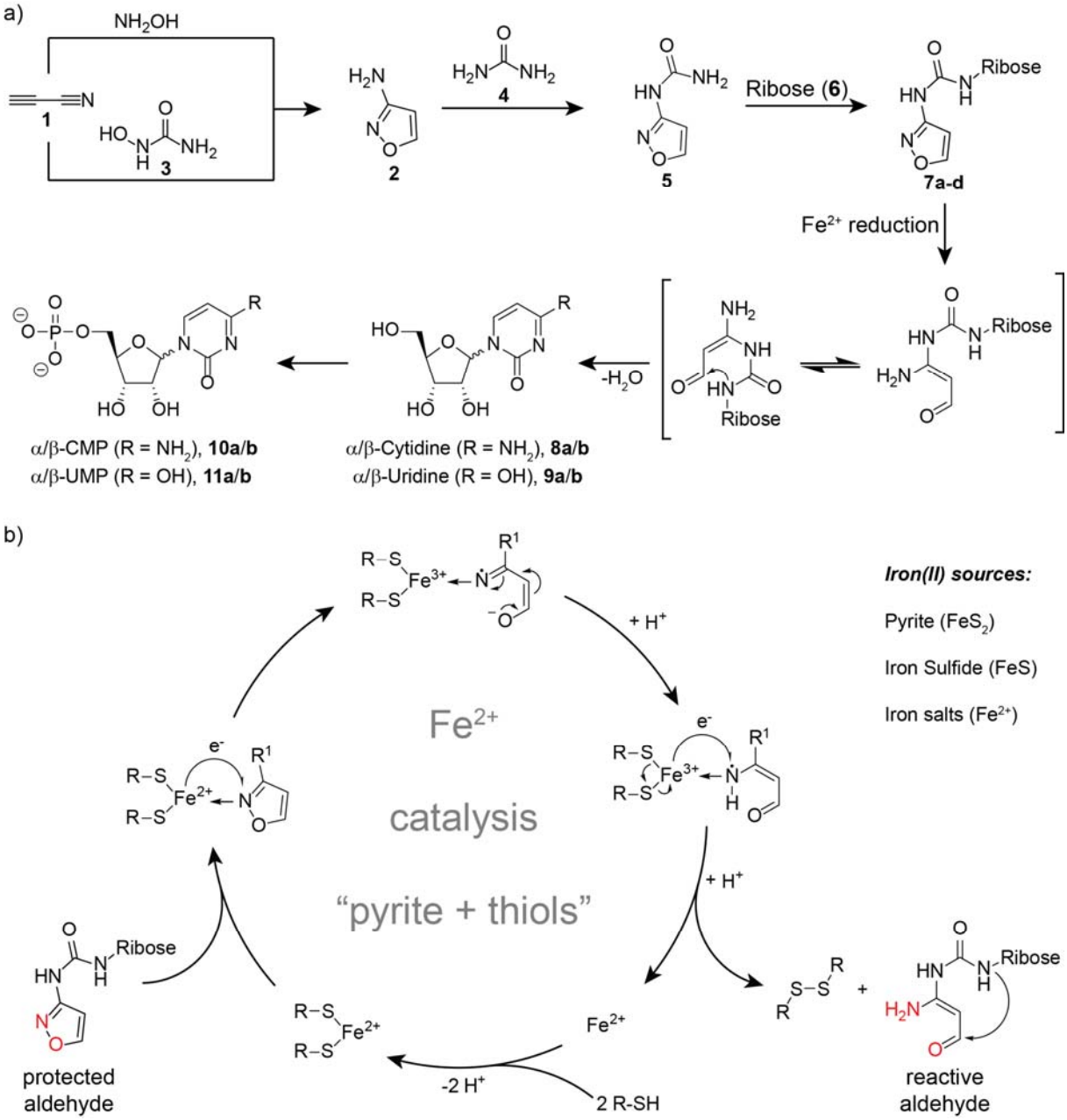


Fig. 2. Pathway to pyrimidine nucleosides and nucleotides. a) The chemical pathway starts with cyanoacetylene **1**, which first reacts to 3-aminoisoxazole **2** and further to *N*-isoxazolyl-urea **5**. The subsequent ribosylation gives *N*-isoxazolyl-urea-riboside **7a-d**. Upon reduction cytidine and uridine are obtained that can be phosphorylated in one-pot to give the corresponding nucleotides **8a**/**b**-**11a**/**b**. b) Proposed catalytic cycle for the Fe^{2+} catalysed reduction of the isoxazole moiety.

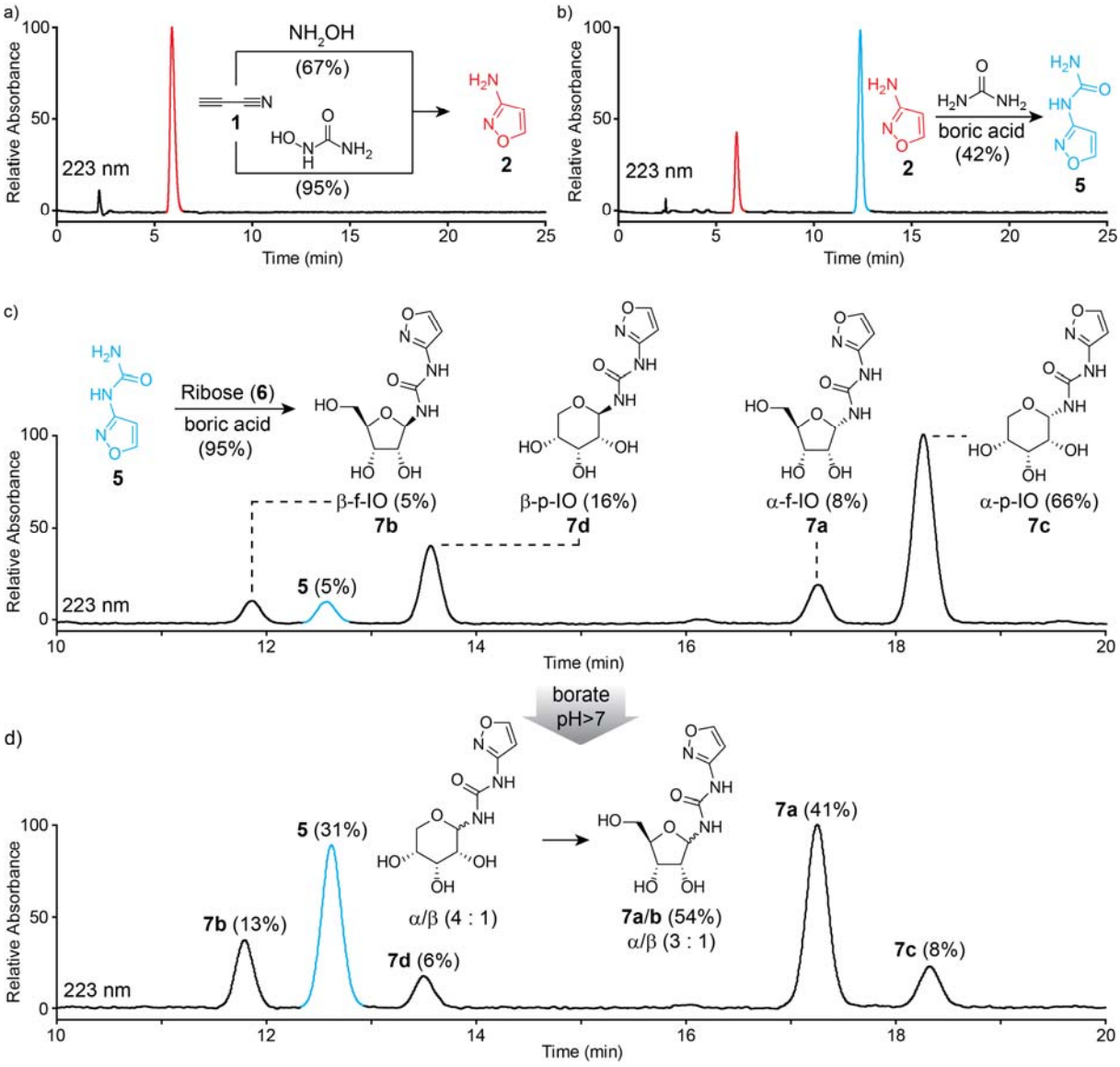


Fig. 3. LC-MS analysis for the formation of **2**, **5** and **7a-d**. Only the UV-chromatograms at 223 nm are shown. a) reaction of cyanoacetylene with hydroxylamine or hydroxylurea, affords 3-aminoisoxazole **2** (in red). b) Compound **3** (in red) is converted by a boric acid catalysed transamination reaction to *N*-isoxazolyl-urea **5** (in blue). c) Boric acid catalysed ribosylation of **6** affords the four expected α and β-anomers of isoxazolyl-urea (IO) as furanoside (f) or pyranoside (p) **7a-d**. d) The mixture obtained in c) was heated in 125 mM borax at 100 °C for 2.5 h. Isomerization is also possible at lower temperature but with longer reaction time. The pyranoside products (**7c** and **d**) are converted into the furanoside products (**7a** and **b**).

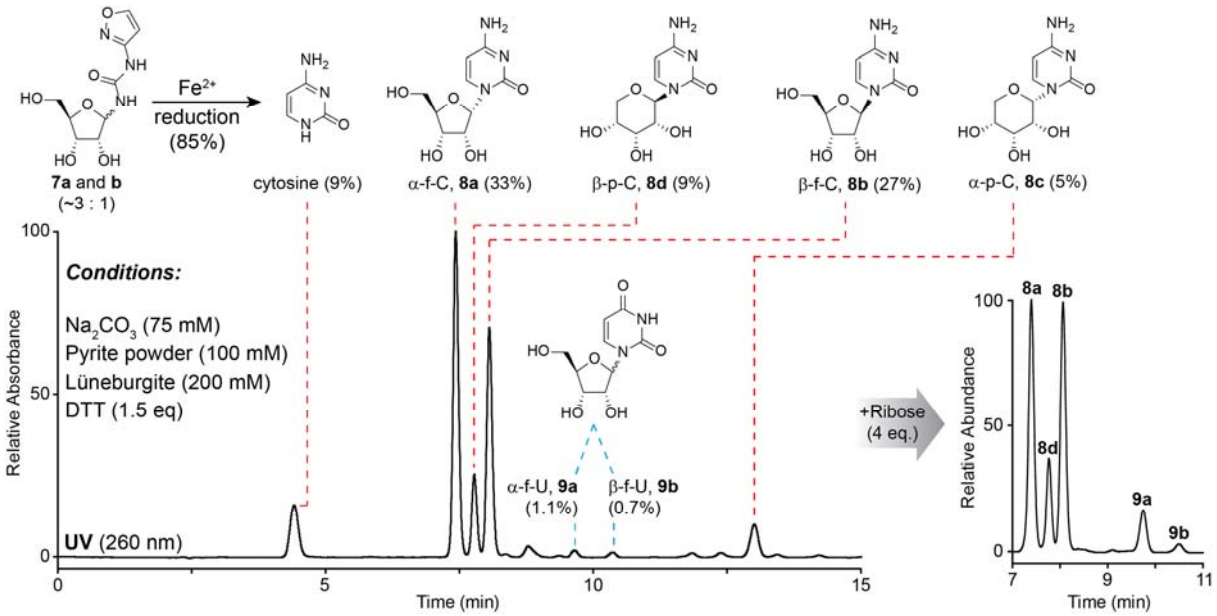


Fig. 4. Formation of pyrimidine nucleosides. LC-MS analysis of pyrimidine nucleoside formation from compound **7a**/**b**. The UV-chromatogram at 260 nm is shown. The reduction of the isoxazole moiety in **7a**/**b** furnished the α and β-anomers as pyranoside (p) and furanside (f) isomers of cytidine (C) or uridine (U) (**8a-d**, **9a**/**b**). If the same reaction is performed with additional ribose (right) the isomer distribution changes in favour of the β-furanoside. At the same time hydrolysis to uridine is also favoured, resulting in increased yields.

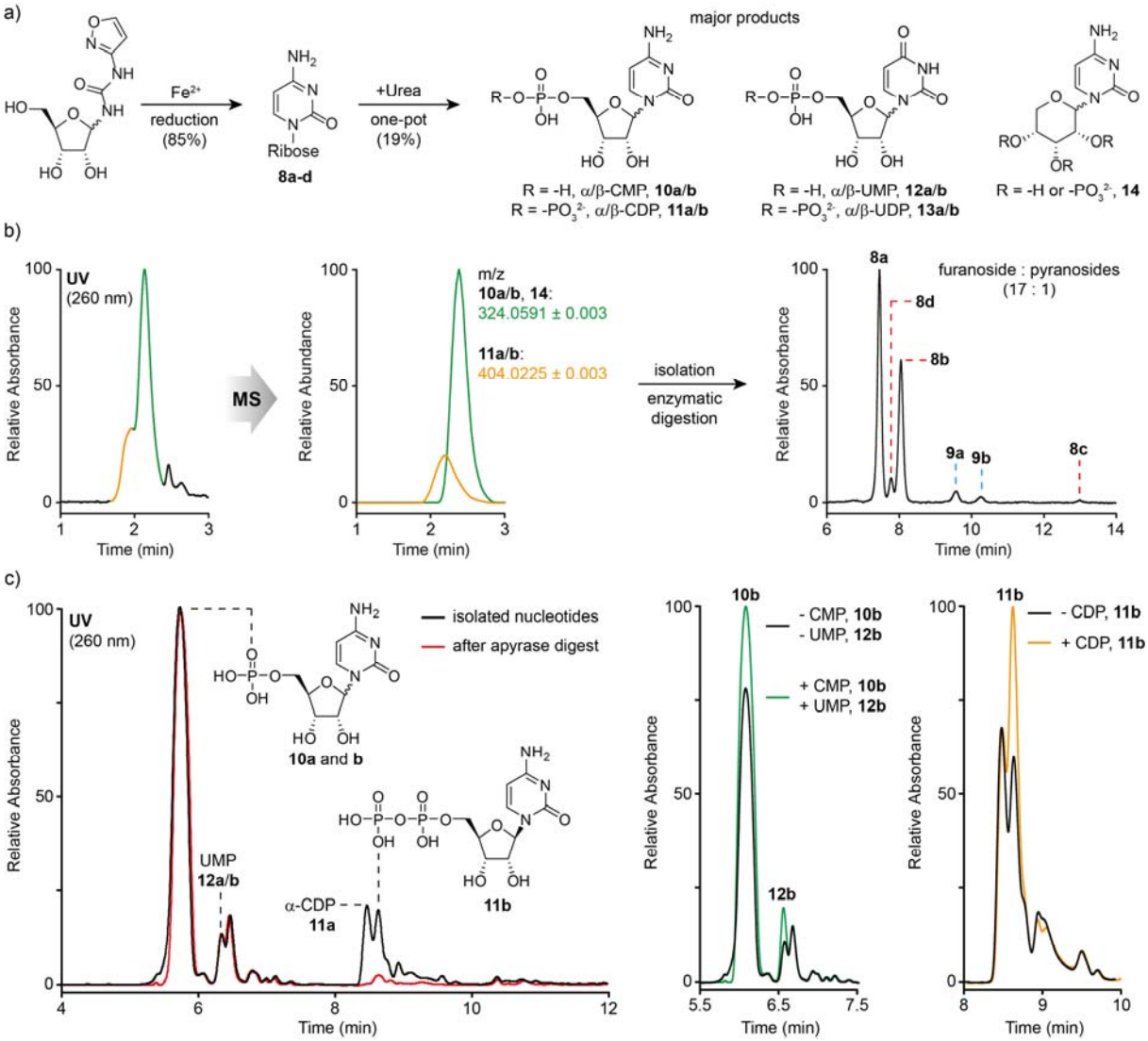


Fig. 5. One-pot nucleotide formation. a) One-pot Synthesis of cytidine and uridine mono- and di-phosphates (**10a**/**b**-**13a**/**b**) after urea was added to the reaction mixture shown in Fig. 4 and letting the mixture dry-down at 85°C for 20 h. b) LC-MS analysis of the corresponding nucleotide peaks and isolation from the prebiotic reaction, followed by an enzymatic removal of the phosphate groups. This confirmed formation of phosphorylated cytidine and uridine derivatives. Phosphorylation of pyranosyl-isomers is just a minor side reaction. c) Chromatographic separation of the different phosphorylated species and confirmation of 5'-pyrophosphates by apyrase digest (left). Co-injection studies confirmed formation of CMP (**10a**/**b**), CDP (**11b**), UMP (**12a**/**b**) and UDP (**13b**). The α- and β-anomers of CMP und UMP were not separable and therefore the UV peaks represent a mixture of both anomers.

