Dissertation zur Erlangung des Doktorgrades

Der Fakultät für Chemie und Pharmazie

Der Ludwig-Maximilians-Universität München

Untersuchungen von Proto-Nukleotiden und Aminosäure-Nukleotiden zum chemischen Ursprung des Lebens

Felix Xu

aus

Henstedt-Ulzburg, Deutschland

2024

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 02.05.2024

……

Felix Xu

Dissertation eingereicht am	24.05.2024
1. Gutachter	Prof. Dr. Thomas Carell
2. Gutachter	Dr. Pavel Kielkowski
Mündliche Prüfung am	06.08.2024

Publikationsliste

Im Rahmen dieser Promotion erstellte Publikationen:

- F. Müller†, L. Escobar†, **F. Xu**, E. Węgrzyn, M. Nainytė, T. Amatov, C.Y. Chan, A. Pichler, T. Carell, A prebiotically plausible scenario of an RNA-peptide world, *Nature* **2022**, 605, 279–284.

- **F. Xu**, A. Crisp, T. Schinkel, R. C. A. Dubini, S. Hübner, S. Becker, F. Schelter, P. Rovó, T. Carell, Isoxazole Nucleosides as Building Blocks for a Plausible Proto-RNA, *Angew. Chem. Int. Ed.* **2022**, *61*, e202211945.

- **F. Xu**†, S. Wiedemann†, J. Feldmann, S. Becker, T. Carell, An Aminoisoxazole-Based Proto-RNA, *ChemistryEurope* **2023**, e202300057.

† Die Autoren haben zu gleichen Teilen zur Publikation beigetragen.

Konferenzbeiträge

08/2021	Molecular Origins of Life, Online, Posterpräsentation
11/2022	Bioinspired Complex Systems from Basic Science to Practical Applications, Neve Ilan, Israel, Posterpräsentation
06/2023	Molecular Origins of Life Munich, Online, Posterpräsentation
10/2023	Origins of Life Donostia Meeting, San Sebastian, Spanien, Posterpräsentation

Weitere Publikationen

- R. Bhimpuria†, K. R. Baddigam†, **F. Xu**, J. A. L. Wells, K. E. Borbas, Synthesis of styrene, phenyl acrylate and allyloxyphenyl-functionalized porphyrins, and preliminary exploration of their utility for assembling multiple porphyrin units, *JPP* **2023**, 27:07n10, 1398-140.

Danksagung

An erster Stelle möchte ich meinem Doktorvater *Prof. Dr. Thomas Carell* danken. Danke, dass du mich in deinen Arbeitskreis aufgenommen und mir ein spannendes und herausforderndes Thema zur Verfügung gestellt hast. Besonders danken möchte ich dir für die immensen Freiheiten und das damit verbundene Vertrauen.

Dr. Pavel Kielkowski danke ich für die Übernahme des Zweitgutachtens. Ebenso bedanke ich mich bei allen Mitgliedern der Prüfungskommission für das Beisitzen der mündlichen Prüfung.

Dr. Markus Müller danke ich für die vielen wissenschaftlichen und nicht-wissenschaftlichen Diskussionen, die bedingungslose Hilfsbereitschaft in allen Belangen und die vielen kreativen Ratschläge.

Frau Gärtner und *Dr. Nada Raddaoui* danke ich mich für die Hilfe bei allen organisatorischen Aufgaben. Bei *Kerstin Kurz* und *Claudia Scherübel* bedanke ich mich für die Unterstützung bei technischen Aufgaben, wie der Bestellung von Chemikalien und der Wartung diverser Laborgeräte.

Für das gewissenhafte Korrekturlesen meiner Arbeit bedanke ich mich bei *Dr. Markus Müller*, *Dr. Stefan Wiedemann* und *Annika Tölke.*

Meinen Kollaborationspartnern *Dr. Romeo Dubini, Thea Schinkel, Dr. Petra Rovó* und *Manuel Loos* danke ich für die tolle Zusammenarbeit und die vielen inspirierenden Diskussionen.

Ich bedanke mich bei dem allen ehemaligen und aktuellen Mitgliedern des Arbeitskreises für die unvergessliche Zeit und die großartige Atmosphäre. Ich hatte das Glück meine Promotionszeit mit vielen wundervollen Menschen verbringen zu dürfen, die mir sehr ans Herz gewachsen sind. Ihr seid der Grund, warum ich mich jeden Tag darauf gefreut habe an die Arbeit zu fahren. In diesem Zusammenhang bedanke ich mich bei: *Dr. Giacomo Ganazzoli, Fabian Hernichel, Andreas Reichl, Dr. Dilara Özdemir, Katerina Pappa, Alexander Pichler, Dr. Felix Müller, Dr. Jonas Feldmann, Dr. Stefan Wiedemann, Dr. Eva Korytiaková, Dr. Luis Montiel* und *Annika Tölke.*

Der größte Dank geht an meine Eltern und meinen Bruder für ihren bedingungslosen Rückhalt über all die Jahre meines Studiums und der Doktorarbeit. Ich danke euch die moralische und finanzielle Unterstützung

Inhaltsverzeichnis

Abstract

The RNA world hypothesis postulates that RNA, a versatile molecule capable of both storing information (genotype) and catalyzing biochemical reactions (phenotype), played a central role in the early evolution of life. In the RNA world scenario, self-replicating RNAs are believed to have developed from the primordial environment and developed increasingly complex chemical structures through evolutionary processes. The revelation that ribosomal peptide synthesis is catalyzed by ribozymes further strengthened the acceptance of this theory. However, comprehending the abiotic synthesis of RNA nucleotides continues to be challenge. The lack of evidence concerning the physical, chemical, and locational boundary conditions of this process compels us to 're-invent' the origins of life. To this end, numerous prebiotically plausible synthetic routes to RNA nucleosides and nucleotides have been recently proposed. Nevertheless, the ideal functionality and structure of RNA suggests that RNA itself is a product of evolution rather than a product of abiotic processes. It likely developed from a precursor (proto-RNA) that was more readily accessible in a prebiotic world.

In this work, we propose, that *N*-isoxazolyl-urea and formamidopyrimidine (FaPy) are plausible precursors of the contemporary base pairing system consisting of the *Watson-Crick* pairs: A-U and G-C. Recently it was demonstrated, that *N*-isoxazolyl-ureas and formamidopyrimidines could have played crucial roles as intermediates in the prebiotic synthesis of RNA nucleosides. Their ability to transform to the canonical nucleosides via intramolecular reactions gives us a compelling mechanism to explain the evolutionary transition of an proto-RNA world to an RNA world. To this end, synthetic routes to the corresponding phosphoramidite building blocks were developed. The building blocks were subsequently incorporated into RNA to investigate the base pairing properties of *N*-isoxazolyl-ureas. In addition, we demonstrated that the stereochemical configuration of the anomeric center formed by the intramolecular reaction cascade of *N*-isoxazolyl-urea-nucleotides to U and C is highly dependent on the configuration of the neighboring bases. Furthermore, an alternative scenario in which non-genetic-polymers formed first and then transitioned to proto-RNAs was investigated. Encouraging preliminary results are reported, demonstrating that under prebiotic conditions formamidopyrimidines can be regioselectively attached to abasic sites of oligonucleotides via glycosylation.

In the second part of this thesis a primitive form of RNA-based peptide synthesis was investigated, employing amino-acid modified, non-canonical nucleosides, regarded as relics of the RNA world. The transition from the RNA world to a system where proteins became the primary catalysts, as observed in contemporary life, remains one of the major mysteries

regarding the origin of life. Here we show a system, where duplexes of oligonucleotides carrying amino acid modified *N*^{6}-carbamoyl adenine (m^{6}aa^{6}A) and (methyl)aminomethyl uridine ((m)nm^{5}U) can undergo peptide bond formation to produce hairpin-like structures. Cleavage of the carbamoyl bond would furnish RNA with a peptide connected to a (m)nm^{5}U, enabling us to effectively grow or transfer peptides on RNA. These results provide evidence for the concept proposing the early existence of an RNA-peptide world.

Zusammenfassung

Die RNA-Welt-Hypothese geht davon aus, dass RNA eine zentrale Rolle in der frühen Evolution des Lebens spielte. Diese Annahme basiert auf der Vielseitigkeit von RNA als ein Molekül, das sowohl in der Lage ist Informationen zu speichern (Genotyp) als auch biochemische Reaktionen zu katalysieren (Phänotyp). Im Modell der RNA-Welt wird angenommen, dass zuerst selbstreplizierende RNA auf der frühen Erde entstanden ist und sich daraus über evolutionäre Prozesse immer komplexere chemische Strukturen entwickelt haben. Die Entdeckung, dass die ribosomale Peptidsynthese durch Ribozyme katalysiert wird, hat die Akzeptanz dieser Theorie weiter gestärkt. Allerdings bleibt die fehlende Erklärung der RNA-Welt-Hypothese über die abiotische Synthese von RNA-Nukleotiden eine fortwährende Herausforderung. Der Mangel an Beweisen für die physikalischen, chemischen und räumlichen Randbedingungen dieses Prozesses zwingt uns dazu, den Ursprung des Lebens "neu zu erfinden". Diesbezüglich wurden bereits zahlreiche präbiotisch plausible Syntheserouten zu RNA-Nukleosiden und -Nukleotiden postuliert. Die ideale Funktionalität und Struktur der RNA legen jedoch nahe, dass die RNA selbst ein Produkt der Evolution und nicht ein Produkt abiotischer Prozesse ist. Es ist wahrscheinlich, dass sich die RNA aus einem Vorläufer (Proto-RNA) entwickelt hat, der auf der frühen Erde leichter zugänglich war.

Im Rahmen dieser Arbeit untersuchten wir ein Szenario, in dem *N*-Isoxazolyl-Harnstoff und Formamidopyrimidine (FaPy) plausible Vorläufer des heutigen Basenpaarsystems gewesen sein könnten, das aus den *Watson-Crick*-Paaren A-U und G-C besteht. Es konnte bereits gezeigt werden, dass *N*-Isoxazolyl-Harnstoff und Formamidopyrimidine eine entscheidende Rolle als Zwischenstufen bei der präbiotischen Synthese von RNA-Nukleosiden gespielt haben könnten. Die Möglichkeit, diese Moleküle durch intramolekulare Reaktionen in die kanonischen Nukleoside umzuwandeln, liefert einen überzeugenden Mechanismus zur Erklärung des evolutionären Übergangs von einer Proto-RNA-Welt zu einer RNA-Welt. Zu diesem Zweck wurden synthetische Routen zu den entsprechenden Phosphoramidit-Bausteinen entwickelt. Diese wurde anschließend in RNA eingebaut, um die Basenpaarungseigenschaften von *N*-Isoxazolyl-Harnstoffen zu untersuchen. Außerdem wurde gezeigt, dass die stereochemische Konfiguration des anomeren Zentrums, das durch die intramolekulare Reaktionskaskade von *N*-Isoxazolyl-Harnstoff-Nukleotiden zu U und C geformt wird, stark von der Konfiguration der benachbarten Basen abhängt. Darüber hinaus wurde ein alternatives Szenario untersucht, in dem zunächst nicht-genetische Polymere gebildet wurden, die dann in Proto-RNAs übergehen. Es wurden vielversprechende vorläufige Ergebnisse erzielt, die zeigen, dass unter präbiotischen Bedingungen Formamidopyrimidine regioselektiv an abasischen Stellen von Oligonukleotiden über Glykosylierungen gekoppelt werden können.

Im zweiten Teil dieser Arbeit wurde eine primitive Form der RNA-basierten Peptidsynthese untersucht, bei der Aminosäure-modifizierte, nicht-kanonische Nukleoside, die als Relikte der RNA-Welt gelten, verwendet werden. Der Übergang von der RNA-Welt zu einem System, in dem Proteine die Funktion als primäre Katalysatoren übernommen haben, wie es im heutigen Leben zu beobachten ist, bleibt eines der größten Rätsel in Bezug auf den Ursprung des Lebens. In dieser Arbeit zeige ich, dass ein System, in dem Duplexe von Oligonukleotiden, die Aminosäure-modifiziertes *N*^{6}-Carbamoyladenin (m^{6}aa^{6}A) und (Methyl)aminomethyluridin ((m)nm^{5}U) tragen, Peptidbindungen eingehen können, um haarnadelartige Strukturen zu bilden. Durch die Spaltung der Carbamoylbindung wird RNA mit einem an (m)nm^{5}U gebunden Peptid erhalten. Dies eröffnet uns die Möglichkeit, Peptide auf RNA effektiv wachsen zu lassen oder zu übertragen. Diese Ergebnisse liefern Beweise für die Idee der frühen Existenz einer RNA-Peptid-Welt.