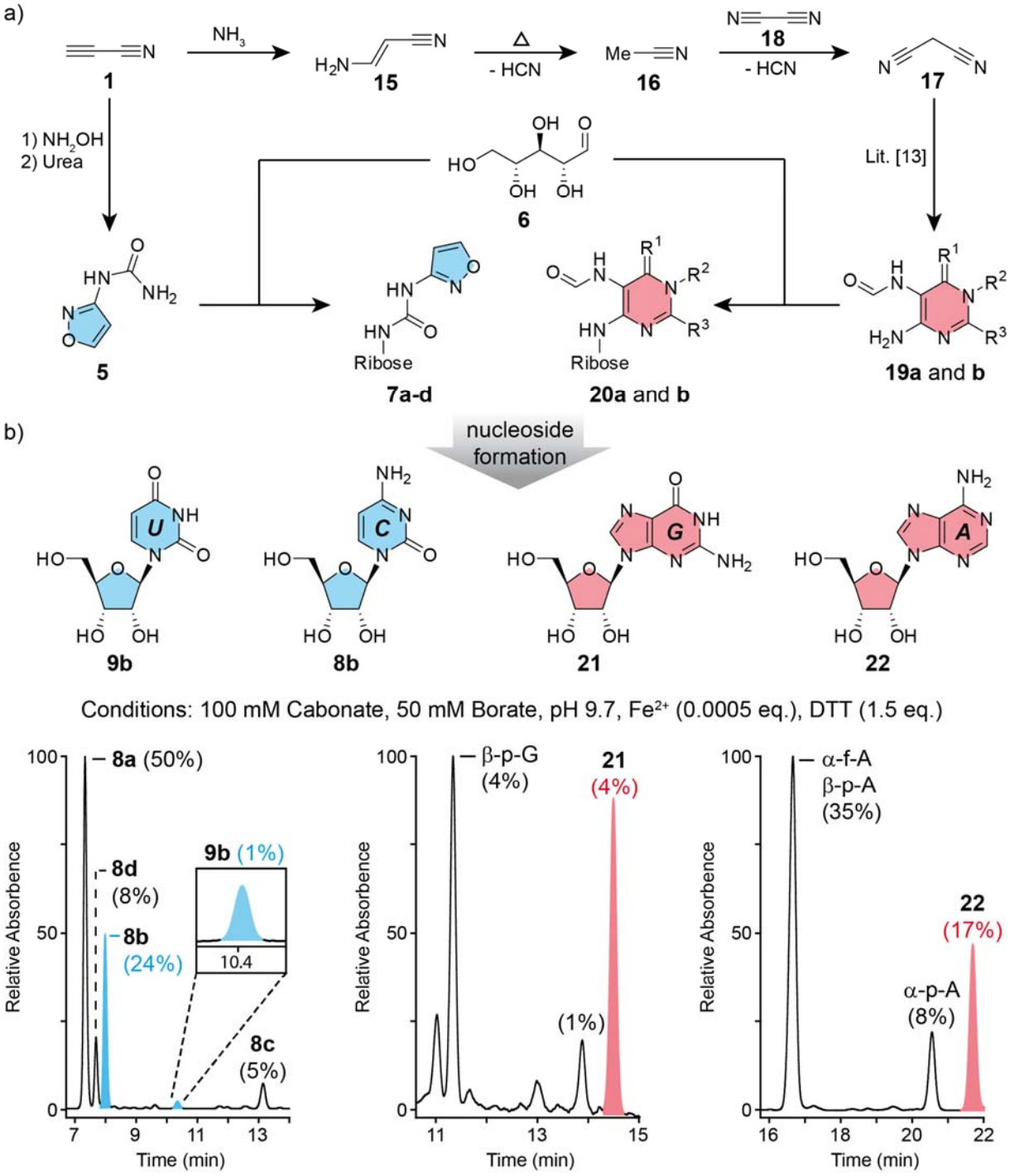


Fig. 6. Parallel formation of pyrimidine and purine nucleosides. a) Chemical pathway for the formation of purine and pyrimidine precursors, both starting from cyanoacetylene **1**. The pyrimidine precursors (blue, **5** and **7a-d**) and purine precursors (red, **19a**/**b** and **20a**/**b**) are highlighted. The compounds containing formamidopyrimidines (FaPys) contain the following substitution patterns; **19a**/**20a** (FaPyG): R^1 = O, R^2 = H, R^3 = NH_2; **19b**/**20b** (FaPyA): R^1 = NH, R^2 = H, R^3 = H. b) Formation of C (**8b**), U (**9b**), G (**21**), and A (**22**) under the same conditions. The UV-chromatograms are shown at 260 nm. The canonical pyrimidine (blue) and purine (red) building blocks are highlighted.

Supplementary Information

Materials and Methods

General Information

Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR, Carbosynth, TCI or Acros organics and used without further purification. Solutions were concentrated *in vacuo* on a Heidolph rotary evaporator. The solvents were of reagent grade or purified by distillation. Chromatographic purification of products was accomplished using flash column chromatography on Merck Geduran Si 60 (40-63 µm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on Merck 60 (silica gel F254) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or standard staining solutions. ^{1}H- and ^{13}C-NMR spectra were recorded in deuterated solvents on Varian Oxford 200, Bruker ARX 300, Varian VXR400S, Varian Inova 400, Bruker AMX 600 and Bruker AVIIIHD 400 spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. High-resolution ESI spectra were obtained on the mass spectrometers Thermo Finnigan LTQ FT-ICR. IR measurements were performed on Perkin Elmer Spectrum BX FT-IR spectrometer with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were measured on a Büchi B-540 device. For preparative HPLC purification a Waters 1525 binary HPLC Pump in combination with a Waters 2487 Dual Absorbance Detector was used, with a Nucleosil 100-7 C18 reversed phase column. The prebiotic reactions were analyzed by LC-ESI-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system. All chromatographic separations except for nucleotides were performed on an Interchim Uptisphere120 3HDO C18 column with a flow of 0.15 ml/min and a constant column temperature of 30 °C. Eluting buffers were buffer A (2 mM $HCOONH_4$ in H_2O (pH 5.5)) and buffer B (2 mM $HCOONH_4$ in H2O/MeCN 20/80 (pH 5.5)). The gradient for isoxazole containing compounds and pyrimidine nucleosides was 0 → 20 min, 0% → 4% buffer B. The gradient for purine nucleosides was 0 → 45 min, 0% → 15% buffer B. Chromatographic separation for nucleotides were performed on a YMC-Triart C18 column with a flow of 0.20 ml/min and a constant column temperature of 40 °C. Eluting buffers were buffer A (10 mM NH_4HCO_3 and 5 mM dibutylamine in H_2O (pH 9.1)) and buffer B (MeCN). The gradient was 0 → 10 min, 0% → 20%; 10 → 20 min, 20 → 20% buffer B. The elution was monitored at 223 nm and 260 nm (Dionex Ultimate 3000 Diode Array Detector). The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 80-500 with a resolution of 30000. Nucleotides were scanned by use of a negative polarity mode over a full-scan range of m/z 120-1000 with a resolution of 30000. Parameters of the mass spectrometer were tuned with a freshly mixed aqueous solution of adenosine (5 µM, positive mode) or cytidine-mono-phosphate (10 µM, negative mode). The synthetic standards for the co-injection experiments were synthesized in our lab (see synthetic procedures or according to reported literature[1]) or purchased from Sigma-Aldrich, Fluka, ABCR, Carbosynth, TCI or Acros organics. XRD measurements were performed on a STOE powder diffractometer in transmission geometry (Cu-Kα1, λ = 1.5406 Å) with a step size of 2°2θ (30 seconds per step) and equipped with a position-sensitive Mythen-1K detector.

Degradase digestion

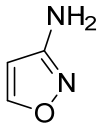
Isolated nucleotides in 42 µL H_2O were digested as follows: 10X DNA Degradase™ Reaction Buffer (5 µL), together with 8 U DNA Degradase Plus™ (2 µL, Zymo Research) and 3.5 U of Benzonase® Nuclease (1 µL, Merck, purity >90%) was added and the mixture was incubated at 37 °C for 2 h. The sample was directly analysed by LC-MS according to the general information.

Apyrase digestion

Isolated nucleotides in 17.6 µL H_2O were digested as follows: 10X Apyrase reaction buffer (2 µL, NEB) was added together with 0.2 U of apyrase (0.4 µL, NEB). The mixture was incubated at 30 °C for 2 h. The sample was directly analysed by LC-MS according to the general information.

Synthetic Procedures

3-aminoisoxazole (2)



Reaction with hydroxyurea:

To the ice-cold solution of *N*-hydroxyurea (61 mg, 0.8 mmol) and sodium hydroxide (53 mg, 1.3 mmol) in 2 ml of H_2O was added cyanoacetylene (31 µl, 0.5 mmol), and the reaction mixture was stirred overnight at room temperature. The yield was determined by LC-MS measurement with the calibration curve prepared using commercially available 3-aminoisoxazole (Yield: 95%). For the confirmation of the structural integrity, the reaction mixture was extracted with diethylether, dried over $MgSO_4$, concentrated and the crude residue purified by flash column chromatography (CH_2Cl_2:MeOH = 20:1). The NMR spectrum of the isolated compound was identical to that of the reference compound.

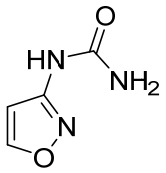
Reaction with hydroxylamine:

To the ice-cold solution of hydroxylamine hydrochloride (112 mg, 1.6 mmol) and sodium hydroxide (106 mg, 2.6 mmol) in 4 ml of H_2O was added cyanoacetylene (31 µl, 0.5 mmol), and the reaction mixture was stirred overnight at room temperature. The yield was determined by LC-MS measurement with the calibration curve prepared using commercially available 3-aminoisoxazole (Yield: 67%). For the confirmation of the structural integrity, the reaction mixture was extracted with diethylether, dried over $MgSO_4$, concentrated and the crude residue purified by flash column chromatography (CH_2Cl_2:MeOH = 20:1). The NMR spectrum of the isolated compound was identical to that of the reference compound.

The analytical data is in agreement with a commercial sample.

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.32 (d, *J* = 1.7 Hz, 1H, HC5), 5.88 (d, *J* = 1.7 Hz, 1H, HC4), 5.57 (s, 2H, NH_2). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 163.76 (C5), 158.73 (C3), 97.71 (C4).

***N*-isoxazolyl-urea (5)**



Prebiotic Synthesis:

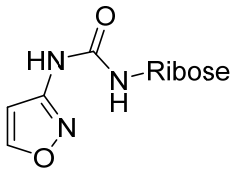
To a mixture of urea (60 mg, 1.0 mmol) and boric acid (3 mg, 0.05 mmol) was added 3-aminoisoxazole (21 mg, 0.25 mmol) and H_2O (5 µl, 0.25 mmol). The vial was sealed tightly and heated at 150°C for 12 h. The reaction mixture was dissolved in 1 ml of H_2O, and the yield was determined by LC-MS measurement with the calibration curve prepared using synthetically prepared reference compound (Yield: 42%). For the confirmation of the structure, the reaction mixture was extracted with CH_2Cl_2, dried over $MgSO_4$, concentrated and the crude residue purified by flash column chromatography (CH_2Cl_2:MeOH = 20:1 → 10:1). The NMR spectrum of the isolated compound was identical to that of the reference compound.

Synthetic reference:

3-amino-isoxazole (1.14 g, 1.00 ml, 13.5 mmol) was dissolved in dry THF (20 ml) under inert atmosphere at 0°C. 2,2,2-trichloroacetylisocyanate (2.45 g, 1.61 ml, 13.5 mmol) was slowly added to the solution and the reaction was stirred at rt for 2 h. The reaction was quenched with MeOH (10 ml) and the solvent removed *in vacuo*. After co-evaporation with EtOH (2 x 20 ml) the product was obtained as a colorless solid (3.44 g, 12.6 mmol, 93 %). The crude product (3.12 g, 11.4 mmol) was dissolved in methanolic ammonia (10 mL, 7 M) and stirred for 1.5 h at rt. MeOH (20 mL) und EtOH (20 mL) were added to the reaction to obtain a clear solution. Et_2O (40ml) was added to the clear solution to precipitate the product. The crude product was filtered off and was subsequently dissolved in H_2O (20 ml). After filtering through celite, the solvent was removed by freeze-drying to obtain the product as a colorless solid (1.16 g, 9.15 mmol, 80%).

mp: 175 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 9.41 (s, 1H, HN4), 8.65 (d, ^3J = 1.7 Hz, 1H, HC5), 6.70 (d, ^3J = 1.7 Hz, 1H, HC4), 6.28 (s, 2H, NH_2).**^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 159.4 (C5), 158.6 (C3), 154.6 (CO), 98.2 (C4). **HRMS** (ESI+): calc.: $[C_4H_6N_3O_2]^+$ 128.0455, found: 128.0455 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3486 (vw), 3389 (w), 3259 (vw), 3179 (w), 3194 (w), 3033 (w), 2958 (w), 1752 (m), 1730 (m), 1686 (w), 1631 (w), 1592 (vs), 1537 (w), 1486 (m), 1416 (s), 1378 (s), 1343 (w), 1298 (vw), 1284 (w), 1123 (m), 1070 (w), 1039 (vw), 981 (vw), 928 (w), 891 (w), 797 (s), 756 (vs), 654 (w).

***N*-isoxazolyl-*N'*-ribosyl-urea (7a-d)**



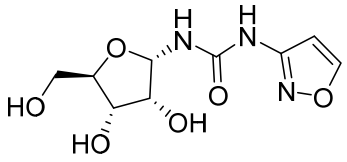
Prebiotic synthesis:

N-isoxazolylurea (6.4 mg, 0.05 mmol) was thoroughly ground up with ribose (37.5 mg, 0.25 mmol, 5 eq.) and boric acid (0.8 mg, 0.013 mmol, 0.25 eq.). The mixture was heated overnight in an oven at 95 °C. Alternatively the reaction can also be performed by a dry-down method, where a solution of *N*-isoxazolylurea (500 ul, 0.05 mmol, 100 mM) is mixed with a ribose (83 µl, 0.25 mmol, 3 M) and boric acid (25 ul, 0.013 mmol, 500 mM) solution. The mixture was kept in an oven for 20 h at 95 °C. The sample was dissolved in H_2O (2 ml) and analyzed by LC-MS. To confirm the structural integrity, the different isomers were isolated by reversed phase HPLC in pure form.

Synthetic reference:

1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (17.0 g, 33.8 mmol) was dissolved in dry DCM (300 ml) under inert atmosphere. $TiCl_4$ (4.46 mL, 40.7 mmol) was added and stirred for 2 h at rt. H_2O (250 ml) was added and filtered through celite. The organic layer was separated and the aqueous phase was extracted with DCM (3 x 100 ml). The combined organic layers were washed with sat. NaCl (300 ml), dried over $MgSO_4$, filtered and the solvent was removed *in vacuo*. The crude product was dissolved in dry toluene (300 ml) and AgNCO (6.25 g, 41.6 mmol) was added. After refluxing for 2.5 h, the reaction mixture was filtered through celite and washed with dry toluene. To the clear solution was added 3-aminoisoxazole (3 ml, 40.6 mmol) and stirred for 16 h at rt. The solvent was removed *in vacuo* and the product purified by flash column chromatography (DCM:MeOH 100:1). The product was dissolved in methanolic ammonia (330 ml, 7 M) and stirred for 18 h at rt. The solvent was removed in vacuo and the product purified by flash column chromatography (DCM:MeOH 17:3 → 4:1). The product was obtained as a colorless foam and contained a 3:1 mixture of the α- and β-furanoside (4.87 g, 18.8 mmol, 56%). The a- and b-furanosides were isolated in pure form by reversed phase HPLC to confirm the structural integrity.