1. Einleitung

1.1. Die Entstehung des Lebens

Die Frage nach der Entstehung des Lebens ist eine der grundlegendsten und zugleich am wenigsten verstandenen Fragen der Menschheit. Die Entwicklung einer intellektuell fortgeschrittenen Spezies aus lebloser Materie, die das Bewusstsein erlangt hat, um über dieses Rätsel nachzudenken, stellt ein Phänomen dar, welches die Begeisterung von Wissenschaftlern aus allen Fachrichtungen weckt. Im Laufe der Menschheitsgeschichte wurden zahlreiche wissenschaftliche, philosophische und religiöse Ansätze verfolgt, um die Entstehung des Lebens aus unbelebter Materie zu erklären.

1.2. Was ist Leben?

Bevor wir uns mit der Frage nach der Entstehung beschäftigen, muss eine weitere entscheidende Frage beantwortet werden: *Was ist Leben*? Diese Frage wurde bereits von *Erwin Schrödinger* 1944 mit seinem gleichnamigen Buch gestellt.[1] Ein vereinheitlichtes Merkmal für alles Leben zu finden, wirkt, aufgrund der enormen Artenvielfalt der Erde, wie ein aussichtloses Unterfangen. Allerdings offenbaren uns eine Reihe einfacher Beobachtungen eine bemerkenswerte Ähnlichkeit. Auf molekularer und genetischer Ebene basiert fast alles Leben auf der Erde auf einen vererblichen Fluss von Information, welches von *Francis Crick* als das zentrale Dogma der molekularen Biologie beschrieben wurde (**Abbildung 1**).[2]

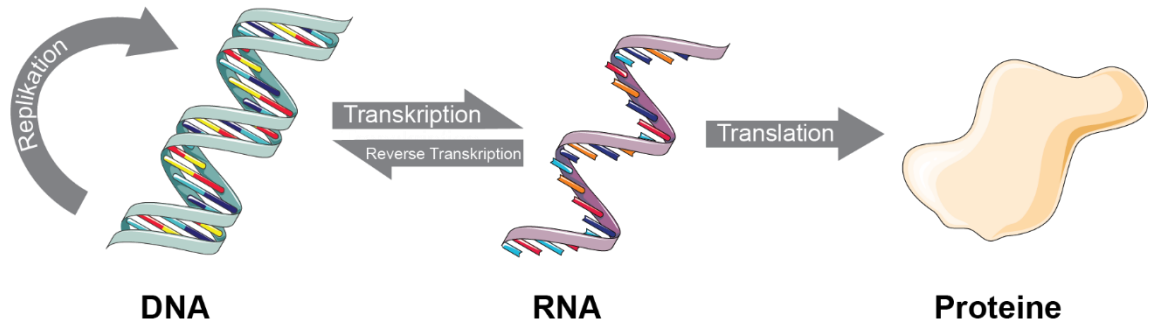


Abbildung 1. Das zentrale Dogma der molekularen Biologie.

Genetische Information wird in der doppelsträngigen Desoxyribonukleinsäure (DNA) gespeichert und durch den Prozess der Replikation vervielfältigt. Darüber hinaus kann die Information über Transkription in einzelsträngige Ribonukleinsäure (RNA) umgeschrieben werden. Diese wird schließlich im Prozess der Translation in Proteine umgewandelt. Da die

einzige Funktion der DNA darin besteht als Speicher der genetischen Information zu dienen, wird sie als genotypisch bezeichnet. Proteine hingegen besitzen strukturelle oder katalytische Funktionen und werden somit als phänotypisch bezeichnet. Die Universalität des zentralen Dogmas und des genetischen Codes deutet darauf hin, dass sich alles Leben aus einem gemeinsamen Vorfahren entwickelt hat.[3-5]

Trotz dieser Erkenntnisse existiert keine klare allgemein anerkannte Definition für den Begriff ‚*Leben*'.[6, 7] Durch einen NASA-Ausschuss wurde 1994 das Leben als „*ein sich selbst erhaltendes chemisches System, das zur darwinistischen Evolution fähig ist*" beschrieben.[8] Doch auch diese Definition ist aus vielerlei Gründen unzureichend. Zunächst wurde der Einfluss und der Informationsfluss der Umwelt vernachlässigt. Außerdem sind Zellen und mehrzellige Organismen nicht selbsterhaltend, sondern immer von externen Faktoren oder anderen Organismen abhängig, um ihr Überleben und ihre Fortpflanzung zu gewährleisten. Darüber hinaus fehlt bei dieser Definition der thermodynamische Aspekt des Lebens als ein System, das weit entfernt vom chemischen Gleichgewicht existiert.[9-11] Trotz dieser Kritiken wird die NASA-Definition von der Forschungsgemeinde der präbiotischen Chemie als ausreichend anerkannt, um die wichtigsten Aspekte des Lebens zu beschreiben.[12-16]

1.3. Präbiotische Chemie

Die Aufgabe der präbiotischen Chemie besteht darin, die Entstehung und die Selbstorganisation einfacher Moleküle zu erklären, welche die Bildung lebender Systeme (Abiogenese) ermöglichte. Bedauerlicherweise existieren heute kaum Hinweise über die planetarischen Bedingungen und verfügbaren Ausgangsmaterialien. Angesichts dieser begrenzten Informationen und zahlreicher weiterer Variablen bleibt uns nur eine hypothetische Rekonstruktion des Prozesses. Hierfür werden Reaktionen unter „präbiotisch plausiblen Bedingungen" durchgeführt, die denen der frühen Erde so nah wie möglich nachempfunden werden.[14] *Albert Eschenmoser* fasste dieses Problem treffend zusammen: „Die Entstehung des Lebens kann nicht ‚entdeckt', sondern muss ‚neu erfunden' werden."[17] Hierzu werden zwei Vorgehensweisen gleichzeitig verfolgt: der *Bottom-Up* Ansatz (geologischer Ansatz) und der *Top-Down* Ansatz (biologischer Ansatz).

1.3.1. *Top-down* Ansatz

Bei dem *Top-down* Ansatz werden phylogenetischen Studien verwendet, um alles Leben zu einem gemeinsamen Ausgangspunkt in der genetischen Geschichte zurückzuverfolgen. Wie bereits in **Abschnitt 1.2.** erwähnt, ist es sehr wahrscheinlich, dass alle drei Domänen des Lebens aus einem gemeinsamen Vorfahren entstanden sind. Dieser hypothetische

gemeinsame Vorfahre wird als *last universal common ancestor* (LUCA) bezeichnet und als Schnittstelle zwischen abiotischer Chemie und hochkomplexem Leben angesehen.[4, 18] *Weiss et al.* identifizierten Gene, die über alle Domänen des Lebens konserviert sind. Basierend darauf postulierten sie, dass LUCA anaerob, CO_2-fixierend, H_2-abhängig, N_2-fixierend und thermophil war. Dies legt nahe, dass LUCA bereits einen einfachen Metabolismus, sowie die Fähigkeit zur Replikation, Transkription und Translation besaß.[4]

Dieser Ansatz liefert uns Informationen über die essenziellen Funktionen des frühen Lebens und somit auch über die synthetischen Ziele der präbiotischen Chemie. Allerdings gibt er keine Aufschlüsse darüber, wie die benötigten Biomoleküle (Membranen, Proteine, RNA und DNA) selbst entstanden sind. Hierzu wird der *Bottom-up* Ansatz benötigt, um Reaktionen zu identifizieren, die unter den geochemischen Bedingungen zu diesen wesentlichen Komponenten geführt haben könnten.

1.3.2. *Bottom-up* Ansatz

Wir wissen heute, dass die Erde etwa 4.53 Milliarden Jahre (4.53 Ga) alt ist.[19] Zu Beginn vergrößerte sich die Erde zunehmend durch Kollisionen mit großen Projektilen und Protoplaneten.[20-22] Zu dieser Zeit, die auch als Hadaikum (4.6-4.0 Ga) bezeichnet wird, gilt die Erde als eine größtenteils geschmolzene, lebensfeindliche Masse mit ständigen vulkanischen Ausgasungen. Die Formation der Erdkruste wurde durch teils massive Einschläge gestört, die stark genug waren, um Ozeane verdampfen zu lassen, die Erdoberfläche zu sterilisieren und den Mond zu bilden.[20, 23, 24] Nach dem Hadaikum trat die Erde in das Archaikum (4.0-2.5 Ga) ein, in dem sich eine stabile Hydro- und Atmosphäre bilden konnte. Da bereits fossile Mikroorganismen beschrieben wurden, die bis zu 3.8 Milliarden Jahre alt sein könnten,[25] bliebe nach der Formation der Erde ein Zeitfenster von etwa einer Milliarde Jahre für die Entstehung des Lebens.

Es gibt kaum geologische Hinweise, die auf eine Vergletscherung während des Archaikums hindeuten, daher wird angenommen, dass die Atmosphäre im Archaikum in der Lage gewesen sein muss, genügend Wärme zu speichern, um die Oberflächentemperatur über 0 °C zu halten. Somit konnte flüssiges Wasser an der Oberfläche gewährleistet sein.[26] Geologische Kohlenstoffzyklusmodelle haben die wahrscheinliche Erdtemperatur im Archaikum auf 0-50 °C und den pH-Wert der Ozeane auf 5-7 ermittelt.[27] Die genaue Zusammensetzung der Atmosphäre zu dieser Zeit ist allerdings eines der größten Streitpunkte der präbiotischen Chemie. Die am weitesten verbreitete Annahme ist, dass die Atmosphäre entweder redox-neutral bzw. nur schwach reduzierend war. So wird angenommen, dass die Atmosphäre hauptsächlich aus einem Gemisch von CO_2, H_2O, N_2

und CO bestand, mit geringen Mengen an H_2, CH_4, SO_2 und H_2S.[28, 29] Zusätzlich müssen weitere mögliche organische Ausgangsmaterialien in Betracht gezogen werden. Informationen zu diesen können über verschiedene Wege gewonnen werden: Spektroskopische Untersuchungen des Interstellaren Raums[30, 31] und andere Himmelskörper unseres Sonnensystems[32-36] können Hinweise über die Beschaffenheit der Erde vor der Abiogenese liefern. Ebenso können massenspektrometrische Analysen von eingeschlagenen Meteoriten zeigen, welche im interstellaren Raum produzierten Verbindungen auch auf der frühen Erde verfügbar waren.[37-40] Zuletzt können über Entladungs- und UV-Bestrahlungs-Experimente die Wirkung von Blitzen und Sonnenstrahlung simuliert werden.[41-44]

Das wohl berühmteste Beispiel eines solchen Experimentes wurde von Stanley *Miller* und *Harold Urey* durchgeführt.[45] Basierend auf einer von *Aleksander Oparin* und *John Haldane* vorgeschlagenen reduzierenden Ur-Atomsphäre[46, 47] aus CH_4, NH_3, H_2 und H_2O konnten sie die Synthese von Aminosäuren unter abiotischen Bedingungen nachweisen.[45, 48] Später zeigte *Miller* über mechanistische Studien, dass sich zunächst Aldehyde und HCN bilden, welche anschließend zu Amino- und Hydroxynitrilen reagieren und in wässriger Phase zu Aminosäuren hydrolysieren. Dieser Vorgang verläuft analog zu der *Strecker* Synthese zur Darstellung von Aminosäuren.[49]

1.4. RNA-Welt Theorie

Replikation, Mutation und natürliche Selektion sind die grundlegenden Evolutionsfaktoren, die die Entwicklung eines komplexen biochemischen Systems von einem primitiven System ermöglichen. Wie bereits in **Abschnitt 1.2.** beschrieben, sind die zwei Grundfunktionen des Lebens auf zwei verschiedene Biopolymere, den Proteinen (Katalyse) und den Nukleinsäuren (Informationsspeicherung), aufgeteilt.