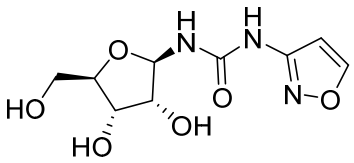
α-furanosyl-isomer **7a**



mp: (131 °C).**^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 9.85 (s, 1H, NH), 8.66 (d, ^3J = 1.7 Hz, 1H, HC5), 7.24 (d, ^3J = 9.5 Hz, 1H, C1'NH), 6.73 (d, ^3J = 1.8 Hz, 1H, HC4), 5.50 (dd, ^3J = 9.5, 4.4 Hz, 1H, HC1'), 5.41 (br, 1H, C3'OH), 5.03 (br, 1H, C2'OH), 4.66 (br, 1H, C5'OH), 3.92 (t, ^3J = 4.4 Hz, 1H, HC2'), 3.87 (dd, ^3J = 6.6, 4.5 Hz, 1H, HC3'), 3.71 (ddd, J = 6.5, 4.9, 3.1 Hz, 1H, HC4'), 3.49 (dd, ^2J = 11.7 Hz, ^3J = 3.1 Hz, 1H, H_aC5'), 3.39-3.35 (m, 1H, H_bC5'). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 159.5 (C5), 158.4 (C3), 153.3 (CO), 98.3 (C4), 81.9 (C4'), 80.6 (C1'), 71.1 (C3'), 70.1 (C2'), 61.6 (C5'). **HRMS** (ESI+): calc.: $[C_9H_{14}N_3O_6]^+$

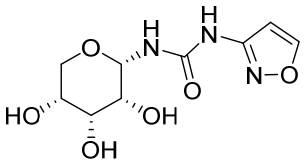
260.0877, found: 260.0877 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3852 (w), 3820 (w), 3648 (w), 3283 (w)m 1843 (vw), 1771 (w), 1717 (m), 1699 (s), 1652 (vs), 1634 (s), 1575 (m), 1558 (vs), 1539 (vs), 1506 (vs), 1456 (s), 1436 (m), 1418 (m), 1032 (m), 971 (m).

β-furanosyl-isomer **7b**



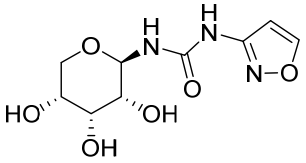
mp: 120 °C (decomp.).**^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 9.63 (s, 1H, NH), 8.67 (d, ^3J = 1.7 Hz, 1H, HC5), 7.13 (d, ^3J = 9.5 Hz, 1H, C1'NH), 6.74 (d, ^3J = 1.7 Hz, 1H, HC4), 5.20 (dd, ^3J = 9.5, 5.4 Hz, 1H, HC1'), 5.12 – 4.68 (m, 3H, C2'OH, C3'OH and C5'OH), 3.85 (t, ^3J = 4.6 Hz, 1H, HC3'), 3.71 (t, ^3J = 5.4 Hz, 1H, HC2'), 3.67 (q, ^3J = 4.2 Hz, 1H, HC4'), 3.45 (dd, ^2J = 11.7 Hz, ^3J = 4.0 Hz, 1H, H_aC5'), 3.42 – 3.39 (m, 1H, H_bC5').**^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 159.6 (C5), 158.3 (C3), 153.4 (CO), 98.3 (C4), 84.4 (C1'), 83.5 (C4'), 74.3 (C2'), 70.3 (C3'), 61.8 (C5'). **HRMS** (ESI+): calc.: $[C_9H_{13}NaN_3O_6]^+$ 282.0697, found: 282.0696 $[M+Na]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3268 (m), 1677 (m), 1599 (s), 1537 (vs), 1478 (s), 1403 (m), 1228 (m), 999 (vs), 893 (s), 777 (vs).

α-pyranosyl-isomer (**7c**)



mp: 190 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 9.97 (s, 1H, NH), 8.67 (d, ^3J = 1.8 Hz, 1H, HC5), 7.47 (d, ^3J = 9.1 Hz, 1H, C1'NH), 6.74 (d, ^3J = 1.8 Hz, 1H, HC4), 5.25 – 4.87 (m, 4H, HC1', C2'OH, C3'OH, C5'OH), 3.70 (t, ^3J = 2.8 Hz, 1H, HC3'), 3.64 – 3.44 (m, 4H, HC2', HC4', H_aC5', H_bC5'). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 159.5 (C5), 158.4 (C3), 153.4 (CO), 98.4 (C4), 77.9 (C1'), 69.4 (C3'), 69.3 (C4'), 68.4 (C5') 67.6 (C2'). **HRMS** (ESI+): calc.: $[C_9H_{13}N_3O_6]^+$ 260.0877, found: 260.0877 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3090 (vw), 2363 (m), 2271 (w), 1956 (w), 1700 (vs), 1683 (s), 1652 (s), 1823 (w), 1732 (w), 1700 (vs), 1683 (s), 1652 (s), 1609 (w), 1559 (m), 1509 (m), 1456 (s), 707 (w), 680 (m), 656 (s).

β-pyranosyl-isomer (**7d**)



mp: 175 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 9.46 (s, 1H, NH), 8.68 (d, ^3J = 1.8 Hz, 1H, HC5), 6.89 (d, ^3J = 9.1 Hz, 1H, C1'NH), 6.74 (d, ^3J = 1.8 Hz, 1H, HC4), 4.90 (t, ^3J = 9.1 Hz, 1H, HC1'), 4.84 (d, ^3J = 3.6 Hz, 1H, C3'OH), 4.82 (d, J = 7.4 Hz, 1H, C2'OH), 4.69 (d, ^3J = 6.5 Hz, 1H, C4'OH), 3.88 (q, J = 2.8 Hz, 1H, HC3'), 3.55 – 3.47 (m, 1H, HC4'), 3.47 – 3.37 (m, 2H, H_aC5, H_bC5'), 3.18 (ddd, J = 9.1, 7.4, 2.8 Hz, 1H, HC2'). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 159.7 (C5), 158.3 (C3), 153.6 (CO), 98.3 (C4),

77.5 (C1‘) 70.9 (C3‘), 69.9 (C2‘), 67.0 (C4‘), 64.1 (C5‘). **HRMS** (ESI+): calc.: $[C_9H_{13}N_3O_6]^+$ 260.0877, found: 260.0877 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3314 (m), 2871 (vw), 1941 (vw), 1705 (vs), 1598 (w), 1540 (vs), 1476 (w), 1432 (vw), 1336 (vw), 1282 (m), 1266 (w), 1208 (w), 1168 (m), 1139 (m), 1088 (w), 1065(m), 1035 (vs), 1004 (m), 950 (s), 916 (w), 886 (m), 804 (m), 789 (vs), 744 (w), 706 (m).

Prebiotic formation of pyrimidine nucleosides and nucleotides

General procedure nucleoside formation

The reactions were handled under inert atmosphere. All solutions were degassed for 1 h with argon before use.

A solution of ribose isoxazole **7a**/**b** as a 3:1 mixture (25.9 mg, 0.1 mmol, 1 eq.) in 75 mM Na_2CO_3 (0.5 ml) was added to a mixture of mineral (1 eq.), dithiol (1.5 eq.) and Fe^{2+} source (0.5 eq.) in a 15 ml falcon tube. The tube was sealed with a PTFE sealing tape and shaked for 5 h at 100 °C in an Eppendorf ThermoMixer®. After cooling to rt, it was centrifuged and a sample (10 µl) was removed and diluted with H_2O to 1 ml. This diluted sample was used for LC-MS analysis according to the general information.

The different minerals, thiols and iron sources used are stated in the main text. Monothiols were used with 3 eq. As water soluble Fe^{2+} salt we used ammonium iron(II) sulfate hexahydrate in a range of 0.001-0.0001 eq. Reactions without minerals were performed in a different buffer (100 mM Na_2CO_3, 50 mM boric acid, pH 9.7).

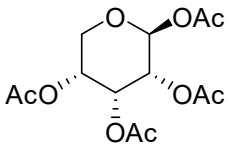
One-pot nucleotide formation

The above described procedure was used for nucleoside formation. We used lüneburgite (50 mg, 0.1 mmol, 1 eq.), DTT (23.1 mg, 0.15 mmol, 1.5 eq.) and FeS_2 (6 mg, 0.05 mmol, 0.5 eq.) as mineral, thiol and Fe^{2+} source, respectively. After nucleoside formation we added solid urea (240 mg, 4 mmol, 40 eq.) and heated to 65 °C until the urea was fully dissolved. The sample was well suspended by vortexing and a sample (25 µl) was transferred into a 2 ml Eppendorf tube. The sample was kept at 85 °C for 20 h in an Eppendorf ThermoMixer® open to the air to allow water to evaporate. The dried sample was taken up in H_2O (1 ml) and analyzed for nucleotides by LC-MS analysis according to the general information.

Formation of purine nucleosides compatible to pyrimidine nucleoside formation

For purine nucleoside synthesis, FaPyA or FaPyG (0.05 mmol, 1eq.) was thoroughly ground up with ribose (37.5 mg, 0.25 mmol, 5 eq.) and boric acid (0.8 mg, 0.013 mmol, 0.25 eq.), identical to the N-isoxazolyl-urea ribosylation. The mixture was heated overnight in an oven at 95 °C. The resulting ribosides were heated at 100 °C in sealed 2 ml Eppendorf tubes under the following conditions: 100 mM Na_2CO_3, 50 mM boric acid, pH 9.7, 1.5 eq. DTT and 0.0005 eq. ammonium iron(II) sulfate hexahydrate at 100 °C. The following reaction times and concentrations were used: FaPyA (3d, 25 mM), FaPyG (2d, 6.25 mM).

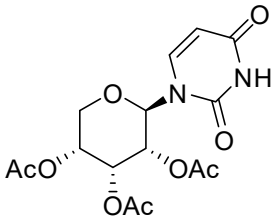
(1R,2R,3R,4S)-tetrahydro-2H-pyran-2,3,4,5-tetrayl tetraacetate (23)



D-ribose (5.00 g, 33.3 mmol) was dissolved in pyridine (20 mL) and cooled to 0°C in an ice water bath. Acetic anhydride (20 ml, 212 mmol) was added and stirred at 0°C overnight. The mixture was poured into ice cold water (200 ml). The solid was filtered off and the crude product was crystallized from a mixture of MeOH and H_2O to obtain the pure product as colorless crystals (5.22 g, 16.4 mmol, 49%).

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm) = 6.02 (d, ^3J = 4.8 Hz, 1H, HC1), 5.47 (t, ^3J = 3.4 Hz, 1H, HC3), 5.19 – 5.11 (m, 1H, HC4), 5.05 – 5.00 (m, 1H, HC2), 4.01 (dd, ^2J = 12.4 Hz, ^3J = 3.4 Hz, 1H, H_aC5), 3.90 (dd, ^2J = 12.4 Hz, ^3J = 5.7 Hz, 1H, H_bC5), 2.12 (s, 3H, CH_3), 2.09 (s, 3H, CH_3), 2.09 (s, 3H, CH_3), 2.08 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm) = 170.0 (CO), 169.9 (CO), 169.6 (CO), 168.9 (CO), 91.0 (C1), 67.40 (C2), 66.3 (C3), 66.2 (C4), 62.8 (C5), 21.0 (CH_3), 20.9 (CH_3), 20.8 (CH_3), 20.8 (CH_3).

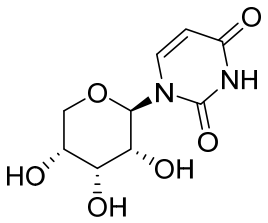
(1'R,2'R,3'R,4'S)-2-(2,4-Dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydro-2H-pyran-3,4,5-triyl triacetate (24)



Uracil (269 mg, 2.40 mmol) was suspended in dry MeCN (14 ml) under inert atmosphere. The mixture was refluxed and bis(trimethylsilyl)acetamid (1.26 mL, 1.05 mg, 5.15 mmol) was added. After the mixture became clear, a solution containing tetraacetyl-β-D-ribose **23** (636 mg, 2.00 mmol, 0.8 eq.) and TMSOTf (0.54 mL ,667 mg, 3.00 mmol) in dry MeCN was added. The reaction was further refluxed for 8 h. After cooling to rt, the reaction was quenched with sat. $NaHCO_3$ (20 ml) and the aqueous phase extracted with EtOAc (5 x 25 ml). The combined organic layers were dried over MgSO4, filtered and the solvent removed *in vacuo*. The crude product was purified by flash column chromatography (DCM/MeOH 97:3) to obtain the pure compound as a yellowish solid (655 mg, 1.77 mmol, 89%).

mp: 164 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ(ppm) = 11.47 (s, 1H, HN3), 7.90 (d, ^3J = 8.2 Hz, 1H, HC6), 5.85 (d, ^3J = 9.7 Hz, 1H, HC1'), 5.70 (dd, ^3J = 8.1 Hz, ^4J = 1.5 Hz, 1H, HC5), 5.66 – 5.63 (m, 1H, HC3'), 5.49 (dd, ^3J = 9.8, 3.1 Hz, 1H, HC2'), 5.23 – 5.15 (m, 1H, HC4'), 3.99 – 3.92 (m, 1H, H_aC5'), 3.85 (dd, ^2J = 11.0 Hz, ^3J = 5.5 Hz, 1H, H_bC5'), 2.19 (s, 3H, CH_3), 1.98 (s, 3H, CH_3), 1.92 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, DMSO-d_6) δ(ppm) = 170.0 (CO), 169.3 (CO), 169.0 (CO), 162.8 (C4), 150.6 (C2), 140.7 (C6), 102.7 (C5), 77.7 (C1'), 67.7 (C3'), 66.5 (C2'), 65.5(C4'), 62.7 (C5'), 20.5 (CH_3), 20.5 (CH_3), 20.3 (CH_3). **HRMS** (ESI+): calc.: $[C_{15}H_{19}N_2O_9]^+$ 371.1090, found: 371.1085 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3066 (vw), 2822 (vw), 2360 (vw), 1747 (s), 1692 (s), 1453 (w), 1371 (s), 1305 (w), 1212 (vs), 1163 (m), 1121 (w), 1079 (s), 1041 (s), 986 (m), 952 (m), 882 (w), 813 (m), 771 (w), 718 (w).

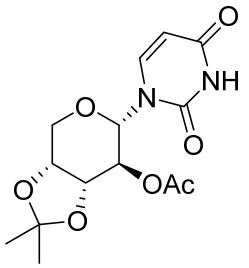
β-D-Ribopyranosyl-uracil (9d)



Nucleoside **24** (56 mg, 0.151 mmol) was dissolved in methanolic ammonia (4 ml, 1 M) and stirred for 19 h at rt. The solvent was removed *in vacuo* and the crude product was purified by flash column chromatography (DCM/MeOH 4:1) to obtain the pure product as a colorless solid (28.0 mg, 0.115 mmol, 76%). For LC-MS measurements, a fraction of the product was further purified by reversed phase HPLC.

mp: 220 °C (decomp.). **^{1}H NMR** (599 MHz, DMSO-d_6) δ (ppm) = 11.25 (s, 1H, HN3), 7.66 (d, ^3J = 8.1 Hz, 1H, HC6), 5.60 (d, ^3J = 8.1 Hz, 1H, HC5), 5.58 (d, ^3J = 9.4 Hz, 1H, HC1'), 5.11 (br, 1H, OH), 5.09 (br, 1H, OH), 4.84 (br, 1H, OH), 3.97 (d, ^3J = 3.2 Hz, 1H, HC3'), 3.68 (d, J = 9.5 Hz, 1H, HC2'), 3.63 (ddd, ^3J = 7.4, 6.0, 2.3 Hz, 1H, HC4'), 3.58 – 3.53 (m, 2H, H_aC5', H_bC5'). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 163.0 (C4), 151.1 (C2), 141.4 (C6), 101.67 (C5), 79.6 (C1'), 71.2 (C3'), 67.6 (C2'), 66.4 (C4'), 65.2 (C5'). **HRMS** (ESI+): calc.: $[C_9H_{13}N_2O_6]^+$ 245.0768, found: 245.0768 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3347 (w), 2357 (vw), 2336 (vw), 1675 (vs), 1363 (w), 1391 (w), 1272 (m), 1250 (m), 1197 (m), 1084 (vs), 1042 (vs), 975 (m), 918 (w), 859 (w), 813 (m), 779 (w), 689 (m), 667 (m).

(1'R,2'S,3'S,4'R)-6-(2,4-Dioxo-3,4-dihydropyrimidin-1(2H)-yl)-2,2-dimethyltetrahydro-4H-[1,3]dioxolo[4,5-c]pyran-7-yl acetate (25)

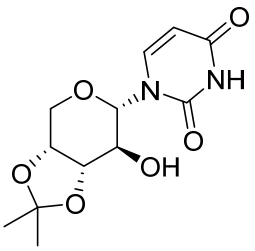


Uracil (224 mg, 2.00 mmol, 1.00 eq.) was suspended in dry MeCN (16 ml) under inert atmosphere. Bis(trimethylsilyl)acetamid (1.26 ml, 1.05 g, 5.00 mmol, 2.50 eq.) was added and stirred at 40 °C until the mixture became clear. TMSOTf (0.50 ml, 0.62 mg, 2.80 mmol, 1.40 eq.) was added. 1,2-O-acetate-3,4-O-isopropylidene-D-arabinose1 (716 mg, 2.60 mmol, 1.30 eq.) was separately dissolved in dry MeCN (5 ml) and slowly added to the reaction mixture within 30 min. After stirring 2 h at 40 °C, the reaction was quenched with sat. $NaHCO_3$ (20 ml) and extracted with DCM (3 x 25 ml). The combined organic layers were washed with sat. NaCl (30 ml), dried over $Na2SO_4$, filtered and the solvent removed *in vacuo*. The product was purified by flash column chromatography (iHex/acetone, 7:3 → 1:1) to yield the product as a colorless solid (565 mg, 1.73 mmol, 87%).

mp: 225 °C; **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm) = 8.54 (s, 1H, HN3), 7.43 (d, ^3J = 8.2 Hz, 1H, HC6), 5.78 (dd, ^3J = 8.2 Hz, ^4J 2.3 Hz, 1H, HC5), 5.56 (d, ^3J = 9.4 Hz, 1H, HC1'), 5.10 (dd, ^3J = 9.4, 7.0 Hz, 1H, HC3'), 4.40 (d, ^2J = 13.9 Hz, 1H, H_aC5'), 4.34 (dd, J = 7.1, 5.4 Hz, 1H,HC2'), 4.31 – 4.26 (m, 1H, HC4'), 3.98 (dd, ^2J = 13.9, 2.4 Hz, 1H, H_b5'), 2.06 (s, 3H, CH_3), 1.59 (s, 3H, CH_3), 1.39 (s, 3H, CH_3). **^{13}C NMR** (101 MHz,

$CDCl_3$) δ (ppm) = 170.0 (CO), 162.5 (C4), 150.4 (C2), 139.9 (C6), 110.8 (C_q), 103.3 (C5), 80.4 (C1'), 76.5 (C2'), 73.3 (C4'), 71.1 (C3'), 66.3 (C5'), 27.9 (CH_3), 26.2 (CH_3), 20.8 (CH_3). **HRMS** (ESI+): calc.: $[C_{14}H_{19}N_2O_7]^+$ 327.1187, found: 327.1188 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 2985 (vw), 2882 (vw), 2359 (vw), 2053 (vw), 1741 (m), 1711 (s), 1675 (vs), 1470 (w), 1426 (w), 1375 (m), 1296 (m), 1249 (m), 1218 (vs), 1170 (w), 1130 (s), 1051 (vs), 967 (w), 889 (w), 847 (m), 797 (m), 761 (w), 686 (w).

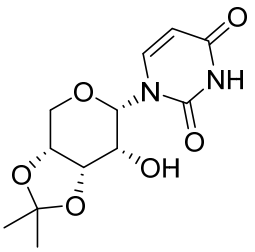
1-((1'R,2'S,3'S,4'R)-7-Hydroxy-2,2-dimethyltetrahydro-4H-[1,3]dioxolo[4,5-c]pyran-6-yl)pyrimidine-2,4(1H,3H)-dion (26)



Nucleoside **25** (355 mg, 1.09 mmol) was dissolved in methanolic ammonia (14 ml, 1 M) and kept at rt for 20 h. The solvent was removed *in vacuo* and the crude product was purified by flash column chromatography (DCM/MeOH, 92:8) to obtain the product as colorless solid (284 mg, 1.00 mmol, 92%).

mp: 220 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 11.36 (s, 1H, HN3), 7.65 (d, ^3J = 8.1 Hz, 1H, HC6), 5.64-5.59 (m, 2H, HC5, HO), 5.26 (d, ^3J = 9.7 Hz, 1H, HC1'), 4.23 – 4.18 (m, 1H, HC4'), 4.15 (d, ^2J = 13.5 Hz, 1H, H_aC5'), 4.11 (dd, ^3J = 7.0, 5.5 Hz, 1H, HC3'), 3.92 (dd, ^2J = 13.5 Hz, ^3J = 2.6 Hz, 1H, H_bC5'), 3.71 – 3.63 (m, 1H, HC2'), 1.49 (s, 4H, CH_3), 1.29 (s, 4H, CH_3). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 163.0 (C4), 151.0 (C2), 141.3 (C6), 108.7 (C_q), 102.0 (C5), 81.9 (C1'), 79.1 (C3'), 73.3 (C4'), 69.7 (C2'), 65.3 (C5'), 28.0 (CH_3), 26.2 (CH_3). **HRMS** (ESI+): calc.: $[C_{12}H_{17}N_2O_6]^+$ 285.1081, found: 285.1082 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3378 (w), 2989 (w), 2888 (vw), 2827 (vw), 2360 (vw), 1678 (vs), 1623 (m), 1469 (w), 1416 (w), 1392 (m), 1372 (m), 1336 (w), 1288 (m), 1275 (w), 1249 (s), 1216 (s), 1202 (s), 1167 (w), 1138 (s), 1108 (m), 1086 (vs), 1046 (s), 1022 (m), 975 (m), 958 (m, 936 (w), 874 (s), 846 (s), 826 (m), 796 (m), 780 (w), 761 (s), 732 (m), 684 (m).

1-((1'R,2'R,3'S,4'R)-7-Hydroxy-2,2-dimethyltetrahydro-4H-[1,3]dioxolo[4,5-c]pyran-6-yl)pyrimidine-2,4(1H,3H)-dion (27)

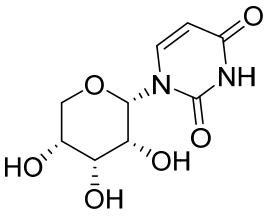


Nucleoside **26** (150 mg, 0.53 mmol) was dissolved in DCM (8 ml) together with Dess-Martin Periodinan (336 mg, 0.79 mmol) and $NaHCO_3$ (89 mg, 1.06 mmol). The reaction mixture was reacted for 2 h at 45 °C and quenched with solid $Na_2S_2O_3$ (0.55 g) together with sat. $NaHCO_3$ (5 ml). It was rigorously stirred until the organic phase became clear. The organic layer was separated and the aqueous layer

was extracted with DCM (5 x 15 ml) and EtOAc (5 x 20 ml). The combined organic layers were dried over $MgSO_4$, filtered and the solvent removed *in vacuo*. The crude product was dissolved in a mixture of DCM/EtOAc/MeOH (2:1:1, 10 mL) and $NaBH_4$ (35 mg, 0.93 mmol) was added. The mixture was stirred for 1 h at 0 °C. The solvent was removed *in vacuo*. The residue was dissolved in EtOAc (20 ml) and washed with sat. NaCl (30 ml). The aqueous phase was extracted with EtOAc (8 x 25 ml). The combined organic layers were dried over $MgSO_4$, filtered and the solvent removed *in vacuo*. The crude product was purified by flash column chromatography (DCM/MeOH, 96:4) to obtain the product as a colorless solid (60 mg, 0.21 mmol, 40%).

mp: 205 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 11.32 (s, 1H, HN3), 7.79 (d, ^3J = 8.1 Hz, 1H, HC6), 5.64 (d, ^3J = 3.0 Hz, 1H, HC1'), 5.56 (d, ^3J = 8.1 Hz, 1H, HC5), 5.40 (d, ^3J = 5.9 Hz, 1H, OH), 4.34 (dd, ^3J = 6.7, 4.5 Hz, 1H, HC3'), 4.24 – 4.19 (m, 1H, HC4'), 4.06 (dd, ^2J = 12.6 Hz, ^3J = 3.1 Hz, 1H, H_aC5'), 3.95 (dd, ^2J = 12.7 Hz, ^3J = 3.9 Hz, 1H, H_bC5'), 3.91 – 3.87 (m, 1H, HC2'), 1.46 (s, 3H, CH_3), 1.30 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 163.2 (C4), 150.3 (C2), 143.2 (C6), 109.1 (C_q), 100.0 (C5), 80.2 (C1'), 72.8 (C3'), 70.8 (C4'), 65.4 (C5'), 64.60 (C2'), 26.0 (CH_3), 25.5. (CH_3). **HRMS** (ESI+): calc.: $[C_{12}H_{17}N_2O_6]^+$ 285.1081, found: 285.1081 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 2989 (vw), 2890 (vw), 2359 (vw), 2215 (vw), 1708 (s), 1674 (vs), 1454 (w), 1425 (w), 1381 (m), 1211 (s), 1155 (m), 1126 (s), 1084 (s), 1064 (vs), 1042 (vs), 991 (w), 882 (w), 846 (m), 814 (m), 763 (m), 737 (m), 678 (w).

α-D-Ribopyranosyl-Uracil (α-p-U, 9c)

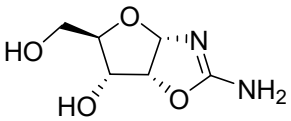


Nucleoside **27** was dissolved in 70% acetic acid (5 ml) and stirred for 5 h at 60 °C. The solvent was removed *in vacuo* and remaining AcOH was co-evaporated with EtOH (2 x 30 ml) and MeCN (3 x 30 ml). The crude product was purified by flash column chromatography (DCM/MeOH 88:12) to obtain the product as white solid (32.0 mg, 0.151 mmol, 83%). For LC-MS measurements, a fraction of the product was further purified by reversed phase HPLC.

mp: 243 °C (decomp.) **^{1}H NMR** (599 MHz, DMSO-d_6) δ (ppm) = 11.34 (s, 1H, HN3), 7.70 (d, ^3J = 8.2 Hz, 1H, HC6), 5.57 (d, ^3J = 8.1 Hz, 1H, HC5), 5.47 (d, ^3J = 1.2 Hz, 1H, HC1'), 5.27 (br, 1H, C2'OH), 5.17 – 5.12 (br, 1H, C3'OH), 5.10 (br, 1H, C4'OH), 3.96 (dd, ^2J = 12.4 Hz, ^3J = 1.6 Hz, 1H, H_aC5'), 3.75 (d, ^2J = 12.2 Hz, 1H, H_bC5'), 3.72 (d, ^3J = 7.4 Hz, 1H, HC2'), 3.70 – 3.66 (m, 2H, HC3', HC4'). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 163.1 (C4), 150.0 (C2), 142.5 (C6), 100.1 (C5), 81.6 (C1'), 70.9 (C2'), 70.3 (C5'), 68.6 (C4'), 67.2 (C3'). **HRMS** (ESI+): calc.: $[C_9H_{13}N_2O_6]^+$ 245.0768, found: 245.0768 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ =3359 (w), 3295 (w), 2934 (vw), 1769 (vw), 1716 (w), 1692 (m), 1665 (s), 1461 (m), 1432 (w), 1398 (w), 1385 (w), 1373 (w), 1322 (vw), 1295 (m), 1251 (m), 1195 (w), 1159 (w), 1110 (m), 1084 (vs), 996 (w), 968 (w), 911 (w), 889 (w), 834 (w), 818 (s), 769 (vs), 714 (s), 670 (s).

Synthesis pathway for α-furanosyl-uridine

(1'S,2'R,3'S,4'R)-2-amino-5-(hydroxymethyl)-3a,5,6,6a-tetrahydrofuro[2,3-d]oxazol-6-ol (28)

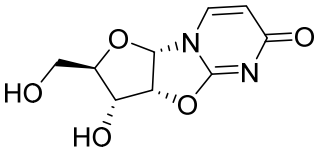


D-Ribose (14.4 g, 96 mmol) and cyanamide (8.06 g, 192 mmol) were dissolved in aqueous ammonia (16 ml, 1M) in a pressure tube. The mixture was heated at 30 °C until the solution became clear. The reaction was further stirred at rt until a precipitate started to form. The reaction was heated for 30 min at 60 °C and the solvent removed *in vacuo*. The residue was taken up in hot MeOH (32 ml) and filtrated immediately. The warm filtrate was kept at -20 °C overnight. The yellowish solid was filtered off and washed with MeOH and Et_2O to obtain the pure product as a colorless solid (12.6 g, 72.3 mmol, 75%).[2]

The analytical data is in agreement with reported literature.[3]

^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm) = 6.14 (d, ^3J = 1.8 Hz, 2H), 5.57 – 5.12 (m, 1H), 4.29 (dd, ^3J = 18.4, 1.9 Hz, 2H), 2.00 – 1.78 (m, 2H), 1.74 – 1.66 (m, 1H), 1.51 – 1.29 (m, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 164.0, 98.1, 80.9, 77.8, 71.1, 60.4.

(1'S,2'R,3'S,4'R)-3-Hydroxy-2-(hydroxymethyl)-2,3,3a,9a-tetrahydro-6H-furo[2',3':4,5]oxazolo[3,2-a]pyrimidin-6-on (29)

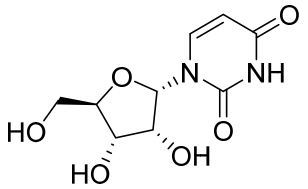


Compound **28** (10.0 g, 75.4 mmol) and methyl propiolate (9.60 g, 9.65 mL, 114 mmol) were dissolved in H_2O (130 mL) and stirred at 100 °C for 30 min. The solvent was removed *in vacuo* and the residue taken up in hot MeOH (65 ml). The solution was filtered hot and the filtrate kept at -20 °C for 48 h. The formed yellow crystals were filtered off to obtain the pure product (6.21 g, 27.5 mmol, 48%).[2]

The analytical data is in agreement with reported literature.[4]

^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm) = 7.85 (d, ^3J = 7.4 Hz, 1H), 6.20 (d, ^3J = 5.3 Hz, 1H), 5.88 (d, ^3J = 7.4 Hz, 1H), 5.77 (d, ^3J = 6.8 Hz, 1H), 5.23 (t, ^3J = 5.4 Hz, 1H), 4.91 – 4.86 (m, 1H), 4.05 (ddd, ^3J = 9.1, 6.8, 5.4 Hz, 1H), 3.69 (ddd, ^2J = 12.3 Hz, ^3J = 4.9, 2.0 Hz, 1H), 3.56 (ddd, ^3J = 9.2, 5.0, 1.9 Hz, 1H), 3.50 – 3.42 (m, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 171.2, 160.8, 137.0, 108.9, 88.7, 81.5, 80.8, 69.7, 59.5.

α-furanosyl-uracil (9a)

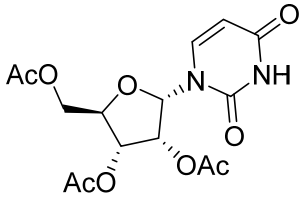


Compound **29** (1.0 g, 4.40 mmol) dissolved in aqueous HCl (2.2 ml, 0.2 M) was refluxed for 2 h. The solvent was removed in vacuo and the residue was taken up in H_2O. The solution was neutralized until pH ~6 with Dowex 1X8. It was filtered and the resin was washed with H_2O. The filtrate was freeze dried to obtain the product as a colorless powder (1.02 g, 4.18 mmol, 95%). For LC-MS analysis, a small fraction of the product was further purified by reversed phase HPLC.

mp: 202 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 11.18 (s, 1H, HN3) 7.61 (d, ^3J = 8.1 Hz, 1H, HC6), 6.01 (d, ^3J = 4.6 Hz, 1H, HC1'), 5.56 (d, ^3J = 8.1 Hz, 1H, HC5), 5.49 (br, 1H, OH), 5.12 (br, 1H, OH), 4.80 (d, ^3J = 9.0 Hz, 1H, OH), 4.16 (t, ^3J = 4.5 Hz, 1H, HC2'), 4.06 – 3.98 (m, 2H, HC3', HC4'), 3.58 (dd, ^2J = 12.0 Hz, ^3J = 2.7 Hz, 1H, H_aC5), 3.42 (dd, ^2J = 12.2 Hz, ^3J = 4.1 Hz, 1H, H_bC5). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 163.3 (C4), 150.6 (C2), 142.8 (C6), 99.8 (C5), 85.1 (C1'), 84.0 (C4'), 70.4 (C3'), 70.3 (C2'), 61.2 (C7). **HRMS** (ESI+): calc.: $[C_9H_{13}N_2O_6]^+$ 245.0768, found: 245.0768 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3301 (w), 3058 (w), 2930 (w), 2805 (vw), 2342 (vw), 1659 (vs), 1463 (m), 1394 (m), 1269 (s), 1197 (m), 1096 (s), 1037 (vs), 1021 (vs), 990 (s), 927 (m), 857 (m), 806 (s), 763 (s), 718 (m).