Die Frage, welches dieser beiden Makromoleküle zuerst entstanden ist, erinnert an ein Henne-Ei-Problem. *Woese*[50], *Crick*[51] und *Orgel*[52] spekulierten, dass die RNA zuerst entstanden sei. Ihre Begründung war, dass Nukleinsäuren im Gegensatz zu Proteinen einen plausiblen Mechanismus zur Selbstreplikation besitzen. Durch Watson-Crick-Paarung der Nukleobasen können Nukleinsäuren als Template für die Synthese komplementärer Stränge dienen. Darüber hinaus besitzt die RNA aufgrund der 2‘-Hydroxygruppen die Fähigkeit, komplexe tertiäre Strukturen in der einzelsträngigen Form auszubilden.[14, 50] Die Entdeckung von katalytisch aktiven RNA Molekülen (Ribozyme) in den 1980er Jahren durch *Cech* und *Altman* verstärkte das Interesse an dieser Theorie.[53, 54] Bei der Untersuchung des ribosomalen RNA Gens von *Tetrahymena thermophilia* entdeckte *Cech* ein Intron, das in der

Abwesenheit von Enzymen herausgespleißt werden konnte. Er schloss daraus, dass die Struktur der RNA-Sequenz diese Reaktion katalysiert.[53, 55] Unabhängig davon untersuchte *Altman* die katalytischen Eigenschaften des Enzyms Ribonuklease P, welches aus einer Protein- und einer RNA-Untereinheit besteht. Als die Proteinkomponente entfernt wurde, konnte er feststellen, dass die prä-tRNAs trotzdem in die aktiven tRNAs umgewandelt wurden.[54, 56, 57] Somit konnte erstmals gezeigt werden, dass ein einziges Biomolekül die beiden grundlegenden Funktionen des Lebens, den Genotyp (Informationsspeicher) und den Phänotyp (Katalysator), besitzen kann.

Diese Entdeckungen führten schließlich zur Formalisierung der RNA-Welt-Theorie durch *Walter Gilbert.*[58] Diese geht davon aus, dass genetisch kodierte Proteine nicht als Katalysatoren in der RNA-Welt beteiligt waren, die genetische Kontinuität in der ersten Phase der molekularen Evolution durch RNA-Replikation gesichert wurde und dass *Watson-Crick*-Basenpaarung eine entscheidende Rolle in diesem Replikationsprozess spielte.[59] Die RNA-Welt-Hypothese ermöglicht eine plausible Erklärung für den Übergang eines primitiven Systems zum heutigen zentralen Dogma. RNA-Moleküle könnten sich durch zufällige Mutation und Rekombination weiterentwickelt haben und schließlich verschiedene katalytische Funktionen erlangt haben. Diese Funktionen könnten anfangs unter anderem durch Koenzyme erweitert worden sein, welche als molekulare Fossile der RNA-Welt gelten.[60, 61] Im Laufe der Zeit erwarben diese Ribozyme durch die Evolution die Fähigkeit Proteine zu synthetisieren und die Enzymkatalyse setzte sich durch, wodurch Ribozyme langsam an Nutzen verloren. Ein Hinweis hierfür findet sich in der Struktur des Ribosoms, das keine Proteinmotive am aktiven Zentrum besitzt.[62]

Die RNA-Welt-Hypothese gilt nicht als unumstritten, trotzdem wird sie aktuell als die plausibelste Theorie zur Entstehung des Lebens angesehen.[59, 63] Es gilt jedoch zunächst zu klären, wie Nukleotide ohne Einfluss von Proteinen unter präbiotischen Bedingungen entstanden sein könnten. Im Anschluss daran stellen sich die Fragen, wie Nukleotide enzymfrei zu zufällige RNA Sequenzen polymerisieren konnten und wie diese ebenfalls enzymfrei repliziert werden konnten.[14] Ebenso sollte in Betracht gezogen werden, dass sich das erste genetische Polymer chemisch deutlich von der aktuellen RNA unterschied.[59]

1.5. Der Ursprung von RNA

RNA besteht aus Nukleosiden, die an der 3‘- und 5‘-Position über ein Phosphodiester-Rückgrat miteinander verknüpft sind. Die einzelnen Nukleoside setzten sich aus dem Zucker Ribose und den Purin-Nukleobasen Adenin (A) und Guanin (G) oder den Pyrimidin-Nukleobasen Cytosin (C) und Uracil (U) zusammen (**Abbildung 2**). Der präbiotische Zugang zu den Nukleosiden scheint aus retrosynthetischer Sicht offensichtlich zu sein: Die separate

Synthese der zwei einzelnen Bausteine, gefolgt von der Verknüpfung über eine Glykosylierungsreaktion.

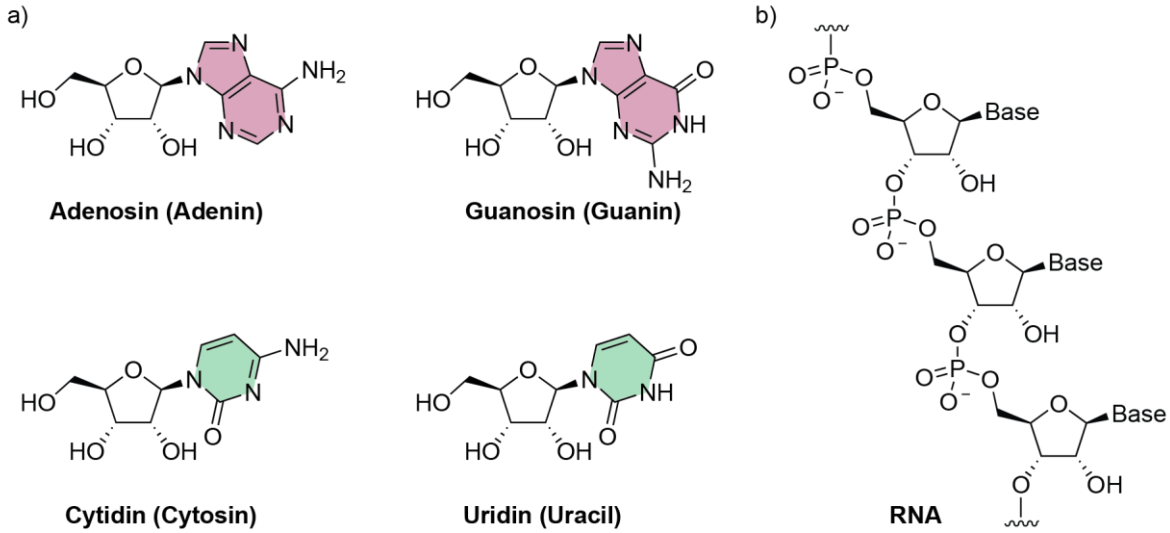
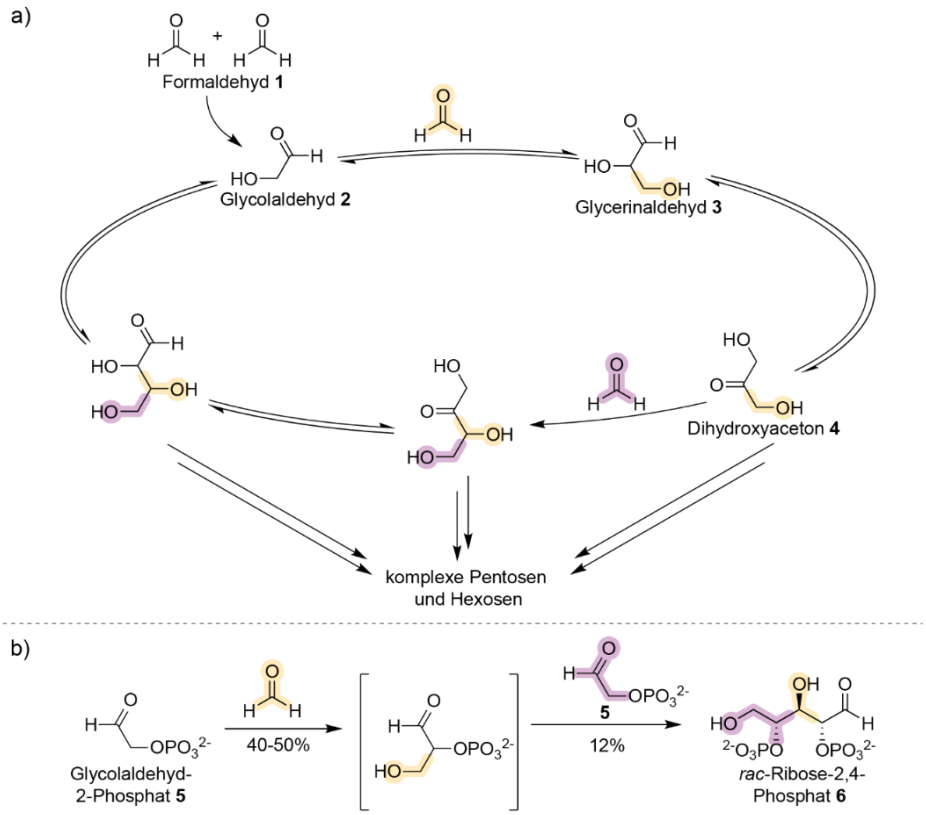


Abbildung 2. a) Strukturformeln der RNA Nukleoside Adenosin, Guanosin, Cytidin und Uridin, zusammengesetzt aus D-Ribose und den vier Nukleobasen Adenin (A), Guanin (G), Cytosol (C) und Uracil (U). b) Molekularer Aufbau von RNA.

Die 1861 von *Butlerov* entdeckte *Formosereaktion* wird als die wichtigste präbiotische Syntheseroute für Zucker angesehen. Ihm gelang die Herstellung eines komplexen Zuckergemisches (Tetrosen, Pentosen, Hexosen und ihre Abbauprodukte) durch die Polymerisation von Formaldehyd **1** unter basischen Bedingungen in Gegenwart von Calciumhydroxid (Schema 1).[64] Die Reaktion beginnt mit Formaldehyd **1**, welches durch Photolyse aus CO_2 und H_2O hergestellt werden kann.[65] Die initiale Dimerisierung von Formaldehyd **1** zu Glycolaldehyd **2** verläuft über einen noch unbekannten Mechanismus.[66] Im nächsten Schritt kann Glycolaldehyd **2** wiederum über eine Aldolkondensation mit Formaldehyd **1** zum C_3-Zucker Gylcerinaldehyd **3** reagieren. Dieses kann zu Dihydroxyaceton **4** isomerisieren und somit können anschließend über weitere Aldolkondensationen und Umlagerungen schrittweise komplexe lineare oder verzweigte Zucker aufgebaut werden.[67] Analysen von *Decker et al*. zeigten eine riesige Brandbreite an gebildeten Produkten. Aus präbiotischer Sicht ist dieser Mangel an Selektivität eine große Herausforderung. Aufgrund der geringen Stereo- und Regioselektivität liegt die Ausbeute von Ribose bei weniger als 1 %.[68, 69] Angesichts dieser Probleme wurden intensive Anstrengungen unternommen, um die Reaktion auf die Riboseproduktion auszurichten. Zum Beispiel konnten *Zubay et al.* zeigen, dass die Selektivität für Aldopentosen durch die Verwendung einer Suspension aus Magnesiumhydroxid und Blei(II)-Ionen erhöht wird.[70, 71] Ein weiterer Ansatz beruht auf der Fähigkeit von Boraten die cis-Diol-Strukturen von Zuckern

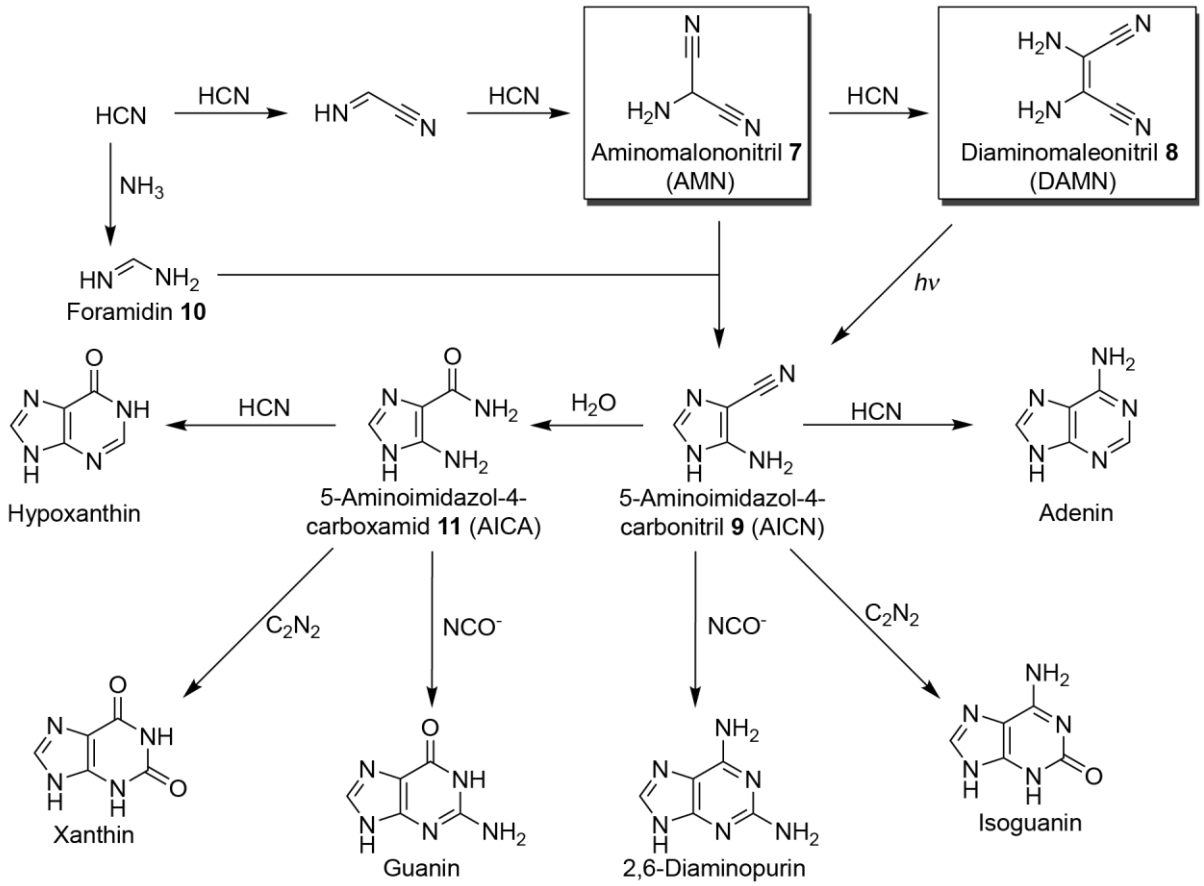
wie Ribose in der zyklischen Form zu komplexieren, wodurch die Enolisierung des Zuckers verhindert und die Stabilität erhöht wird.[72] *Benner et al.* nutzten dies aus, um mit Boratmineralien Ribose und andere Pentosen während der *Formosereaktion* anzureichern.[73-75] Ein Grund für die geringe Ribose-Selektivität der *Formosereaktion* ist die schnelle Isomerisierung von Glycerinaldehyd **3** zu Dihydroxyaceton **4**. *Eschenmoser et al.* konnten diese Isomerisierung durch die Verwendung von Glycolaldehyd-2-phosphat **5** verhindern, um racemisches Ribose-2,4-phosphat **6** als Hauptprodukt mit einer Ausbeute von 12 % zu erhalten (**Schema 1b**).[76, 77] Allerdings ist dieses Ribose-Derivat aus präbiotischer Sicht nicht von Bedeutung und Methoden, dieses zu den nützlicheren 3- oder 5-Phosphaten umzuwandeln, existieren bisher nicht. Die *Formosereaktion* wird in Hinsicht auf die selektive präbiotische Bildung von Ribose weiterhin als unzureichend angesehen. Trotzdem ist diese, aufgrund der Einfachheit C1-Moleküle in C5-Moleküle umzuwandeln, weiterhin von großer Bedeutung für das Forschungsfeld.[66]



Schema 1. a) Vorgeschlagener Mechanismus der Formosereaktion durch *Breslow*[67], ausgehend von Formaldehyd **1**, um komplexe lineare oder verzweigte Zucker aufzubauen. b) Selektive Synthese des Ribose-Derivates **6** ausgehend von Glycolaldehyd-2-Phosphat **5** nach *Eschenmoser et al.*[76]

Die Annahme, dass eine einfache Kondensation zwischen Zucker und Base das gewünschte Nukleosid ergeben würde, regte die Entwicklung vieler synthetischer Routen zu den Nukleobasen an. Die erste dieser Syntheserouten war die 1960 von *Oró* entwickelte Purin-

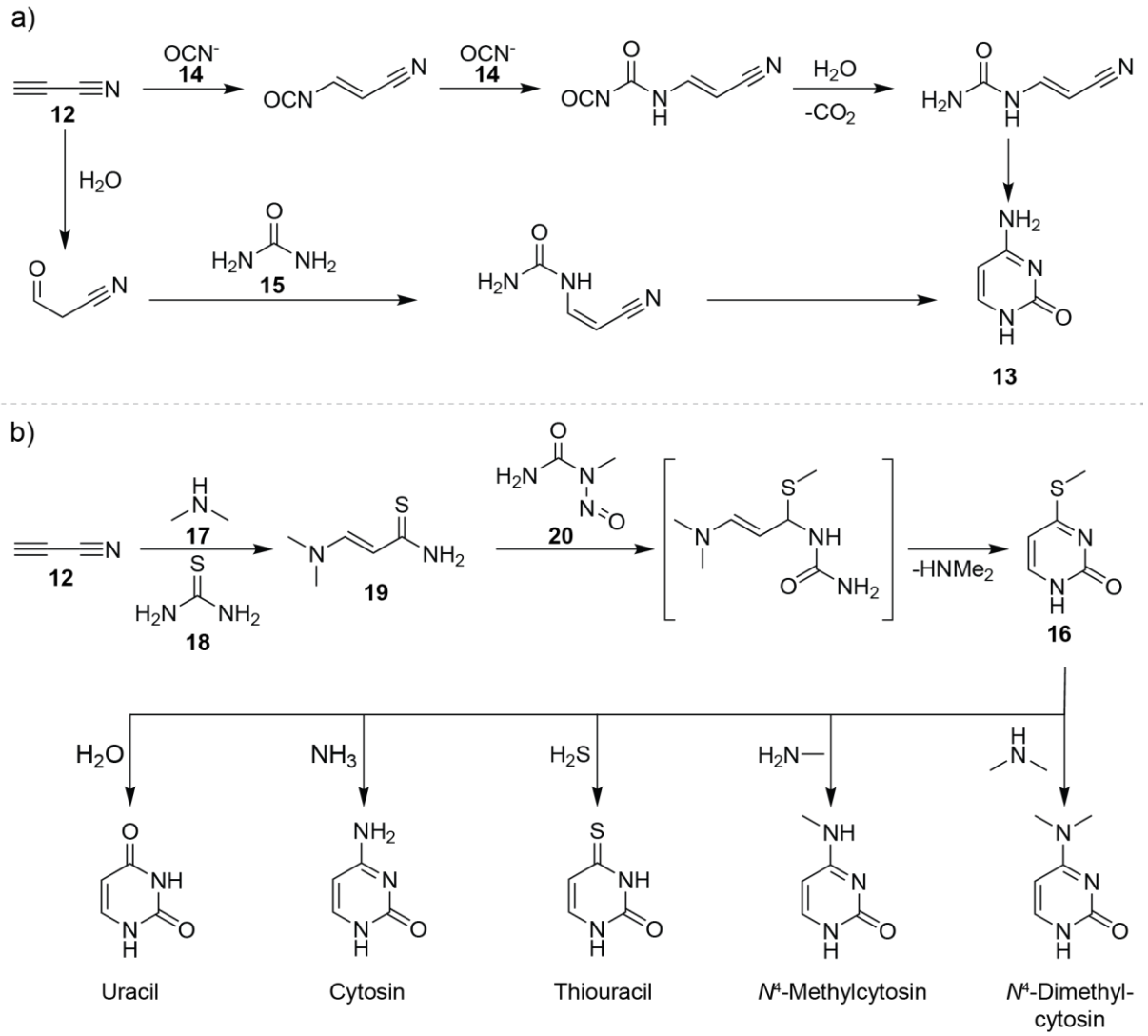
Synthese mittels HCN-Polymerisation. Durch das Erhitzen einer Ammoniumcyanid-Lösung auf 70 °C konnte er die Bildung von Adenin mit einer Ausbeute von 0.5 % nachweisen.[78] Diese Ausbeute konnte später durch die Zugabe von Glycolnitril[79] oder durch die Verwendung von HCN und flüssigem Ammoniak verbessert werden.[80] Mechanistische Studien durch *Ferris* und *Orgel* zeigten (**Schema 2**), dass das HCN-Trimer Aminomaleonitril **7** (AMN) und das Tetramer Diaminomaleonitril **8** (DAMN) wichtige Zwischenprodukte der Synthese sind. Die Bildung von 5-Aminoimidazol-4-carbonitril **9** (AICN) kann entweder über UV-Bestrahlung von DAMN **8** oder über die Reaktion von AMN mit Formamidin **10** verlaufen.[81] AICN **9** kann anschließend zu 5-Amino-4-imidazolcarboxamid **11** (AICA) hydrolysieren, beide Moleküle können schließlich durch die Reaktion mit verschiedenen Kohlenstoffquellen eine Reihe an Purinen bilden.[82-85]



Schema 2. Purinsynthese über HCN-Polymerisation nach *Oró*[78] und später modifiziert durch *Ferris* und *Orgel*.[81-85]

Pyrimidine können ausgehend von Cyanacetylen **12** dargestellt werden, welches durch Gasentladung in einem Methan- und Stickstoff-Gemisch, durch das Erhitzen von Acetylen in Stickstoff oder aus Ethin und Cyanid entstehen kann. *Orgel et al.* zeigten die Synthese von Cytosin **13** und dessen anschließende Hydrolyse zu Uracil durch die Reaktion von

Cyanacetylen mit Cyanat **14** oder Harnstoff **15** (**Schema 3a**).[41, 86, 87] Eine Alternative präbiotisch plausible Pyrimidinsynthese über das Zwischenprodukt *S*-Methylpyrimidinon **16** wurde von *Carell* und Mitarbeiter entwickelt (**Schema 3b**). Diese startet im ersten Schritt ebenfalls mit Cyanacetylen **12**, welches mit Dimethylamin **17** und Thiourea **18** zum Thioamid **19** reagiert. Dieses wird im nächsten Schritt mit *N*-Methyl-*N*-nitrosoharnstoff **20** zu *S*-Methylpyrimidinon **16** umgesetzt, welches mit unterschiedlichen Nukleophilen zu den jeweiligen Pyrimidin-Modifikationen reagieren kann.[88] Die präbiotische Relevanz einiger dieser Modifikationen wird in einem späteren Teil dieser Arbeit diskutiert (**Abschnitt 1.6**).