Synthesis of α-furanosyl-cytidine

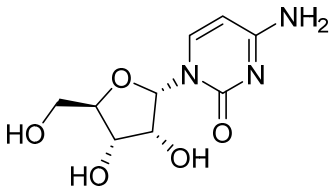
(1'S,2'R,3'S,4'R)-2-(Acetoxymethyl)-5-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-3,4-diyl diacetate (30)



Compound **9a** (500 mg, 2.05 mmol) was dissolved in pyridine (5 ml) and acetic anhydride (5 ml) at 0 °C. The mixture was reacted for 18 h and the solvent removed *in vacuo*. It was co-evaporated with MeCN (20 ml). The residue was purified by flash column chromatography (DCM/MeOH 97:3) to afford the product as a colorless solid (611 mg, 1.65 mmol, 80%).

mp: 70 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 11.38 (s, 1H, HN3), 7.65 (d, ^3J = 8.2 Hz, 1H, HC6), 6.34 (d, ^3J = 4.7 Hz, 1H, HC1'), 5.64 (d, ^3J = 8.2 Hz, HC5), 5.54 (t, ^3J = 4.7 Hz, 1H, HC2'), 5.35 (dd, ^3J = 6.2, 5.1 Hz, 1H, HC3'), 4.66 – 4.60 (m, 1H, HC4'), 4.26 (dd, ^2J = 12.2 Hz, ^3J = 3.4 Hz, 1H, H_aC5'), 4.16 (dd, ^2J = 12.2 Hz, ^3J = 5.7 Hz, 1H, H_bC5'), 2.06 (s, 3H, CH_3), 2.03 (s, 3H, CH_3), 1.97 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 170.1 (CO), 169.4 (CO), 168.8 (CO), 163.1 (C4), 150.2 (C2), 141.0 (C6), 100.9 (C5), 83.5 (C1'), 78.8 (C4'), 70.6 (C3') 69.8 (C2'), 63.1 (C5'), 20.6 (CH_3), 20.4 (CH_3), 20.1 (CH_3). **HRMS** (ESI+): calc.: $[C_{15}H_{18}N_2NaO_9]^+$ 393.0905, found: 393.0904 $[M+Na]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3059 (vw), 2159 (vw), 1743 (s), (1682 (vs), 1455 (w), 1371 (m), 1211 (vs), 1109 (s), 1075 (m), 1030 (s), 944 (w), 902 (w), 813 (m), 766 (w), 714 (w), 674 (vw).

α-furanosyl-cytidine (8a)

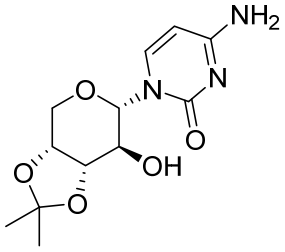


Compound **30** (100 mg, 0.27 mmol) was dissolved in dry DCM. It was added Et_3N (105 µL, 78.1 mg, 0.77 mmol), DMAP (6.00 mg, 0.05 mmol) and 2,4,6-TIPBS (164 mg, 0.54 mmol) and stirred for 20 h at rt. The reaction was quenched with sat. aqueous ammonia (6 ml) and further stirred for 3 h at rt. The solvent was removed *in vacuo* and residual water co-evaporated with MeCN (2 x 25 ml). The residue was purified by flash column chromatography (DCM/MeOH 88:12) and reversed phase HPLC to afford the product as a colorless solid (15 mg, 0.06 mmol, 23%)

mp: 190 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 7.52 (d, ^3J = 7.4 Hz, 1H, HC6), 7.05 (br, 1H, H_aN), 6.96 (br, 1H, H_bN), 6.01 (d, ^3J = 3.7 Hz, 1H, HC1'), 5.66 (d, ^3J = 7.4 Hz, 1H, HC5), 5.27 (s, 1H, OH), 4.97 (s, 1H, OH), 4.77 (s, 1H, OH), 4.07 – 4.01 (m, 2H, HC2', HC3'), 3.98 – 3.92 (m, 1H, HC4'), 3.61 (dd, ^2J = 12.1 Hz, ^3J = 2.6 Hz, 1H, HaC5'), 3.42 (dd, ^2J = 12.2 Hz, ^3J = 4.6 Hz, 1H, HbC5').**^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 165.6 (C4), 155.2 (C2), 143.1 (C6), 92.3 (C5), 85.6 (C1'), 83.1 (C4'), 70.6 (C2'), 70.1 (C3'), 61.1 (C5'). **HRMS** (ESI+): calc.: $[C_9H_{14}N_3O_5]^+$ 244.0928, found: 244.0927 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3267 (w), 2360 (vs), 2222 (s), 1842 (m), 1589 (m), 1311 (w), 1210 (w), 1033 (m), 793 (m), 668 (vs).

Synthesis of α-pyranosyl-cytidine

4-Amino-1-((1'R,2'S,3'S,4'R)-7-hydroxy-2,2-dimethyltetrahydro-4H-[1,3]dioxolo[4,5-c]pyran-6-yl)pyrimidin-2(1H)-on (31)

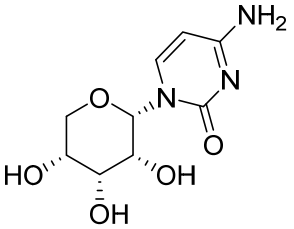


Compound **25** (143 mg, 0.44 mmol) was dissolved in dry DCM (9 ml). It was added triethylamine (118 mg, 162 ml, 1.60 mmol), DMAP (8 mg, 0.07 mmol, 0.15 eq.) and 2,4,6-TIPBS (265 mg, 0.88 mmol). The mixture was reacted for 72 h at rt and quenched with sat. ammonia in H_2O (3 ml). The reaction was further stirred for 5 h at rt. The solvent was removed *in vacuo* und residual H_2O was co-evaporated with MeCN (2 x 25 ml). The crude mixture was purified by flash column chromatography (DCM/MeOH 12:1 → 4:1, containing 0.1% Et_3N) to obtain the product as a white solid (67 mg, 0.21 mmol, 48%).

mp: 210 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 7.54 (d, ^3J = 7.5 Hz, 1H, HC6), 7.23 (br, 1H, H_aN), 7.11 (br, 1H, H_bN), 5.71 (d, ^2J = 7.4 Hz, 1H, HC5), 5.41 (d, ^3J = 6.4 Hz, 1H, OH), 5.36 (d, ^3J = 9.8 Hz, 1H, HC1'), 4.22 – 4.18 (m, 1H, HC4'), 4.14 – 4.07 (m, 2H, H_aC5', HC3'), 3.85 (dd, ^2J = 13.6 Hz, ^3J = 2.5

Hz, 1H, H_bC5'), 3.67 – 3.57 (m, 1H, HC3'), 1.48 (s, 3H, CH_3), 1.29 (s, 3H, CH_3). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 165.4 (C4), 155.5 (C2), 141.7 (C6), 108.5 (C_q), 94.2 (C5), 82.3 (C1'), 79.4 (C3'), 73.3 (C4'), 69.9 (C2'), 65.2 (C5'), 28.0 (CH_3), 26.2 (CH_3). **HRMS** (ESI+): calc.: $[C_{12}H_{18}N_3O_5]^+$ 284.1240, found: 284.1241 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3326 (w), 2932 (w), 2819 (vw), 2360 (vw), 1660 (vs), 1525 (vw), 1465 (m), 1395 (m), 1270 (s), 1203 (m), 1096 (s), 1039 (vs), 1022 (s), 928 (w), 869 (m), 808 (s), 763 (m), 719 (m).

α-pyranosyl-cytidine (8c)

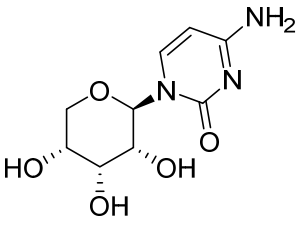


Compound **31** (35 mg, 0.12 mmol) was dissolved in MeCN (5.5 ml). Dess-Martin-Periodinan (104 mg, 0.25 mmol) was added and the reaction mixture was stirred for 3.5 h at rt. The reaction was quenched with $Na_2S_2O_3$ (500 mg) und $NaHCO_3$ (100 mg) and further stirred for 10 min. The mixture was filtered and the solvent removed *in vacuo*. The crude keto product was dissolved in DCM (3 mL), EtOAc (1.5 mL) und MeOH (1.5 mL). $NaBH_4$ (20 mg, 0.529 mmol) was added and stirred for 2.5 h at rt. The solvent was removed *in vacuo*. The crude product was dissolved in 70% acetic acid (5 ml) and stirred for 5 h at 60 °C. The solvent was removed *in vacuo* and co-evaporated with EtOH (2 x 30 ml) and MeCN (3 x 30 ml). The crude mixture was purified by reversed phase HPLC to obtain the product as colorless solid (7.0 mg, 0.029 mmol, 24 %).

mp: 194 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm) = 7.61 (d, ^3J = 7.4 Hz, 1H, HC6), 7.18 (br, 1H, H_aN), 7.02 (br, 1H, H_bN), 5.67 (d, ^3J = 7.4 Hz, 1H, HC5), 5.47 (s, 1H, HC1'), 5.13 (d, ^3J = 6.0 Hz, 1H, OH), 5.10 (d, ^3J = 5.8 Hz, 1H, OH), 5.07 (d, ^3J = 7.7 Hz, 1H, OH), 3.95 (dd, ^2J = 12.2 Hz, ^3J = 1.8 Hz, 1H, H_aC5'), 3.75 – 3.63 (m, 4H, HC2, HC3, HC4, H_bC5').**^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 165.5 (C4), 154.4 (C2), 143.1 (C6), 92.5 (C5), 82.3 (C1'), 70.6 (C2'), 70.3 (C5'), 68.7 (C4'), 67.4 (C3'). **HRMS** (ESI+): calc.: $[C_9H_{14}N_3O_5]^+$ 244.0928, found: 244.0927 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3217 (w), 2360 (vw), 1891 (vw), 1645 (s), 1598 (m), 1525 (w), 1481 (m), 1405 (m), 1299 (m), 1203 (m), 1158 (m), 1084 (vs), 1015 (s), 987 (s), 902 (m), 833 (m), 780 (vs), 752 (s), 668 (vs).

Synthesis of β-pyranosyl-cytidine

β-pyranosyl-cytidine (8d)



Nucleoside **24** (50.0 mg, 0.135 mmol) was dissolved in dry DCM. It was added NEt_3 (27.3 mg, 37.4 µL, 0.270 mmol), DMAP (1.7 mg, 0.014 mmol, 0.1 eq.) and 2,4,6-TIPBS (61.5 mg, 0.203 mmol) and stirred

for 17 h at rt. The reaction was quenched with sat. ammonia in H_2O. The solvent was removed in vacuo and residual water was co-evaporated with MeCN (4 x 20 ml). The crude product was purified by reversed phase HPLC to obtain the pure product as a colorless solid (15 mg, 0.062 mmol, 46%).

mp: 180 °C. **^{1}H NMR** (599 MHz, DMSO-d_6) δ (ppm) = 7.54 (d, ^3J = 7.4 Hz, 1H, HC6), 7.16 (br, 1H, H_aN), 7.05 (br, 1H, H_bN), 5.70 (d, ^3J = 9.6 Hz, 1H, HC1'), 5.68 (d, J = 7.5 Hz, 1H, HC5), 5.01 (d, ^3J = 3.5 Hz, 1H, OH), 4.83 (d, ^3J = 7.2 Hz, 1H, OH), 4.78 (br, 1H, OH), 3.96 (s, 1H, HC3'), 3.64 – 3.57 (m, 2H, HC4', HC2'), 3.54 (d, ^2J = 10.2 Hz, 1H, H_aC5'), 3.50 (dd, ^2J = 10.3 Hz, ^3J = 5.0 Hz, 1H, H_bC5'). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm) = 165.3 (C4), 155.7 (C2), 141.8 (C6), 93.9 (C5), 79.8 (C1'), 71.2 (C3'), 68.0 (C2'), 66.7 (C4'), 65.2 (C5'). **HRMS** (ESI+): calc.: $[C_9H_{14}N_3O_5]^+$ 244.0928, found: 244.0927 $[M+H]^+$. **IR** (cm^{-1}): $\tilde{\nu}$ = 3426 (w), 3210 (m), 2922 (w), 1640 (vs), 1596 (vs), 1526 (m), 1489 (vs), 1400 (m), 1387 (m), 1368 (m), 1322 (w), 1300 (m), 1281 (m), 1269 (m), 1242 (m), 1224 (m), 1204 (s), 1132 (s), 1115 (m), 1094 (s), 1076 (s), 1039 (vs), 1018 (s), 991 (s), 970 (m), 800 (m), 781 (s), 660 (vs).

Synthesis of lüneburgite: $Mg_3(H_2O)_6[B_2(OH)_6(PO_4)_2]$

The synthesis was performed according to literature.[5]

Magnesiumoxide (1 g, 25 mmol), dimagnesiumphosphate trihydrate (8.1 g, 46.5 mmol) and boric acid (9.2 g, 149 mmol) were refluxed in H_2O (200 ml) for 4 days. After cooling to rt, the colorless solid was filtered off to afford the product (9.5 g, 19.2 mmol, 83%). The mineral was analysed by XRD measurement to confirm the correct product (see Fig. S5).

Supplementary Figures

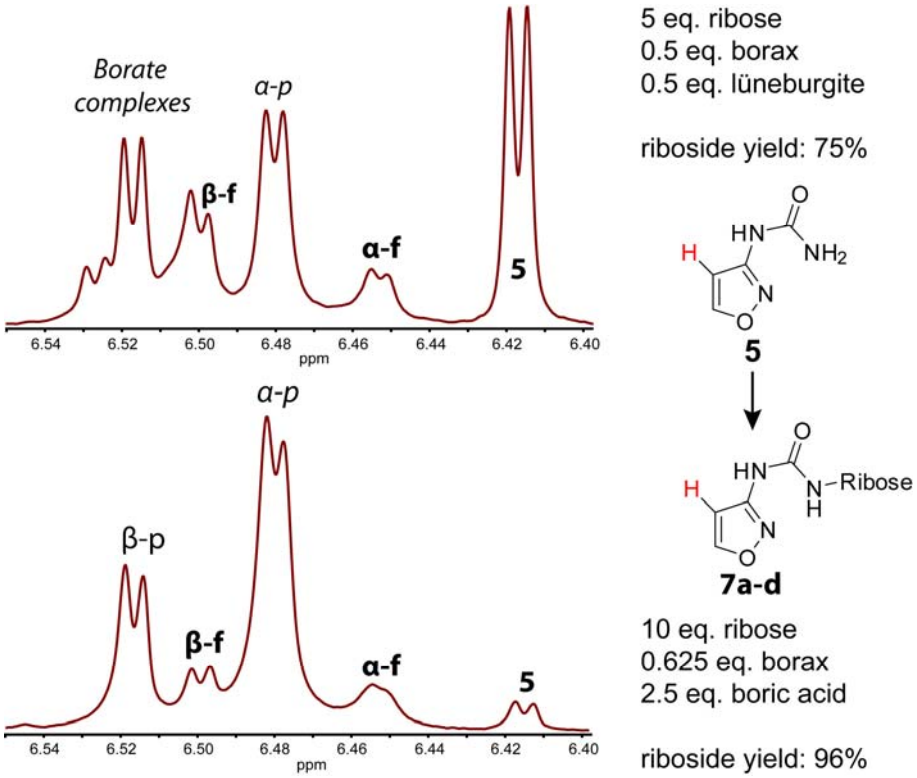


Fig. S1. ^{1}H-NMR in D_2O after **5** was reacted with ribose under different conditions. The ribosylated product **7a-d** is obtain as a mixture of α- and β-anomers as either furanoside (f) or pyranoside (p). The ^{1}H-NMR signals observed correspond to the proton labelled in red. The yields are relative to the starting material **5**.