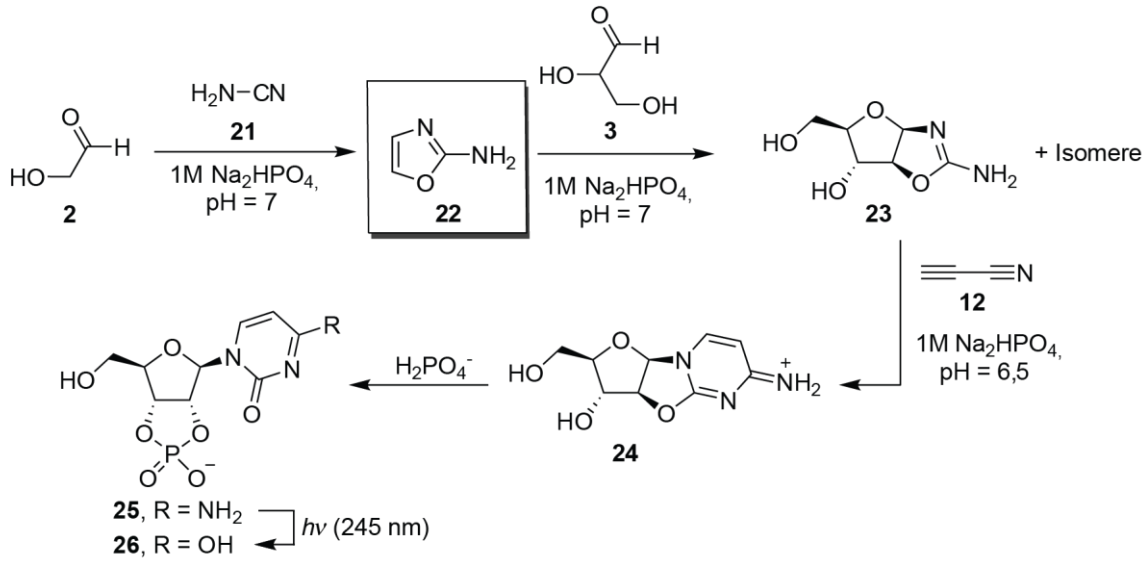


Schema 3. Präbiotisch plausible Pyrimidinsynthesen ausgehend von Cyanacetylen **12** nach a) *Orgel et al.*[86] und b) *Carell et al.*[88]

1.5.1. Präbiotische Synthese der Nukleoside

Es existieren überzeugende Beweise für die präbiotische Bildung der Nukleosid-Bausteine, insbesondere der Nukleobasen. Es bleibt die Frage wie sich aus diesen beiden Bausteinen

Nukleoside bilden konnten. Wie bereits oben beschrieben, wäre die einfachste Lösung die direkte Glykosylierung von Ribose mit der jeweiligen Nukleobase. In den ersten Experimenten durch *Fuller et al.*, die D-Ribose mit Purinen in Anwesenheit von einem Gemisch aus Meerwasser-Salzen erhitzt haben, konnten allerdings nur geringe Ausbeuten der erwünschten kanonischen β-Ribofuranose-Nukleoside erhalten werden (4 % für β-D-Adenosin und 9 % für β-D-Guanosin).[89, 90] Chemisch betrachtet ist dies nicht überraschend. Ribose liegt in wässriger Lösung hauptsächlich in der Pyranoseform vor, während das benötigte α-Furanose Isomer nur 7% des Isomerengemisches ausmacht.[91] Darüber hinaus ist die fehlende Regioselektivität und die geringe Nukleophilie des N^9-Stickstoffes eine weitere Herausforderung. Dies führt dazu, dass bevorzugt die exozyklische N^6-Amino-Gruppe reagiert. Die gleichen chemischen Herausforderungen weisen auch die Pyrimidine auf. Hinzu kommt hier, dass das freie Elektronenpaar des N^1-Stickstoffs im Pyrimidinring delokalisiert ist. Folglich konnten unter den gleichen Reaktionsbedingungen die Bildung von Cytidin und Uridin nicht beobachtet werden.[89, 90] Um diese Probleme, die allgemein als das *Glykosylierungs-Problem* zusammengefasst wurden,[14] zu überwinden, wurden alternative Syntheserouten entwickelt, die auf eine direkte Glykosylierung verzichten.



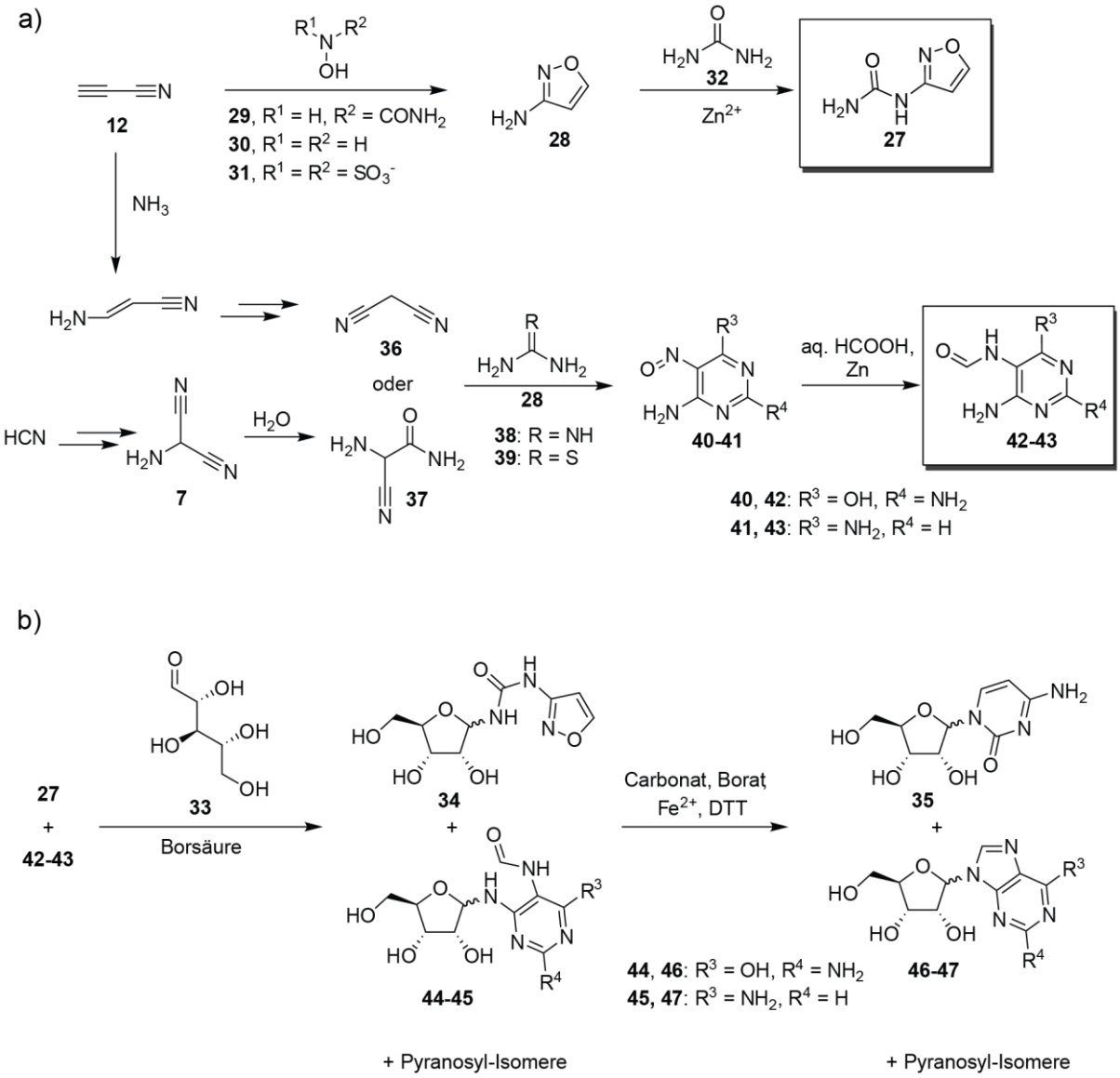
Schema 4. Pyrimidin-Nukleotid (Nukleosid) Synthese nach *Sutherland et al.*[92]

Inspiriert durch die *de novo* Purinbiosynthese wurde von *Sanchez* und *Orgel* ein indirekter stufenweiser Aufbau der Nukleobase an der Ribose als mögliche Lösung vorgeschlagen.[92] Sie gingen von Ribose-5-Phosphat aus, welches sie mit Cyanamid **21** umsetzten. Anschließend reagierte das entstandene Aminooxazolin-Derivat mit Cyanoacetylen **12** und hydrolysierte letztlich zum α-Cytidin. Das natürliche β-Cytidin konnte nach Photoisomerisierung mit einer Ausbeute von 5 % erhalten werden.[93] *Sutherland et al.* erweiterten diese Syntheseroute (**Schema 4**), indem sie von den präbiotisch plausibleren

Startmaterialien Glycolaldehyd **2** und Glycerinaldehyd **3** ausgingen und die Ribose an der Aminooxazol-Struktur aufbauten. Im ersten Schritt reagiert Glycolaldehyd **2** mit Cyanamid **21** zu 2-Aminooxazol **22**. Dies reagiert anschließend in einer Phosphat-katalysierten Reaktion mit Glycerinaldehyd **3** zu Pentose-Aminooxazolinen. Das erwünschte Isomer **23** wird nur mit einer Ausbeute von 15°% gebildet, kann aber in Lösung angereichert werden. Im nächsten Schritt wird das Intermediat mit Cyanoacetylen **12** zum Anhydro-arabinonukleosid **24** umgesetzt und anschließend durch eine Substitution mit Phosphat zum β-Ribocytidin-2‘,3‘-Cyclophosphat **25** umgewandelt. Durch photochemische Bestrahlung von **25** (245 nm) kann das Uridin-Derivat **26** gebildet werden.[92, 94] Diese Synthese war die erste ihrer Art, um Pyrimidinnukleoside in hohen Ausbeuten zu erhalten. Sie zeichnet sich ebenfalls dadurch aus, dass die Probleme um die Verfügbarkeit der Ribose und das *Glykosylierungs-Problem* elegant umgangen werden können. Allerdings bleiben Fragen bezüglich der Plausibilität des Reaktionswegs offen.[95] Zunächst werden chromatographisch aufgereinigte Reagenzien und eine sehr spezifische Abfolge ihrer Zugabe benötigt. Zum Beispiel wird das reaktive und instabile Cyanacetylen in großen Mengen (8 Äquivalente) in einem der letzten Syntheseschritte verwendet. Nennenswert ist zudem die Verwendung von hohen Konzentrationen an Reaktanten und Puffern. Des Weiteren konnte bisher keine Erklärung dafür gefunden werden, wie kanonische Purin-Ribonukleoside unter kompatiblen Bedingungen auf der frühen Erde entstanden sein könnten. Allerdings konnten *Powner* und *Sutherland* die Bildung von 8-Oxo-Purin-Nukleosiden und Purin-Desoxynukleoside über ähnliche Reaktionswege nachweisen.[96, 97]

Eine übereinstimmende Erklärung, wie Pyrimidin- und Purin-Ribonukleoside simultan auf der frühen Erde gebildet wurden, schien weiterhin eine große Herausforderung für die Forschungsgemeinde zu sein. Dieses Problem konnte schließlich 2019 von *Carell et al.* gelöst werden (**Schema 5**).[98, 99] Die postulierte Pyrimidinsynthese nach *Carell* beinhaltet das entscheidende Zwischenprodukt *N*-Isoxazolyl-Harnstoff **27**, welches regioselektiv glykosyliert werden kann. Die Synthese beginnt mit der Bildung von 3-Aminoisoxazol **28**, das durch die Reaktion von Cyanacetylen **12** mit Hydroxylurea **29**, Hydroxylamin **30** oder Nitrosodisulfonat **31** hergestellt werden kann. 3-Aminoisoxazol **28** kann anschließend mit Harnstoff **32** zu **27** umgesetzt werden, welches in einer effizienten lösungsmittelfreien Reaktion mit Ribose **33** zu dem Vorläufer-Nukleosid **34** reagiert. Im letzten Schritt wird in einer Reaktionskaskade die N-O-Bindung der Isoxazol-Untereinheit durch katalytische Mengen an Fe^{2+} und einer Thiolquelle reduktiv aufgespalten. Durch die anschließende intramolekulare Tautomerisierung, Zyklisierung und Wasserabspaltung wird das kanonische β-D-Ribofuranosylcytosin **β-35** (neben anderen Isomeren) mit einer Ausbeute von 27 % gebildet. Die Purinsynthese nach *Carell*, inspiriert durch die Arbeit von *Trinks*[100] und *Koch*[101],

verwendet *N*-Formamidopyrimidine (FaPys) um die kanonischen Purin-Nukleoside zu generieren. Die Syntheseroute beginnt mit der Reaktion von Malononitril **36** oder 2-Amino-2-Cyanoacetamid **37** und Amidinen **38**, die präbiotisch über einfache Moleküle wie HCN, Cyanamid und Cyanacetylen **12** zugänglich sind, zu den Nitrosopyrimidinen **40** und **41**. Überraschenderweise kann diese Reaktion auch gelöst in 3-Aminoisoxazol **28** ablaufen. Nach der Formylierung von **40**-**41**, können die FaPys **42**-**43** regioselektiv mit Ribose **33** zu den Vorläufer-Nukleosiden **44**-**45** reagieren und über eine intramolekulare Kondensation zu den Purin-Nukleosiden **46**-**47** umgewandelt werden. Die letzten zwei Schritte der Pyrimidin- und Purinsynthesen können unter den gleichen Bedingungen ablaufen.[98, 99] Die Regioselektivität während der Ribosylierung ist ein großer Vorteil der *Carell*-Nukleosidsynthesen. Diese geht bei den FaPys auf die C_{2v} Symmetrie und die Anwesenheit der zwei nukleophilen N^4- und N^6-Amino-Aminogruppen zurück. Bei *N*-Isoxazolyl-Harnstoff **32** ist diese Selektivität der Abwesenheit weiterer nukleophiler Zentren geschuldet. Ein Nachteil dieser Synthese ist eben dieser Schritt der Ribosylierung. Durch die direkte Glykosylierung mit den Vorläufer-Nukleosiden entsteht ein Gemisch aus vier möglichen Isomeren: den α/β-Furanosiden und α/β-Pyranosiden. Zudem ist dieser Reaktionsweg von der Verfügbarkeit vollständig gebildeter Ribose in einem späteren Stadium der Synthese abhängig.