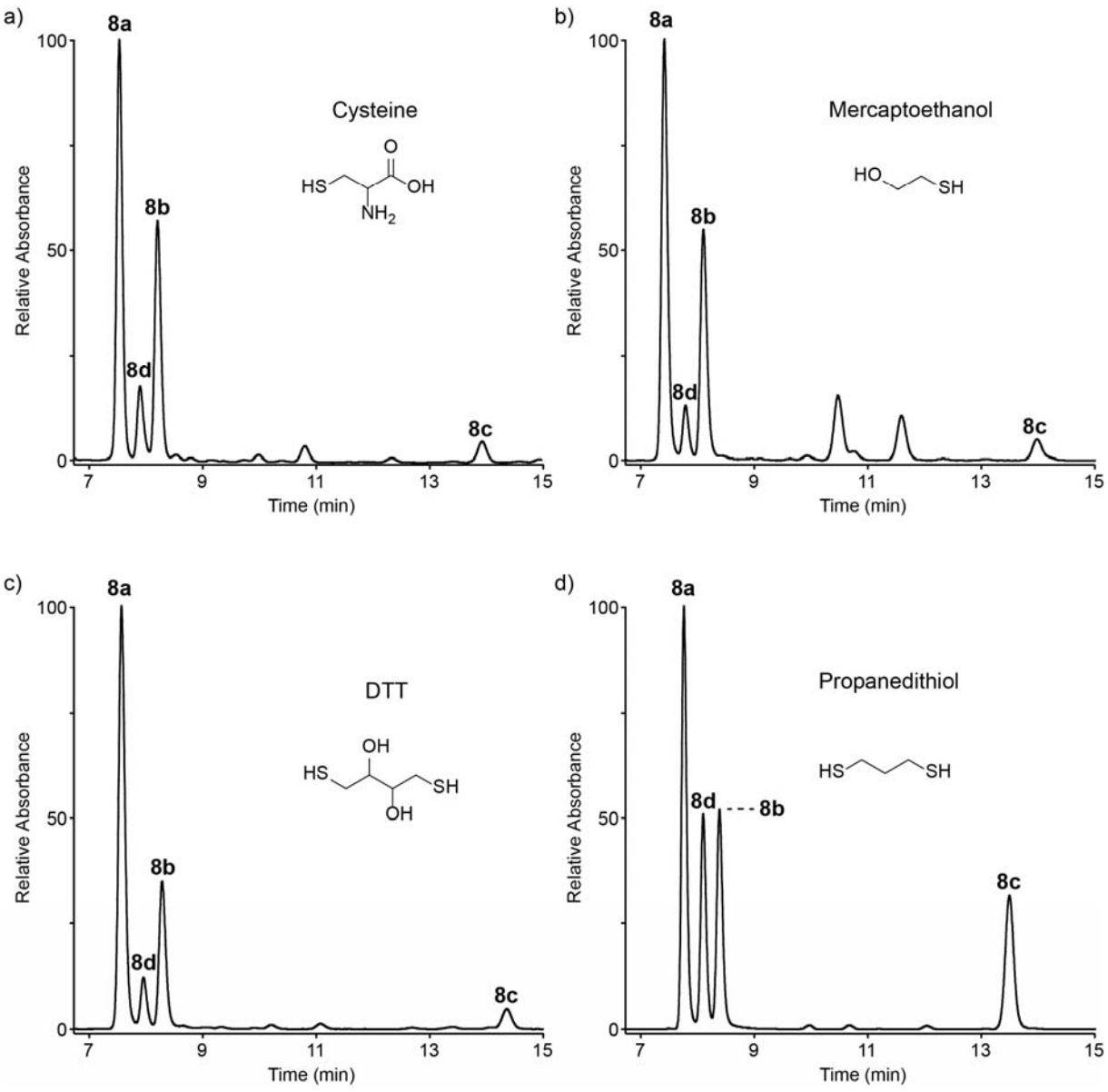


Fig. S2. Cytidine formation in the presence of different thiols. All reductions were performed with 0.001 eq. of soluble Fe^{2+} in 100 mM sodium carbonate, 50 mM borate buffer (pH 9.7). Reduction in the presence of different thiols: a) cysteine, b) mercaptoethanol, c) dithiothreitol (DTT) and d) propanedithiol.

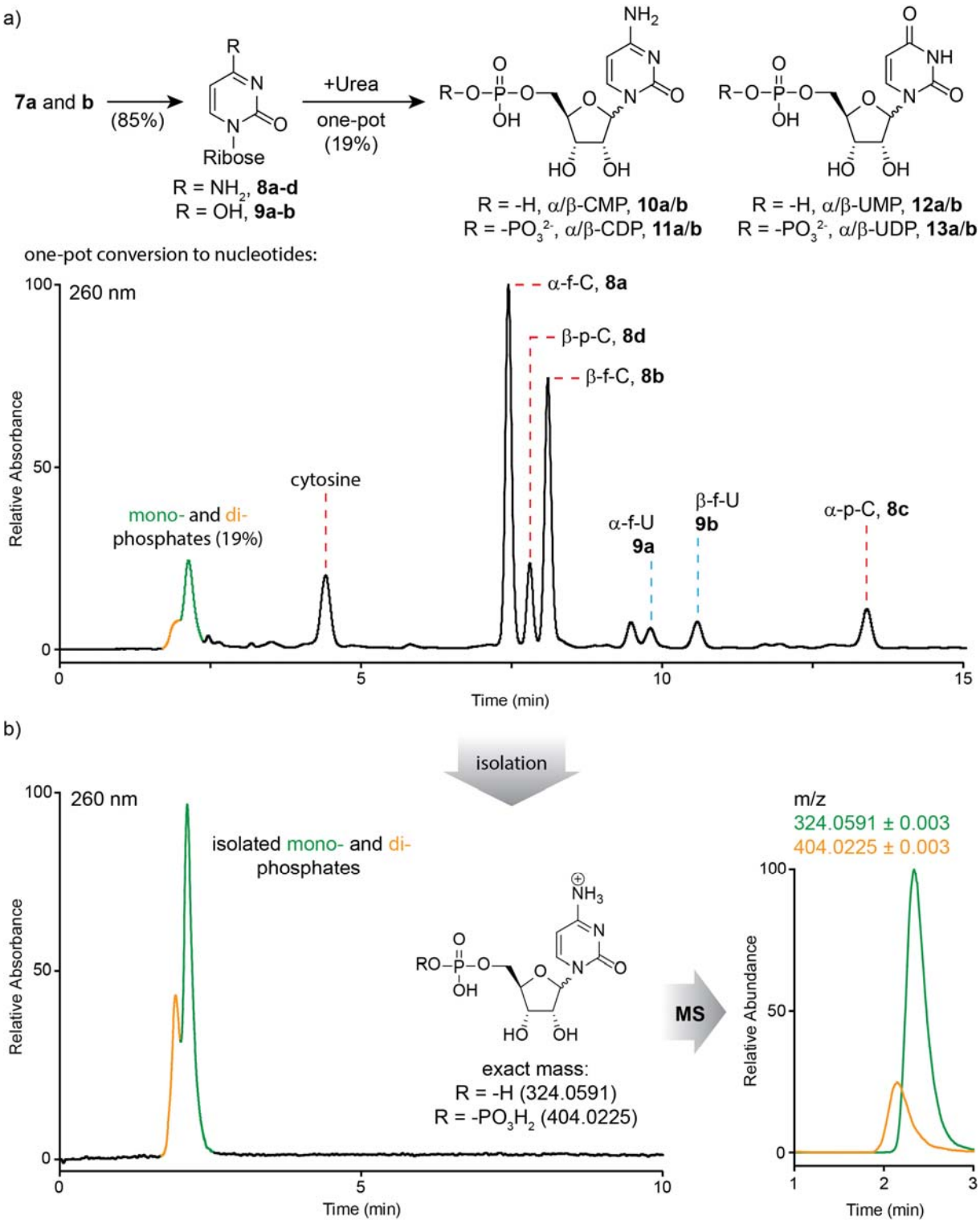


Fig. S3. Prebiotic nucleotide synthesis. a) Reaction scheme and full UV-chromatogram after nucleotide formation. An extract was shown in Fig. 5b in the main text. The yield was calculated relative to all cytidine isomers. The mono- (green) and di-nucleotide (orange) peaks are highlighted. b) UV-chromatogram and corresponding mass (ESI+) for the isolated nucleotides.

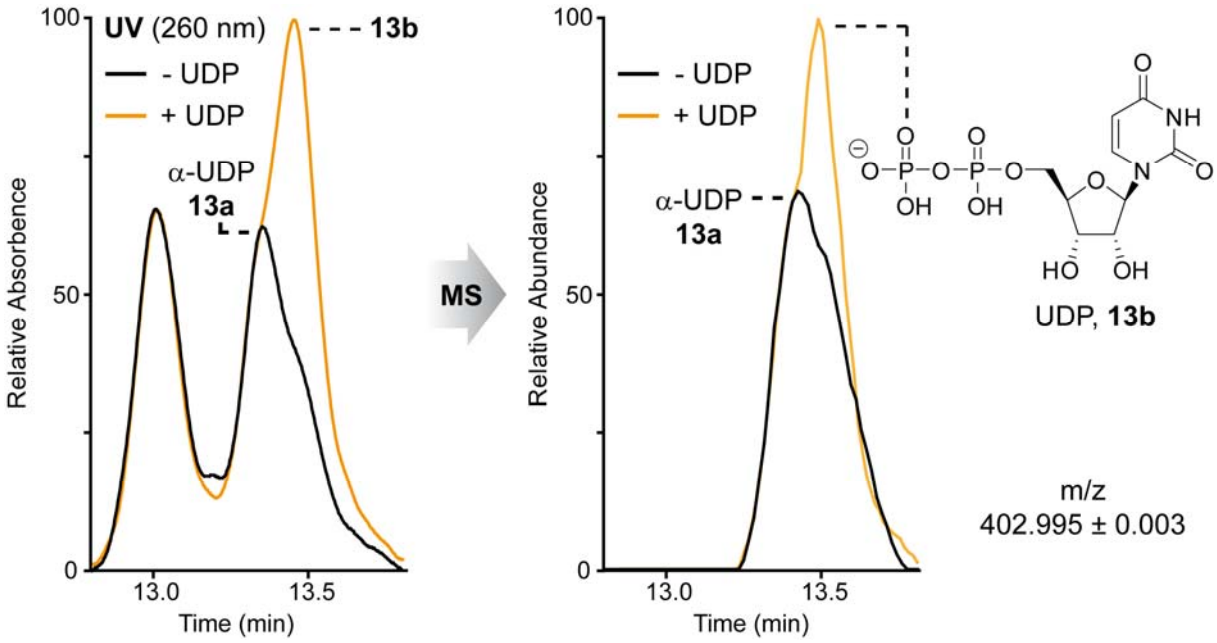


Fig. S4. Co-injection study for 5'-uridine-di-nucleotide **13b**. For this experiment we used the same buffers as stated in the general information, but a different gradient: 0 → 15 min, Buffer B 0% → 10%. Co-injection was followed by UV and MS detection with commercially available material.

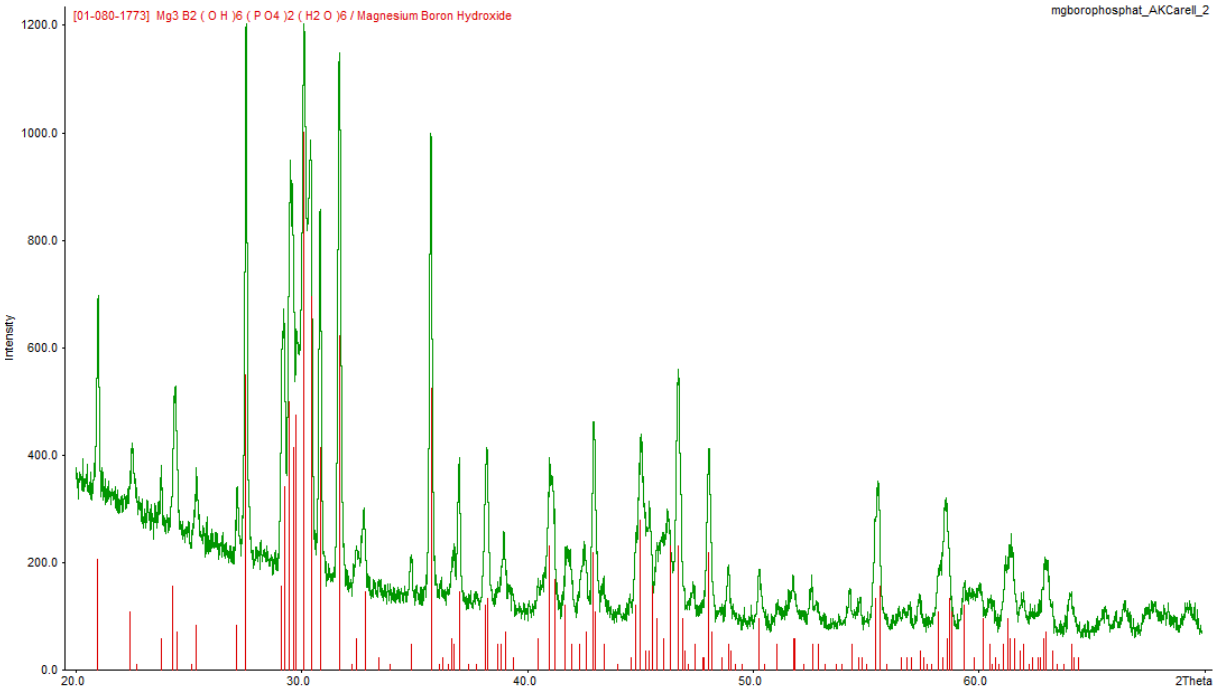


Fig. S5. XRD measurement of synthetic lueneburgite (green) and comparison with database values (red).

1 Becker, S. *et al.* A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833-836 (2016).
2 Shannahoff, D. & Sanchez, R. *J. Org. Chem.*, 593-598 (1973).
3 Davidson, R., White, E., Margolis, S. & Coxon, B. *Carbohydr. Res.*, 239-254 (1983).
4 Lemieux, T., Nagabhushan, T. & Paul, B. *Can. J. Chem.* **50**, 773-776 (1972).
5 Berdesinski, W. Synthetische Darstellung von Lüneburgit. *Naturwissenschaften* **38**, 476-477 (1951).