Schema 5. a) Synthese von *N*-Isoxazolyl-Harnstoff **27** und *N*-Formamidopyrimidine **42-43** unter präbiotischen Bedingungen b) Pyrimidin-Nukleosid und Purin-Nukleosid-Synthese unter einheitlichen Bedingung nach *Carell et al.* [99]

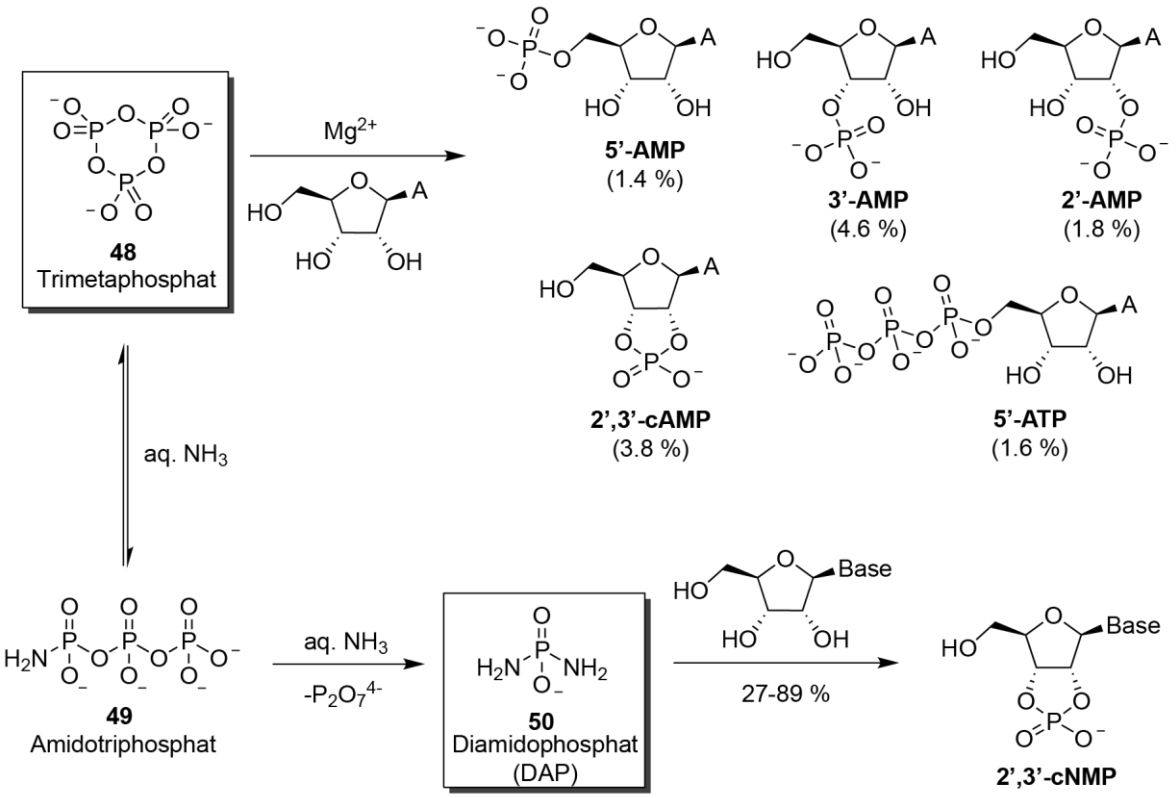
In den letzten Jahren wurde die Bildung von Pyrimidin-Nukleosiden und Purinen-Nukleosiden unter präbiotisch plausiblen Bedingungen mit einer zunehmenden Zahl an experimentellen Fakten belegt. Der nächste logische Schritt ist die Suche nach einer Lösung für die Frage: Wie können diese Nukleoside phosphoryliert und miteinander verknüpft (oligomerisert) werden, um RNA-Moleküle zu bilden?

1.5.2. Phosphorylierung

Die Verfügbarkeit von anorganischem Phosphat auf der frühen Erde ist stark umstritten. Orthophosphat (PO_4^{3-}), die häufigste Form von Phosphor auf der Erde, kommt größtenteils in

Apatitmineralen vor, die in den meisten geologischen Gegebenheiten relativ unreaktiv und unlöslich sind.[102] Die Schwierigkeit konzentriertes Phosphat zu bilden, zusammen mit seiner zentralen Rolle in präbiotischen Synthesen, wurde seit jeher als *Phosphat-Problem* bezeichnet.[103-105] Dennoch konnten plausible Ansätze und Bedingungen für Phosphorylierungen gefunden werden.

Der am meisten untersuchte Ansatz für präbiotische Phosphorylierungen ist die elektrophile Aktivierung von Orthophosphaten. Hierfür wurden zahlreiche Aktivierungsmittel wie Cyanat[106], Cyanacetylen **12**[107], Cyanamid[108] und Dicyan[109] betrachtet. Phosphorylierungsreaktionen in Wasser sind thermodynamisch ungünstig, da das Wasser mit dem Reaktionspartnern konkurriert und somit konnten nur geringe Ausbeuten durch diese Ansätze erhalten werden.[106, 108, 110] Um dieses Problem zu umgehen, wurden wasserfreie Bedingungen in Betracht gezogen. Auf diese Weise konnten Nukleotide mit Ausbeuten von bis zu 50 % durch Reaktionen in Formamid[111, 112] und 16 % unter lösungsmittelfreien Bedingungen erhalten werden.[113] Die selektive Phosphorylierung der 5‘-Position konnte ebenfalls vor kurzem in *dry-down*-Experimenten unter wasserfreien Bedingungen in Ameisensäure gezeigt werden.[114] Als Alternative wurden auch reaktivere Phosphorquellen untersucht. Reduzierte Phosphorspezies in Form des meteoritischen Minerals Schreibersit ($(Fe,Ni)_3P$) können in Wasser korrodieren (Hydrolysieren) und ein Gemisch aus Phosphit (HPO_3^-), Phosphat, Hypophosphat ($P_2O_6^{4-}$), Pyrophosphat ($P_2O_7^{4-}$), Triphosphat und Trimetaphosphat **48** (TMP) bilden.[115-117] *Pasek et al.* zeigten, dass diese Korrosion mit der Phosphorylierung von Nukleosiden gekoppelt werden kann, um geringe Mengen an Nukleotiden zu gewinnen.[118] Das Korrosionsprodukt Trimetaphosphat **48** kann ebenfalls durch magmatische Prozesse und durch die Kondensation von Phosphaten gewonnen werden.[119-121] Die zyklische Phosphat-Verbindung ist wasserlöslich und kann durch die energetisch begünstigte Ringöffnung diverse Nukleotide bilden.[110] 2‘-AMP und 3‘-AMP können mit einer Ausbeute von 31 % (1:1 Gemisch) durch die Reaktion von Adenosin mit Natriumtrimetaphosphat **48** unter basischen Bedingungen gebildet werden.[122] Durch den Zusatz von Metallkatalysatoren können ebenfalls 2‘,3‘-cAMP, 5‘-AMP und das biologisch relevante 5‘-ATP gebildet werden (**Schema 6**).[123] Letztere Verbindung galt bis dato als sehr schwierig zu synthetisieren.[124, 125] Darüber hinaus kann Trimetaphosphat **48** mit Ammoniak reagieren und Amidotriphosphat **49** (AmTP) und das Phosphorylierungsreagenz Diamidophosphat **50** (DAP) bilden.[126] *Krishnamurthy* und seine Mitarbeiter konnten zeigen, dass **50** mit Ribonukleosiden zu 2‘,3‘-cNMPs mit Ausbeuten von bis zu 89 % reagierten (**Schema 6**).[127, 128] Außerdem konnten die Phosphorylierungen von Zuckern[129, 130], Fettsäuren[129] und die Oligomerisierung von Aminosäuren[131] durch DAP nachgewiesen werden.



Schema 6. Phosphorylierung von Nukleosiden unter präbiotisch plausiblen Bedingungen mit Trimetaphosphat **44** und Diamidophosphat **50**.[123, 127]

1.5.3. Oligomerisierung

Wie bereits in **Abschnitt 1.4**. beschrieben gilt die nicht-enzymatische Oligomerisierung von Nukleotiden und die anschließende nicht-enzymatische Replikation als eine der entscheidenden Argumente zur Begründung der RNA-Welt-Theorie. Allerdings ist die Kondensation (Oligomerisierung) von Nukleotiden in Wasser eine thermodynamisch sowie kinetisch ungünstige Reaktion, die nicht spontan auftreten kann.[14] Dies, zusammen mit der geringen Regioselektivität der Reaktion (2'-5'- oder 3'-5'-Phosphodiesterbindungen sind die möglichen Produkte), sind die größten Probleme bei der Oligomerisierung zur Bildung von RNA. Trotz dieser Herausforderungen konnten beachtliche Erfolge erzielt werden. In den ersten Experimenten wurden Nukleotidmonophosphate (NMP) lösungsmittelfrei erhitzt. Dadurch wurde ein komplexes Gemisch an kurzen Oligonukleotiden mit zufälligen 2'-5'- oder 3'-5'-Phosphodiesterbindungen erhalten.[132] Versuche, Nukleotide unter wässrigen Bedingungen zu polymerisieren, benötigten zwangsläufig Aktivierungsmittel. Hierfür wurden zum Beispiel Cyanamid und andere wasserlösliche Carbodiimide verwendet, allerdings konnten nur geringen Ausbeuten an Dinukuleotiden und sehr kurze Oligonukleotide erhalten werden.[14, 133, 134] Im Gegensatz zu den Monophosphaten sind zyklische Phosphate durch die

thermodynamisch begünstigte Hydrolyse reaktiver.[135] Für die Oligomerisierung können 2‘,3‘-zyklische Monophosphate[136, 137] oder 3‘,5‘-zyklische Monophosphate[138, 139] unter lösungsmittelfreien Bedingungen erhitzt werden. Im Fall von 3‘,5‘-zyklischen GMP konnten Oligomere mit Längen von bis zu 40 Nukleotiden erhalten werden.[139] Alternativ wurden auch aktivierte 5‘-Phosphate in Form von Phosphorimidazolen als Substrate in Betracht gezogen. Phosphorimidazole von Nukleotiden können unter präbiotisch plausiblen Bedingungen aus 5‘-Nukleosidmonophoshaten oder 5‘-Nukleosidpolyphosphaten und Imidazol gebildet werden.[140-142] In Gegenwart von Metallionen (z.B. Zn^{2+}, Pb^{2+}, UO_2^{2+}, Lu^{3+}) konnten aus den aktivierten 5‘-Phosphaten bis zu 16mere (katalysiert durch UO_2^{2+}) erhalten werden. Allerdings wurden durch diese Reaktionen hauptsächlich 2‘,5‘-verknüpfte Oligomere und Pyrophosphate erhalten.[143-147] *Ferris et al.* konnten daraufhin zeigen, dass Mineralien die Regioselektivität der Kondensation steuern können. Durch Reaktionen an der Oberfläche von Montmorillonit wurden bevorzugt 3‘-5‘ verknüpfte Oligomere gebildet, wobei Längen von bis zu 50 Nukleotid-Einheiten erhalten wurden.[148-151] *Deamer et al.* erweiterten dies, indem sie diese Reaktion in multi-lamellenartigen Lipid-Vesikel durchführten. Die Lipid Matrix beschleunigte die Polymerisation, so konnten durch Hydrations- und Dehydrations-Zyklen Polymere mit 25-100 Einheiten erhalten werden.[152, 153]

Nach erfolgreicher Bildung von Oligonukleotiden mit verschiedenen Sequenzen können die Einzelstränge über nicht-enzymatische, templat-vermittelte Polymerisation die Synthese der komplementären Stränge steuern und somit ein selbstreplizierendes System aufbauen. Für diesen Prozess wurden bereits attraktive Ansätze diskutiert.[66, 154]

1.6. RNA-Modifikationen und der Ursprung der Translation

Bevor die drei Domänen des Lebens und frühe Zellentwicklung entstehen konnte, musste sich ein primitives Translationssystem in der RNA-Welt entwickelt haben.[155] Die Evolution der Translation hat wahrscheinlich mit der Synthese von kurzen und zufälligen Peptidsequenzen begonnen, die eine begrenzte Anzahl an präbiotisch plausiblen Aminosäuren (A, D, E, G, I, L, P, S, T, V) besaßen.[156-159] Die Tatsache, dass Proteine ausschließlich in der Peripherie des Ribosoms zu finden sind, während RNA in Form von ribosomaler RNA (rRNA) das Peptidyltransferase-Zentrum bildet und die Funktion der tRNA-Selektion ausübt, deutet darauf hin, dass das erste Translationssystem ausschließlich aus RNA bestand.[62, 156] Neben dem Translationsapparat musste auch der genetische Code bereits früh existiert haben. Dies würde bedeuten, dass Transfer-RNAs (tRNA) ebenfalls bereits in einem frühen Stadium der Evolution gebildet wurden.[155]

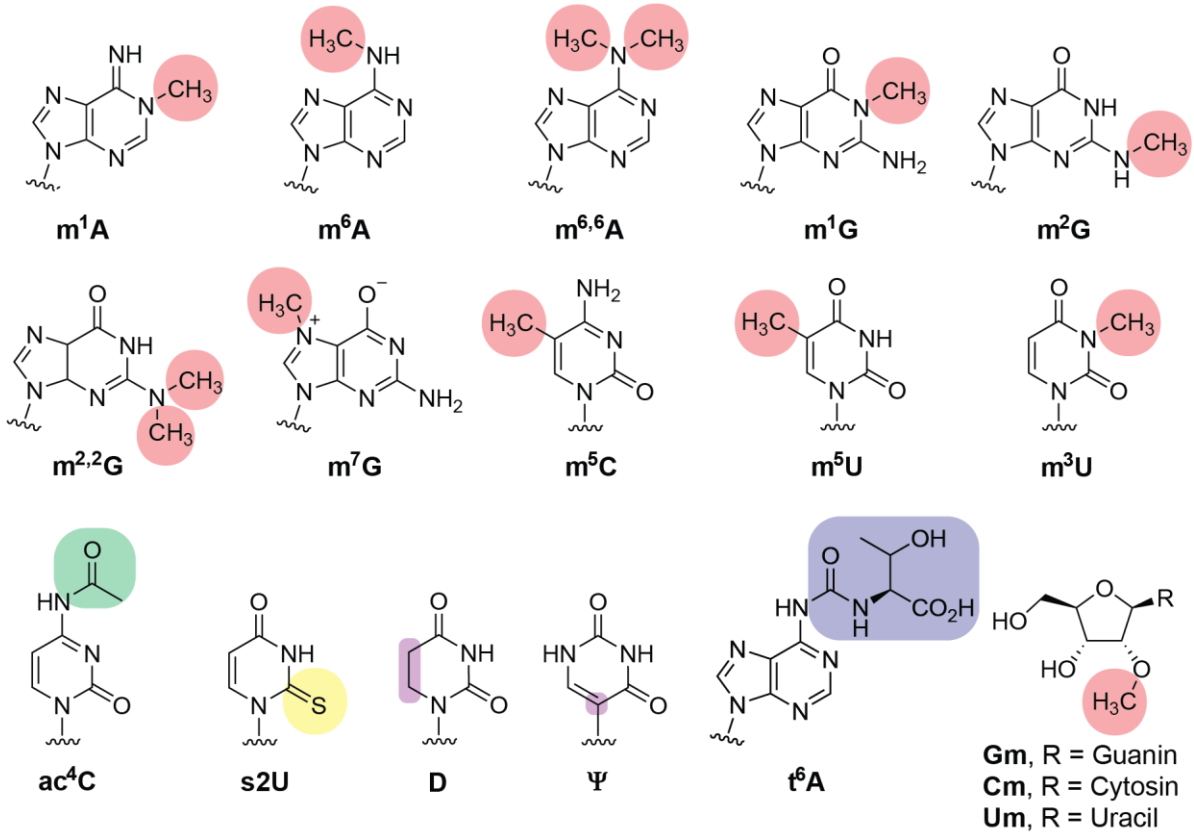


Abbildung 3. Strukturen von modifizierten Nukleosiden, die durch phylogenetische Studien als molekulare Fossile von LUCA identifiziert wurden sind.[160]

Um nach Hinweisen für ein primitives Translationssystem zu suchen, können die Strukturen der rRNAs und tRNAs genauer betrachtet werden. Vor allem in tRNAs können eine Vielzahl an Modifikationen an den Nukleobasen und der Ribose gefunden werden. 18 dieser Modifikationen sind in den tRNAs aller drei Domänen des Lebens vorhanden und können somit als molekulare Fossile von LUCA betrachtet werden (**Abbildung 3**).[4, 160] Diese Modifikationen sind aus chemischer Sicht meistens relativ einfach. Es handelt sich meistens um Methylierungen an der 2'-Position der Ribose oder an den Nukleobasen. Weitere Modifikationen umfassen Acetylierungen (**ac^{4}C**), Thiolierungen (**S^{5}U**), die Reduktion der Doppelbindungen von Uridin (Dihydrouridin, **D**) oder die Isomerisierung der Uracil-Ribose-Verknüpfung (Pseudouridin, **Ψ**). Die komplexeste Modifikation, *N*^{6}-Threonylcarbamoyl-Adenosin (**t^{6}A**), enthält die Aminosäure Threonin, die über eine Harnstoffbrücke an Adenosin gebunden ist. [161-163] Diese Modifikation kann ausschließlich in der Position 37 nahe der Anticodon-Schleife gefunden werden und ist für die korrekte Ausbildung der Anticodon-Schleife während der Translation verantwortlich.[164, 165] Es ist vorstellbar, dass Aminosäure-modifizierte Nukleoside wie **t^{6}A** Teil des primitiven Translationssystems waren. Diese Hypothese wird durch *Carell et al.* unterstützt, die zeigen konnten, dass *N*^{6}-

Glycinylcarbamoyl-Adenosin (**G^6A**) und **t^6A** unter präbiotisch plausiblen Bedingungen gebildet werden können.[166] In dem heutigen Translationsprozess werden Aminosäuren über 5'-Aminoacyladenyate (aa-AMP) aktiviert und anschließend über eine Ester-Bindung mit der 3'-Hydroxygruppe der Ribose verknüpft.[167] Diese Bindung kann unter basischen oder sauren Bedingungen leicht hydrolysiert werden.[168, 169] Es ist somit unwahrscheinlich, dass dieses System unter den harschen Bedingungen der frühen Erde in einem primitiven Translationsprozess zum Einsatz kam.

Die Fragen, wie die Translation auf der frühen Erde entstanden sein könnte und wie die RNA-Welt in eine RNA-Peptid-Welt übergehen konnte, bleiben weiterhin einige der zentralen ungeklärten Fragen.[170] Vor kurzem wurden Prozesse entdeckt, die die nicht kodierte Verlängerung von Peptiden unter präbiotisch plausiblen Bedingungen erklären könnten.[171] Darüber hinaus berichteten *Richert et al.* über die templatgesteuerte Bildung von Peptidbindungen mit Hilfe von Aminosäuren-beladenen Nukleotiden. Allerdings konnten über diesen Mechanismus nur maximal Tripeptide gebildet werden.[172] Eine weitere Möglichkeit, wie Peptide mittels Aminosäure-modifizierter Nukleoside an der RNA synthetisiert werden können, wird in **Abschnitt 3.1.** diskutiert.

1.7. Alternative zur RNA

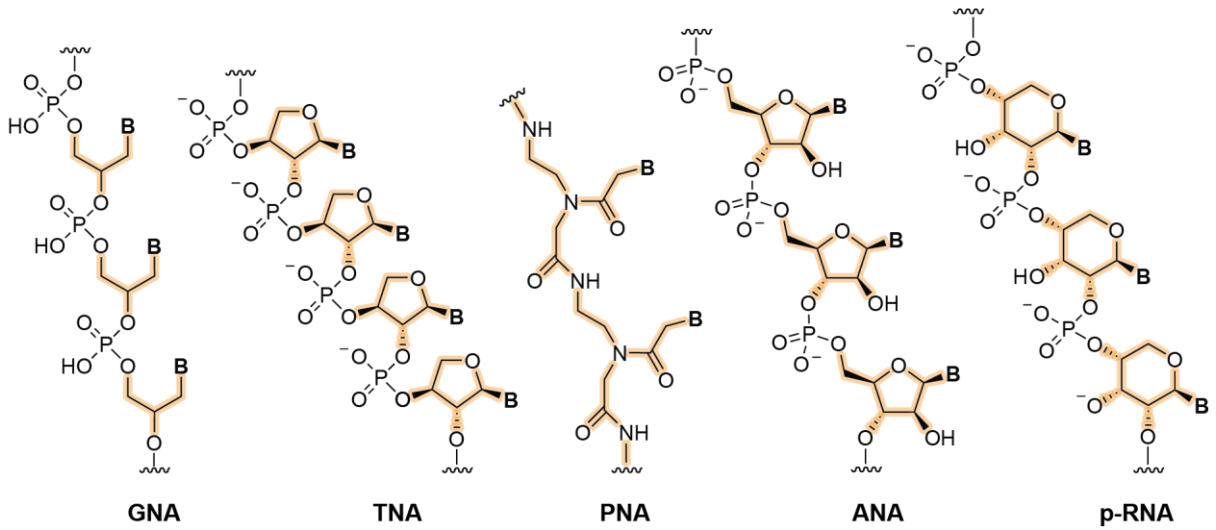


Abbildung 4. Strukturen von möglichen Vorläufer-Polymeren.

In Anbetracht der zahlreichen Herausforderungen, die die Synthese von RNA unter präbiotischen Bedingungen mit sich bringt, sollte die Möglichkeit in Erwägung gezogen werden, dass RNA-Moleküle nicht die ersten genetischen Systeme auf der frühen Erde waren. Die Idee, dass die RNA ein Produkt chemischer Evolution ist und somit ein Nachfolger einer einfacheren „Proto-RNA", wurde bereits 1987 von *Orgel* und *Joyce*

vorgeschlagen.[173] Nach dieser Theorie besteht die Möglichkeit, dass am Anfang der chemischen Evolution einfachere und besser verfügbare Moleküle allen drei strukturellen Komponenten der RNA vorrausgingen. Durch schrittweisen Austausch der Komponenten im Laufe der RNA-Evolution haben sich die vier kanonischen Nukleobasen, Ribose und Phosphat als Bausteine durchgesetzt. Es ist denkbar, dass während dieser Evolution eine Komponente mehrfach ausgetauscht wurde und auch mehrere Komponenten simultan ausgetauscht wurden. Demnach könnte eine Vielzahl intermediärer Proto-RNAs im Laufe der Entwicklung zur spezialisierten (kanonischen) RNA existiert haben.[174-177] Dies bedeutet, dass der chemische Raum der Proto-RNA Kandidaten fast grenzenlos erscheint. Ein Szenario, das einst von *Leslie Orgel* als eine „düstere Aussicht“ (*gloomy prospect*) beschrieben wurde.[178] Seitdem wurden bereits vielversprechende Vorschläge für potenzielle Vorläufer Strukturen entwickelt.