7 Abkürzungsverzeichnis

A	Adenin oder Adenosin
Abb.	Abbildung
BA	Barbitursäure oder Barbitursäurenukleosid
C	Cytosin oder Cytidine
D	Dextrorotation
DA	Diaminopurin oder Diaminopurinnukleosid
DNA	Desoxyribonukleinsäure
FaPy	Formamidopyrimidin
G	Guanin oder Guanosin
L	Levorotation
LUCA	letzter allgemeiner gemeinsamer Vofahr
m^2A	2-Methyladeonsin
m^1G	*N*-1-Methylguanosin
m^2G	*N*-2-Methylguanosin
$m^2{}_2G$	*N*-2,2-dimethylguanosine
mRNA	messenger RNA
ms^2A	2-Thiomethyladenosin
pH	potentia Hydrogenii
quant.	quantitativ
RNA	Ribonukleinsäure
rRNA	ribosomale RNA
s^2U	2-Thiouridin
SELEX	Systematische Evolution von Liganden durch exponentielle Anreicherung
T	Thymin oder Thymidin
t^6A	*N*-6-Threonylcarbamoyladenosin
tRNA	transfer RNA
U	Uracil oder Uridin
UV	ultraviolett

8 Literaturverzeichnis

[1] M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 16116.
[2] J. D. Sutherland, *Nat. Rev. Chem.* **2017**, *1*, 0012.
[3] F. Schüth, *Angew. Chem. Int. Ed.* **2016**, *55*, 14878-14879.
[4] P. Ball, *Nature* **2006**, *442*, 500-502.
[5] M. S. Dodd, D. Papineau, T. Grenne, J. F. Slack, M. Rittner, F. Pirajno, J. O'Neil, C. T. S. Little, *Nature* **2017**, *543*, 60-64.
[6] A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821-12844.
[7] A. I. Oparin, *The Origin of Life: Transl. with Annot. by Sergius Morgulis*, Dover Publ., **1953**.
[8] A. I. Oparin, *Proiskhozhedenie Zhizni; Moscow: Mosckovskii Rabochii (1924)*, Reprinted and translated. In: Bernal JD (ed). The origin of life. London: Weidenfeld and Nicolson (1967).
[9] J. B. S. Haldane, *Rationalist Annual* **1929**, *148*, 3-10.
[10] S. L. Miller, *Science* **1953**, *117*, 528-529.
[11] F. Wöhler, *Ann. Phys. Chem.* **1824**, *3*, 177-182.
[12] F. Wöhler, *Ann. Phys.* **1828**, *88*, 253-256.
[13] A. Strecker, *Liebigs Ann.* **1850**, *75*, 27-45.
[14] A. Strecker, *Liebigs Ann.* **1854**, *91*, 349-351.
[15] A. Butlerow, *Liebigs Ann.* **1861**, *120*, 295-298.
[16] W. Löb, *Z. Elektrochem. Angew. Phys. Chem.* **1906**, *12*, 282-312.
[17] W. Löb, *Chem. Ber.* **1913**, *46*, 684-697.
[18] S. L. Miller, *J. Am. Chem. Soc.* **1955**, *77*, 2351-2361.
[19] S. L. Miller, H. C. Urey, *Science* **1959**, *130*, 245-251.
[20] S. L. Miller, *Ann. N. Y. Acad. Sci.* **1957**, *69*, 260-275.
[21] P. H. Abelson, *Proc. Natl. Acad. Sci. USA* **1966**, *55*, 1365-1372.
[22] S. Moorbath, R. K. O'Nions, R. J. Pankhurst, *Nature* **1973**, *245*, 138-139.
[23] M. Schidlowski, P. W. U. Appel, R. Eichmann, C. E. Junge, *Geochim. Cosmochim. Acta* **1979**, *43*, 189-199.
[24] A. N. N. Henderson-Sellers, A. J. Meadows, *Nature* **1977**, *270*, 589-591.
[25] T. Owen, R. D. Cess, V. Ramanathan, *Nature* **1979**, *277*, 640-642.
[26] J. C. G. Walker, *Evolution of the atmosphere*, Macmillan, New York, **1977**.
[27] J. Kasting, *Science* **1993**, *259*, 920-926.
[28] R. Stribling, S. L. Miller, *Orig. Life Evol. Biosph.* **1987**, *17*, 261-273.
[29] G. Schlesinger, S. L. Miller, *J. Mol. Evol.* **1983**, *19*, 376-382.
[30] J. P. Ferris, W. J. Hagan, *Tetrahedron* **1984**, *40*, 1093-1120.
[31] J. Oró, *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407-412.
[32] J. Oro, S. S. Kamat, *Nature* **1961**, *190*, 442-443.
[33] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **2015**, *7*, 301-307.
[34] D. Ritson, J. D. Sutherland, *Nat. Chem.* **2012**, *4*, 895-899.
[35] W. L. Chameides, J. C. G. Walker, *Orig. Life* **1981**, *11*, 291-302.
[36] Y. L. Yung, M. B. McElroy, *Science* **1979**, *203*, 1002-1004.
[37] J. F. Kasting, *Orig. Life Evol. Biosph.* **1990**, *20*, 199-231.
[38] D. P. Summers, S. Chang, *Nature* **1993**, *365*, 630-633.
[39] V. S. Airapetian, A. Glocer, G. Gronoff, E. Hebrard, W. Danchi, *Nat. Geosci.* **2016**, *9*, 452-455.
[40] D. P. Summers, R. C. B. Basa, B. Khare, D. Rodoni, *Astrobiology* **2012**, *12*, 107-114.
[41] B. Fegley, R. G. Prinn, H. Hartman, G. H. Watkins, *Nature* **1986**, *319*, 305-308.
[42] C. Chyba, C. Sagan, *Nature* **1992**, *355*, 125-132.
[43] S. Sugita, P. H. Schultz, *Geophys. Res. Lett.* **2009**, *36*, L20204.
[44] F. Tera, D. A. Papanastassiou, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **1974**, *22*, 1-21.
[45] B. A. Cohen, T. D. Swindle, D. A. Kring, *Science* **2000**, *290*, 1754-1756.
[46] T. Goldin, *Nat. Geosci.* **2012**, *5*, 309-309.
[47] C. Chyba, P. Thomas, L. Brookshaw, C. Sagan, *Science* **1990**, *249*, 366-373.
[48] Y. Furukawa, T. Sekine, M. Oba, T. Kakegawa, H. Nakazawa, *Nat. Geosci.* **2009**, *2*, 62-66.
[49] R. Hayatsu, *Science* **1964**, *146*, 1291-1293.
[50] Z. Martins, O. Botta, M. L. Fogel, M. A. Sephton, D. P. Glavin, J. S. Watson, J. P. Dworkin, A. W. Schwartz, P. Ehrenfreund, *Earth Planet. Sci. Lett.* **2008**, *270*, 130-136.

[51] A. S. Burton, J. C. Stern, J. E. Elsila, D. P. Glavin, J. P. Dworkin, *Chem. Soc. Rev.* **2012**, *41*, 5459-5472.
[52] S. Pizzarello, E. Shock, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.
[53] M. A. Sephton, *Nat. Prod. Rep.* **2002**, *19*, 292-311.
[54] M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House, J. P. Dworkin, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13995-13998.
[55] N. C. Wickramasinghe, *Nature* **1974**, *252*, 462-463.
[56] L. M. Ziurys, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12274-12279.
[57] P. Ehrenfreund, S. B. Charnley, *Annu. Rev. Astron. Astrophys.* **2000**, *38*, 427-483.
[58] E. Herbst, *Chem. Soc. Rev.* **2001**, *30*, 168-176.
[59] J. E. Elsila, D. P. Glavin, J. P. Dworkin, *Meteorit. Planet. Sci.* **2009**, *44*, 1323-1330.
[60] F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Krüger, L. Le Roy, A. J. MacDermott, S. McKenna-Lawlor, U. J. Meierhenrich, G. M. M. Caro, F. Raulin, R. Roll, A. Steele, H. Steininger, R. Sternberg, C. Szopa, W. Thiemann, S. Ulamec, *Science* **2015**, *349*.
[61] K. Altwegg, H. Balsiger, A. Bar-Nun, J.-J. Berthelier, A. Bieler, P. Bochsler, C. Briois, U. Calmonte, M. R. Combi, H. Cottin, J. De Keyser, F. Dhooghe, B. Fiethe, S. A. Fuselier, S. Gasc, T. I. Gombosi, K. C. Hansen, M. Haessig, A. Jäckel, E. Kopp, A. Korth, L. Le Roy, U. Mall, B. Marty, O. Mousis, T. Owen, H. Rème, M. Rubin, T. Sémon, C.-Y. Tzou, J. Hunter Waite, P. Wurz, *Sci. Adv.* **2016**, *2*.
[62] R. Saladino, G. Botta, S. Pino, G. Costanzo, E. Di Mauro, *Chem. Soc. Rev.* **2012**, *41*, 5526-5565.
[63] S. L. Miller, A. Lazcano, *J. Mol. Evol.* **1995**, *41*, 689-692.
[64] J. L. Bada, A. Lazcano, *Science* **2002**, *296*, 1982-1983.
[65] M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7933-7938.
[66] R. Shapiro, *Orig. Life Evol. Biosph.* **1995**, *25*, 83-98.
[67] J. L. Bada, C. Bigham, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1248-1250.
[68] E. R. Garrett, J. Tsau, *J. Pharm. Sci.* **1972**, *61*, 1052-1061.
[69] R. Shapiro, R. S. Klein, *Biochemistry* **1966**, *5*, 2358-2362.
[70] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8158-8160.
[71] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196-196.
[72] J. B. Lambert, S. A. Gurusamy-Thangavelu, K. Ma, *Science* **2010**, *327*, 984-986.
[73] S. Miyakawa, H. James Cleaves, S. L. Miller, *Orig. Life Evol. Biosph.* **2002**, *32*, 195-208.
[74] R. Sanchez, J. Ferris, L. E. Orgel, *Science* **1966**, *153*, 72-73.
[75] D. O. Gough, *Sol. Phys.* **1981**, *74*, 21-34.
[76] M. J. NEWMAN, R. T. ROOD, *Science* **1977**, *198*, 1035-1037.
[77] N. B. John, M. H. Pinsonneault, B. Sarbani, *Astrophys. J.* **2001**, *555*, 990.
[78] A. E. Ringwood, *Geochim. Cosmochim. Acta* **1961**, *21*, 295-296.
[79] S. Miyakawa, H. J. Cleaves, S. L. Miller, *Orig. Life Evol. Biosph.* **2002**, *32*, 209-218.
[80] S. A. Wilde, J. W. Valley, W. H. Peck, C. M. Graham, *Nature* **2001**, *409*, 175-178.
[81] S. J. Mojzsis, T. M. Harrison, R. T. Pidgeon, *Nature* **2001**, *409*, 178-181.
[82] D. R. Lowe, *Annu. Rev. Earth Planet. Sci.* **1980**, *8*, 145-167.
[83] J. C. G. Walker, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **1982**, *40*, 1-11.
[84] T. M. Harrison, *Annu. Rev. Earth Planet. Sci.* **2009**, *37*, 479-505.
[85] R. K. Ulrich, *Science* **1975**, *190*, 619-624.
[86] J. F. B. Mitchell, *Rev. Geophys.* **1989**, *27*, 115-139.
[87] C. Sagan, G. Mullen, *Science* **1972**, *177*, 52-56.
[88] J. C. G. Walker, K. J. Zahnle, *Nature* **1986**, *320*, 600-602.
[89] R. Arevalo Jr, W. F. McDonough, M. Luong, *Earth Planet. Sci. Lett.* **2009**, *278*, 361-369.
[90] A. S. Endal, K. H. Schatten, *J. Geophys. Res. Oceans* **1982**, *87*, 7295-7302.
[91] N. H. Sleep, K. J. Zahnle, J. F. Kasting, H. J. Morowitz, *Nature* **1989**, *342*, 139-142.
[92] S. R. Taylor, S. McLennan, *Planetary Crusts: Their Composition, Origin and Evolution*, Cambridge University Press, **2009**.
[93] L. P. Knauth, S. Epstein, *Geochim. Cosmochim. Acta* **1976**, *40*, 1095-1108.
[94] J. Karhu, S. Epstein, *Geochim. Cosmochim. Acta* **1986**, *50*, 1745-1756.
[95] H. D. Holland, *The Chemical Evolution of the Atmosphere and Oceans*, Princeton University Press, **1984**.
[96] M. T. Hren, M. M. Tice, C. P. Chamberlain, *Nature* **2009**, *462*, 205-208.
[97] R. E. Blake, S. J. Chang, A. Lepland, *Nature* **2010**, *464*, 1029-1032.
[98] K. Kashefi, D. R. Lovley, *Science* **2003**, *301*, 934-934.
[99] D. A. Cowan, *Trends Microbiol.* **2004**, *12*, 58-60.

[100] N. G. Holm, *Marine Hydrothermal Systems and the Origin of Life: Report of SCOR Working Group 91*, Springer Netherlands, **2012**.
[101] W. Martin, J. Baross, D. Kelley, M. J. Russell, *Nat. Rev. Microbiol.* **2008**, *6*, 805-814.
[102] W. Martin, M. J. Russell, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2007**, *362*, 1887-1926.
[103] U. Trinks, A. Eschenmoser, ETH Zürich (Zürich), **1987**.
[104] R. Lohrmann, *J. Mol. Evol.* **1972**, *1*, 263-269.
[105] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25-33.
[106] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249-257.
[107] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833-836.
[108] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.
[109] C. de Duve, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8253-8256.
[110] G. Wächtershäuser, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 200-204.
[111] I. A. Berg, D. Kockelkorn, W. H. Ramos-Vera, R. F. Say, J. Zarzycki, M. Hügler, B. E. Alber, G. Fuchs, *Nat. Rev. Microbiol.* **2010**, *8*, 447-460.
[112] G. Wächtershäuser, *Microbiol. Rev.* **1988**, *52*, 452-484.
[113] H. J. Morowitz, J. D. Kostelnik, J. Yang, G. D. Cody, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7704-7708.
[114] K. B. Muchowska, S. J. Varma, E. Chevallot-Beroux, L. Lethuillier-Karl, G. Li, J. Moran, *Nat. Ecol. Evol.* **2017**.
[115] E. Camprubi, S. F. Jordan, R. Vasiliadou, N. Lane, *IUBMB Life* **2017**, *69*, 373-381.
[116] R. Braakman, E. Smith, *PLOS Comput. Biol.* **2012**, *8*, e1002455.
[117] A. Roldan, N. Hollingsworth, A. Roffey, H. U. Islam, J. B. M. Goodall, C. R. A. Catlow, J. A. Darr, W. Bras, G. Sankar, K. B. Holt, G. Hogarth, N. H. de Leeuw, *Chem. Commun.* **2015**, *51*, 7501-7504.
[118] C. He, G. Tian, Z. Liu, S. Feng, *Org. Lett.* **2010**, *12*, 649-651.
[119] C. Huber, G. Wächtershäuser, *Science* **1997**, *276*, 245-247.
[120] R. Shapiro, *Sci. Am.* **2007**, *269*, 46-53.
[121] L. E. Orgel, *PLOS Biol.* **2008**, *6*, e18.
[122] D. S. Ross, *Orig. Life Evol. Biosph.* **2007**, *37*, 61-65.
[123] V. Vasas, E. Szathmáry, M. Santos, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1470-1475.
[124] M. Eigen, *Naturwissenschaften* **1971**, *58*, 465-523.
[125] C. Woese, *The genetic code : the molecular basis for genetic expression*, Harper & Row, New York, **1967**.
[126] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381-393.
[127] H. B. White, *J. Mol. Evol.* **1976**, *7*, 101-104.
[128] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849-857.
[129] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147-157.
[130] T. A. Steitz, P. B. Moore, *Trends Biochem. Sci.* **2003**, *28*, 411-418.
[131] W. Gilbert, *Nature* **1986**, *319*.
[132] T. Lindahl, *Nature* **1993**, *362*, 709.
[133] D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc. Natl. Acad. Sci. USA* **1967**, *58*, 217-224.
[134] D. P. Horning, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9786-9791.
[135] J. T. Sczepanski, G. F. Joyce, *Nature* **2014**, *515*, 440.
[136] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229-1232.
[137] J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, P. Holliger, *eLife* **2018**, *7*, e35255.
[138] R. Saladino, E. Carota, G. Botta, M. Kapralov, G. N. Timoshenko, A. Y. Rozanov, E. Krasavin, E. Di Mauro, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2746-E2755.
[139] A. Hill, L. E. Orgel, *Orig. Life Evol. Biosph.* **2002**, *32*, 99-102.
[140] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.
[141] A. P. Nutman, V. C. Bennett, C. R. L. Friend, M. J. Van Kranendonk, A. R. Chivas, *Nature* **2016**, *537*, 535.
[142] T. Tashiro, A. Ishida, M. Hori, M. Igisu, M. Koike, P. Méjean, N. Takahata, Y. Sano, T. Komiya, *Nature* **2017**, *549*, 516.
[143] T. Djokic, M. J. Van Kranendonk, K. A. Campbell, M. R. Walter, C. R. Ward, *Nat. Commun.* **2017**, *8*, 15263.
[144] D. J. Ritson, J. D. Sutherland, *Angew. Chem. Int. Ed.* **2013**, *52*, 5845-5847.
[145] R. A. Sanchez, J. P. Ferbis, L. E. Orgel, *J. Mol. Biol.* **1967**, *30*, 223-253.
[146] J. OrÓ, *Nature* **1961**, *191*, 1193.

[147] J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1962**, *96*, 293-313.
[148] M. Levy, S. L. Miller, J. Oró, *J. Mol. Evol.* **1999**, *49*, 165-168.
[149] J. P. Ferris, P. C. Joshi, J. G. Lawless, *BioSystems* **1977**, *9*, 81-86.
[150] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *Science* **1966**, *154*, 784-785.
[151] H. B. Niemann, S. K. Atreya, S. J. Bauer, G. R. Carignan, J. E. Demick, R. L. Frost, D. Gautier, J. A. Haberman, D. N. Harpold, D. M. Hunten, G. Israel, J. I. Lunine, W. T. Kasprzak, T. C. Owen, M. Paulkovich, F. Raulin, E. Raaen, S. H. Way, *Nature* **2005**, *438*, 779.
[152] P. Thaddeus, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2006**, *361*, 1681-1687.
[153] A. D. Keefe, S. L. Miller, *Orig. Life Evol. Biosph.* **1996**, *26*, 111-129.
[154] H. J. Cleaves, J. H. Chalmers, A. Lazcano, S. L. Miller, J. L. Bada, *Orig. Life Evol. Biosph.* **2008**, *38*, 105-115.
[155] M. Ferus, P. Kubelík, A. Knížek, A. Pastorek, J. Sutherland, S. Civiš, *Sci. Rep.* **2017**, *7*, 6275.
[156] S. W. Fox, K. Harada, *Science* **1961**, *133*, 1923-1924.
[157] J. P. Ferris, O. S. Zamek, A. M. Altbuch, H. Freiman, *J. Mol. Evol.* **1974**, *3*, 301-309.
[158] R. Saladino, U. Ciambecchini, C. Crestini, G. Costanzo, R. Negri, E. Di Mauro, *ChemBioChem* **2003**, *4*, 514-521.
[159] C. Menor-Salván, D. M. Ruiz-Bermejo, M. I. Guzmán, S. Osuna-Esteban, S. Veintemillas-Verdaguer, *Chem. Eur. J.* **2009**, *15*, 4411-4418.
[160] M. P. Robertson, S. L. Miller, *Nature* **1995**, *375*, 772.
[161] J. P. Ferris, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1968**, *33*, 693-704.
[162] R. Saladino, C. Crestini, G. Costanzo, R. Negri, E. Di Mauro, *Bioorg. Med. Chem.* **2001**, *9*, 1249-1253.
[163] R. Saladino, C. Crestini, U. Ciambecchini, F. Ciciriello, G. Costanzo, E. Di Mauro, *ChemBioChem* **2004**, *5*, 1558-1566.
[164] S. D. Senanayake, H. Idriss, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1194-1198.
[165] R. Saladino, V. Neri, C. Crestini, G. Costanzo, M. Graciotti, E. Di Mauro, *J. Mol. Evol.* **2010**, *71*, 100-110.
[166] M. Ferus, D. Nesvorný, J. Šponer, P. Kubelík, R. Michalčíková, V. Shestivská, J. E. Šponer, S. Civiš, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 657-662.
[167] T. Wilhelm, *Chem. Ber.* **1900**, *33*, 3035-3056.
[168] R. Breslow, *Tetrahedron Lett.* **1959**, *1*, 22-26.
[169] S. Pallmann, J. Šteflová, M. Haas, S. Lamour, A. Henß, O. Trapp, *New J. Phys.* **2018**, *20*, 055003.
[170] C. Reid, L. E. Orgel, *Nature* **1967**, *216*, 455.
[171] N. W. Gabel, C. Ponnamperuma, *Nature* **1967**, *216*, 453.
[172] H.-J. Kim, A. Ricardo, H. I. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye, S. A. Benner, *J. Am. Chem. Soc.* **2011**, *133*, 9457-9468.
[173] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner, A. Eschenmoser, G. Ohlofjgewidmet, *Helv. Chim. Acta* **1990**, *73*, 1410-1468.
[174] R. Krishnamurthy, G. Arrhenius, A. Eschenmoser, *Orig. Life Evol. Biosph.* **1999**, *29*, 333-354.
[175] R. Krishnamurthy, S. Guntha, A. Eschenmoser, *Angew. Chem. Int. Ed.* **2000**, *39*, 2281-2285.
[176] P. M. Gardner, K. Winzer, B. G. Davis, *Nat. Chem.* **2009**, *1*, 377.
[177] K. Usami, A. Okamoto, *Org. Biomol. Chem.* **2017**, *15*, 8888-8893.
[178] G. Harsch, H. Bauer, W. Voelter, *Liebigs Ann.* **1984**, *1984*, 623-635.
[179] S. Islam, D.-K. Bučar, M. W. Powner, *Nat. Chem.* **2017**, *9*, 584.
[180] J. Kofoed, J.-L. Reymond, T. Darbre, *Org. Biomol. Chem.* **2005**, *3*, 1850-1855.
[181] C. Meinert, I. Myrgorodska, P. de Marcellus, T. Buhse, L. Nahon, S. V. Hoffmann, L. L. S. d'Hendecourt, U. J. Meierhenrich, *Science* **2016**, *352*, 208-212.
[182] G. M. Muñoz Caro, U. J. Meierhenrich, W. A. Schutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W. H. P. Thiemann, A. Brack, J. M. Greenberg, *Nature* **2002**, *416*, 403.
[183] R. M. Hazen, D. A. Sverjensky, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002162.
[184] P. Yin, Z.-M. Zhang, H. Lv, T. Li, F. Haso, L. Hu, B. Zhang, J. Bacsa, Y. Wei, Y. Gao, Y. Hou, Y.-G. Li, C. L. Hill, E.-B. Wang, T. Liu, *Nat. Commun.* **2015**, *6*, 6475.
[185] D. G. Blackmond, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002147.
[186] L. Mascia, T. Cotrufo, M. Cappiello, P. L. Ipata, *Biochim. Biophys. Acta* **1999**, *1472*, 93-98.
[187] M. Camici, F. Sgarrella, P. L. Ipata, U. Mura, *Arch. Biochem. Biophys.* **1980**, *205*, 191-197.
[188] I. Nam, J. K. Lee, H. G. Nam, R. N. Zare, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12396-12400.
[189] M. Akouche, M. Jaber, M.-C. Maurel, J.-F. Lambert, T. Georgelin, *Angew. Chem. Int. Ed.* **2017**, *56*, 7920-7923.
[190] I. Nam, H. G. Nam, R. N. Zare, *Proc. Natl. Acad. Sci. USA* **2017**.
[191] H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11315-11320.

[192] H. D. Bean, Y. Sheng, J. P. Collins, F. A. L. Anet, J. Leszczynski, N. V. Hud, *J. Am. Chem. Soc.* **2007**, *129*, 9556-9557.
[193] H.-J. Kim, S. A. Benner, *Chem. Eur. J.* **2018**, *24*, 581-584.
[194] M. C. Chen, B. J. Cafferty, I. Mamajanov, I. Gállego, J. Khanam, R. Krishnamurthy, N. V. Hud, *J. Am. Chem. Soc.* **2014**, *136*, 5640-5646.
[195] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat. Commun.* **2016**, *7*, 11328.
[196] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239.
[197] J. D. Sutherland, *Angew. Chem. Int. Ed.* **2016**, *55*, 104-121.
[198] A.-A. Ingar, R. W. A. Luke, B. R. Hayter, J. D. Sutherland, *ChemBioChem* **2003**, *4*, 504-507.
[199] J. Xu, M. Tsanakopoulou, C. J. Magnani, R. Szabla, J. E. Šponer, J. Šponer, R. W. Góra, J. D. Sutherland, *Nat. Chem.* **2016**, *9*, 303.
[200] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531-543.
[201] C. Fernandez-Garcia, N. M. Grefenstette, M. W. Powner, *Chem. Commun.* **2018**.
[202] S. Stairs, A. Nikmal, D.-K. Bučar, S.-L. Zheng, J. W. Szostak, M. W. Powner, *Nat. Commun.* **2017**, *8*, 15270.
[203] M. W. Powner, J. D. Sutherland, J. W. Szostak, *J. Am. Chem. Soc.* **2010**, *132*, 16677-16688.
[204] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei, S. A. Benner, *Angew. Chem. Int. Ed.* **2016**, *55*, 15816-15820.
[205] B. Burcar, M. Pasek, M. Gull, B. J. Cafferty, F. Velasco, N. V. Hud, C. Menor-Salván, *Angew. Chem. Int. Ed.* **2016**, *55*, 13249-13253.
[206] C. Gibard, S. Bhowmik, M. Karki, E.-K. Kim, R. Krishnamurthy, *Nat. Chem.* **2017**, *10*, 212.
[207] R. Saladino, B. M. Bizzarri, L. Botta, J. Šponer, J. E. Šponer, T. Georgelin, M. Jaber, B. Rigaud, M. Kapralov, G. N. Timoshenko, A. Rozanov, E. Krasavin, A. M. Timperio, E. D. Mauro, *Sci. Rep.* **2017**, *7*, 14709.
[208] A. C. Rios, Y. Tor, *Isr. J. Chem.* **2013**, *53*, 469-483.
[209] A. S. Tupper, K. Shi, P. G. Higgs, *Life* **2017**, *7*, 41.
[210] M.-C. H. Maurel, Anne-Lise, in *Lectures in Astrobiology, Vol. 1* (Ed.: M. B. Gargaud, B.; Martin, H.; Reisse, J.), Springer-Verlag, **2005**, pp. 571-594.
[211] R. Krishnamurthy, *Isr. J. Chem.* **2015**, *55*, 837-850.
[212] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
[213] S. A. Benner, A. D. Ellington, *Nature* **1987**, *329*, 295.
[214] C. Woese, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6854-6859.
[215] G. E. Fox, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a003483.
[216] N. V. Hud, *Synlett* **2017**, *28*, 36-55.
[217] G. F. Joyce, *Nature* **2002**, *418*, 214.
[218] A. E. Engelhart, N. V. Hud, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.
[219] D. M. Fialho, K. C. Clarke, M. K. Moore, G. B. Schuster, R. Krishnamurthy, N. V. Hud, *Org. Biomol. Chem.* **2018**, *16*, 1263-1271.
[220] E.-K. Kim, V. Martin, R. Krishnamurthy, *Chem. Eur. J.* **2017**, *23*, 12668-12675.
[221] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.*, *57*, 5943-5946.
[222] P. P. Seelam, P. Sharma, A. Mitra, *RNA* **2017**, *23*, 847-859.
[223] C. Roost, S. R. Lynch, P. J. Batista, K. Qu, H. Y. Chang, E. T. Kool, *J. Am. Chem. Soc.* **2015**, *137*, 2107-2115.
[224] M. Chawla, R. Oliva, J. M. Bujnicki, L. Cavallo, *Nucleic Acids Res.* **2015**, *43*, 6714-6729.
[225] R. Oliva, L. Cavallo, A. Tramontano, *Nucleic Acids Res.* **2006**, *34*, 865-879.
[226] J. P. Rife, C. S. Cheng, P. B. Moore, S. A. Strobel, *Nucleic Acids Res.* **1998**, *26*, 3640-3644.
[227] A. R. Cervi, A. Guy, G. A. Leonard, R. Téoule, W. N. Hunter, *Nucleic Acids Res.* **1993**, *21*, 5623-5629.
[228] M. Kroger, B. Singer, *Biochemistry* **1979**, *18*, 91-95.
[229] B. D. Heuberger, A. Pal, F. Del Frate, V. V. Topkar, J. W. Szostak, *J. Am. Chem. Soc.* **2015**, *137*, 2769-2775.
[230] K. Grzeskowiak, T. R. Webb, L. E. Orgel, *J. Mol. Evol.* **1984**, *21*, 81-83.
[231] N. Prywes, J. C. Blain, F. Del Frate, J. W. Szostak, *eLife* **2016**, *5*, e17756.
[232] J. W. Szostak, *J. Sys. Chem.* **2012**, *3*, 2.
[233] M. D. Kirnos, I. Y. Khudyakov, N. I. Alexandrushkina, B. F. Vanyushin, *Nature* **1977**, *270*, 369.
[234] P. M. M. Rae, R. E. Steele, *BioSystems* **1978**, *10*, 37-53.
[235] J. H. Gommers-Ampt, P. Borst, *FASEB J.* **1995**, *9*, 1034-1042.
[236] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322-8323.

[237] J. A. Piccirilli, S. A. Benner, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33.
[238] W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, *Science* **2001**, *292*, 1319-1325.
[239] A. Wochner, J. Attwater, A. Coulson, P. Holliger, *Science* **2011**, *332*.
[240] J. Attwater, A. Wochner, P. Holliger, *Nat. Chem.* **2013**, *5*, 1011.
[241] J. S. Reader, G. F. Joyce, *Nature* **2002**, *420*, 841.
[242] T. M. Tarasow, S. L. Tarasow, B. E. Eaton, *Nature* **1997**, *389*, 54.
[243] G. F. Joyce, *Angew. Chem. Int. Ed.* **2007**, *46*, 6420-6436.
[244] C. Wilson, J. W. Szostak, *Nature* **1995**, *374*, 777.
[245] P. J. Unrau, D. P. Bartel, *Nature* **1998**, *395*, 260.
[246] E. Ekland, J. Szostak, D. Bartel, *Science* **1995**, *269*, 364-370.
[247] L. L. Martin, P. J. Unrau, U. F. Müller, *Life* **2015**, *5*, 247-268.
[248] M.-C. Maurel, J. Ninio, *Biochimie* **1987**, *69*, 551-553.
[249] M. Robertson, S. Miller, *Science* **1995**, *268*, 702-705.
[250] M. Levy, S. L. Miller, *J. Mol. Evol.* **1999**, *48*, 631-637.
[251] J. A. Martínez Giménez, G. T. Sáez, R. T. Seisdedos, *J. Theor. Biol.* **1998**, *194*, 485-490.
[252] V. Anantharaman, E. V. Koonin, L. Aravind, *Nucleic Acids Res.* **2002**, *30*, 1427-1464.
[253] A. Danchin, G. Fang, S. Noria, *Proteomics* **2007**, *7*, 875-889.
[254] J. A. Doudna, J. R. Lorsch, *Nat. Struct. Mol. Biol.* **2005**, *12*, 395.
[255] J. R. Knowles, *Nature* **1991**, *350*, 121.
[256] C. R. Woese, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8742-8747.
[257] A. S. Petrov, B. Gulen, A. M. Norris, N. A. Kovacs, C. R. Bernier, K. A. Lanier, G. E. Fox, S. C. Harvey, R. M. Wartell, N. V. Hud, L. D. Williams, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15396-15401.
[258] C. Davidovich, M. Belousoff, A. Bashan, A. Yonath, *Res. Microbiol.* **2009**, *160*, 487-492.
[259] G. R. Björk, K. Jacobsson, K. Nilsson, M. J. O. Johansson, A. S. Byström, O. P. Persson, *EMBO J.* **2001**, *20*, 231-239.
[260] C. Deutsch, B. El Yacoubi, V. de Crécy-Lagard, D. Iwata-Reuyl, *J. Biol. Chem.* **2012**, *287*, 13666-13673.
[261] C. Yarian, H. Townsend, W. Czestkowski, E. Sochacka, A. J. Malkiewicz, R. Guenther, A. Miskiewicz, P. F. Agris, *J. Biol. Chem.* **2002**, *277*, 16391-16395.
[262] P. F. Agris, F. A. P. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1-13.
[263] K. H. J. Gordon, *J. Theor. Biol.* **1995**, *173*, 179-193.
[264] J. H. Campbell, *J. Mol. Evol.* **1991**, *32*, 3-5.
[265] H. F. Noller, *Nature* **1991**, *353*, 302.
[266] N. Maizels, A. M. Weiner, *Cold Spring Harb. Symp. Quant. Biol.* **1987**, *52*, 743-749.
[267] J. D. Engel, P. H. Von Hippel, *Biochemistry* **1974**, *13*, 4143-4158.
[268] W. A. Decatur, M. J. Fournier, *Trends Biochem. Sci.* **2002**, *27*, 344-351.
[269] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934-4944.
[270] M. Helm, *Nucleic Acids Res.* **2006**, *34*, 721-733.