Orgel und *Joyce* postulierten die Existenz von Vorläufer-Polymeren aus einfachen, flexiblen, möglicherweise prochiralen Nukleotidanaloga, die auf der frühen Erde leicht synthetisiert werden konnte und schnell polymerisieren konnten. Dabei schlugen sie acylische Nukleinsäuren vor, die von Glycerin abgeleitet sind.[173] *Meggers et al.* konnten dies experimentell validieren, indem sie zeigten, dass Glycol-Nukleinsäuren (**GNA**) in der Lage sind mit sich selbst und mit RNA antiparallele Duplexe zu formen (**Abbildung 4**).[179] Die Kreuzhybridisierung gilt als die essentielle Eigenschaft einer potentiellen Proto-RNA, wodurch ein effektiver Transfer von Information zwischen den Polymeren während der Evolution gesichert werden konnte.[180] *Eschenmoser* zeigte, dass Threose-Nukleinsäuren (**TNA**), welche aus 3‘,2‘-verknüpfte α-D-Threofuranosyl-Einheiten bestehen, ebenfalls diese Eigenschaft besitzen und *Watson-Crick*-Paarungen mit sich selbst und auch RNA eingehen können.[181, 182] Der C4-Zucker Threose ist aus präbiotischer Sicht einfacher zugänglich als Ribose, was es zu einem attraktiven Kandidaten macht. In den letzten Jahren wurde dieser Ansatz zudem durch die präbiotischen Synthesen von Threose-Nukleosiden gestärkt.[183, 184] Eines der radikalsten Vorschläge ist wohl die Peptid-Nukleinsäure (**PNA**), aufgebaut aus *N*-(2-Aminoethyl)glycin-Einheiten. Der große Vorteil der **PNA** ist Verwendung einer einzelnen Molekülspezies, um das gesamte Rückgrat des Polymers herzustellen. Auch Strukturen, die eine größere Ähnlichkeit mit Ribose aufweisen können in Betracht gezogen werden. „Kanonische“ Arabino-Nukleoside können präbiotisch in einer Reaktionssequenz mit furanosyl-Selektivität geformt werden.[185] Die Attraktivität der Arabinose-Nukleinsäuren (**ANA**) liegt neben der ähnlichen Tertiärstruktur zu RNA und DNA vor allem in der höheren hydrolytischen Stabilität des Phosphodiester Rückgrates.[186, 187]

Eschenmoser konnte darüber hinaus zeigen, dass pyranosyl-RNA (**p-RNA**) im Vergleich zur kanonischen RNA deutlich stabilere Duplexe formen kann. Diese Beobachtung kann als ein

Argument für den evolutionären Vorteil der RNA interpretiert werden, da eine optimale und nicht maximale Duplexstabilität benötigt wird. Die starke helikale Stabilität könnte zu eine Selbstinhibierung der **p-RNA** geführt haben, welche sie für eine Replikation unbrauchbar gemacht hat. [188, 189]

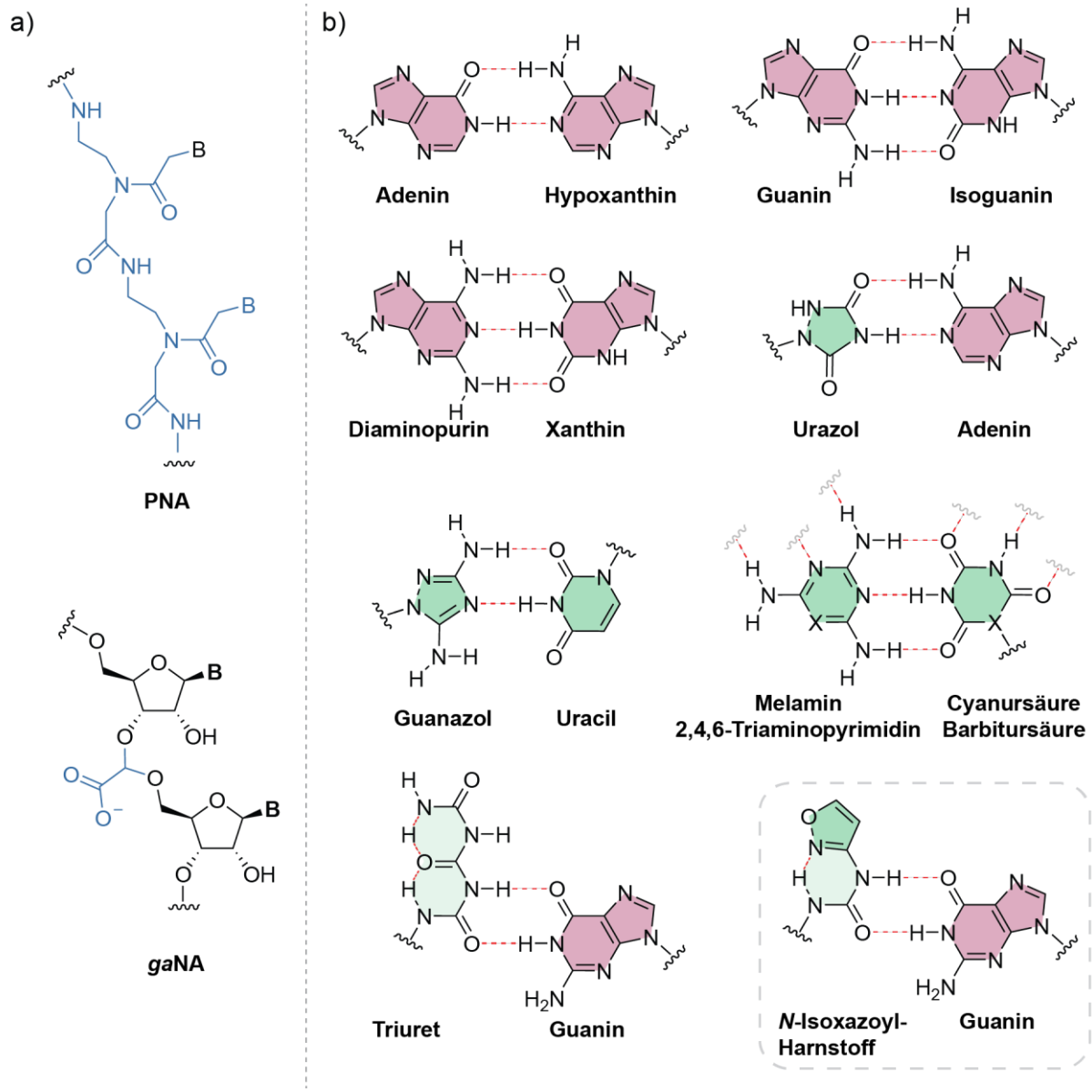


Abbildung 5. Beispiele für plausible proto-RNA Komponenten a) Phosphodiester-Vorläufer b) Vorläufer Basenpaare.

Es wird vermutet, dass die Phosphodiester-Bindung ebenfalls das optimale Produkt der molekularen Evolution ist. Die Esterbindung ist labil genug, damit Enzyme Nukleinsäuren hydrolysieren können und gleichzeitig schützt die negative Ladung das Phosphorzentrum vor nukleophilen Angriffen, um es gegen spontane Hydrolyse zu stabilisieren. Außerdem sorgt die gleichmäßig verteilte negative Ladung für die helikale Struktur der Nukleinsäuren und die Fähigkeit *Watson-Crick*-Basenpaarungen auszubilden.[175] Allerdings lassen die in **Abschnitt**

1.5. beschriebenen Probleme der geringen Verfügbarkeit von Phosphat und der thermodynamisch benachteiligten Kondensation darauf schließen, dass Phosphate nicht die erste Wahl während der Evolution waren. *Hud* präsentierte die Idee eines Proto-Rückgrates bestehend aus Formosereaktionsprodukten. Aldehyde können zwei Moleküle, die OH-Gruppen enthalten, unter Bildung eines Acetals reversibel verknüpfen.[190] Bereits zuvor konnten *Gao* und *Jones* zeigen, dass die sogenannten Acetal-verknüpften Nukleinsäuren (***a*NA**) in der Lage sind Duplexstrukturen mit RNA und DNA zu formen.[191, 192] Je nach Struktur der Aldehyd-Seitenkette können die ***a*NA** ebenfalls negativ geladen sein. Ein Beispiel hierfür sind Glyoxylat-Acetal-Nukleinsäuren (***ga*NA**), die durch das Erhitzen von Nukleosiden, Glyoxylat und mono- oder bivalenten Kationen hergestellt werden können (**Abbildung 5a**).[193] Glyoxylat kann nachweislich unter präbiotisch plausiblen Bedingungen aus Glycolaldehyd **2** gebildet werden.[194] *Hud et al.* konnten zeigen, dass ***ga*NA** Duplexe sehr ähnliche helikale Eigenschaften wie RNA aufweisen und erstaunlich stabil gegenüber spontaner Hydrolyse sind.[193] Auch H-Phosphonate wurden als ein potentieller Vorläufer diskutiert. H-Phosphonatdiester-Bindungen können reversibel gebildet werden und könnten durch eine einfache Oxidation zur Phosphodiester-Bindung umgewandelt werden. Allerdings konnten bisher keine Beweise für eine präbiotisch plausible Bildung einer solchen Nukleinsäure gefunden werden.[175]

Der älteste Vorschlag für ein nicht-kanonisches Proto-RNA Basenpaar stammt von *Crick*, der für ein primitives System, welches nur aus dem Purin-Purin-Basenpaar Adenin und Hypoxanthin bestand, warb (**Abbildung 5b**).[51] Grund dafür war die vergleichsweise einfache Bildung von Adenin unter präbiotischen Bedingungen (siehe **Abschnitt 1.5**)[78, 195] und der Möglichkeit dieses über Desaminierung zu Hypoxanthin zu überführen.[51] Allerdings konnte nachgewiesen werden, dass Duplexe, die aus Homo-Adenosin und Homo-Inosin-Strängen bestehen, relativ instabil sind.[196] Später konnten *Eschenmoser und Switzer* jeweils zeigen, dass aufgrund ihrer Duplex-Stabilität Guanin-Isoguanin und Diaminopurin-Xanthin bessere Kandidaten für einen primitiven Code sind.[197-199] Auch Heterozyklyen, die nur geringe Ähnlichkeiten zu den kanonischen Strukturen aufweisen, können in Betracht gezogen werden. So schlug *Miller* Urazol und Guanazol aufgrund ihrer einfachen präbiotischen Zugänglichkeit und hohen Stabilität gegenüber UV-Strahlungen vor.[200] *Hud et al.* gingen noch einen Schritt weiter und postulierten, dass Information nicht zwangsläufig in Basenpaaren gespeichert wurde. Sie zeigten, dass 2,4,6-Triaminopyrimidin und Melamin mit Barbitursäure oder Cyanursäure in Lösung Hexaden bilden und sich zu Polymeren assemblieren können.[201] Die meisten dieser Beispiele basieren auf heterozyklischen Strukturen, die über mehrstufige präbiotische Synthesen gebildet werden. Es ist vorstellbar, dass das erste System zur Speicherung von Information auf einfacheren Strukturen basiert

war. Aus diesem Grund postulierten *Okamura et al.*, dass Harnstoffe in Form von Biuret und Triuret durch intramolekulare Wasserstoffbrückenbindungen die planare Struktur der kanonischen Nukleobasen imitieren und somit ebenfalls in der Lage sind *Watson-Crick*-Basenpaarungen einzugehen.[202]

Somit steht eine Vielzahl an möglichen Kandidaten für alle drei strukturellen Komponenten einer potenziellen proto-RNA zur Verfügung. Allerdings besteht die übergreifende Frage, wie der Austausch der Komponenten stattgefunden hat. Eine mögliche Lösung wird im Rahmen dieser Arbeit diskutiert (**Abschnitt 3**).

2. Zielsetzung

Der primäre Fokus dieser Arbeit lag auf der Untersuchung von plausiblen proto-RNA Basenpaarungssystemen. Nachdem gezeigt werden konnte, dass das *N*-Isoxazol-3-yl-Harnstoff-Ribosid und die Formamidopyrimidin-Riboside durch intramolekulare Reaktionen zu den kanonischen Nukleosiden Cytidin, Adenosin und Guanosin umgewandelt werden können,[99] stellte sich die Frage, ob diese Strukturen Teil einer potentiellen Proto-RNA gewesen sein könnten (**Abbildung 6**). Diese könnten als ein alternatives Modell dienen, um den Übergang von einer Proto-RNA-Welt zu einer RNA-Welt zu erklären, bei dem ein Austausch der Basen-Komponente nicht notwendig war.

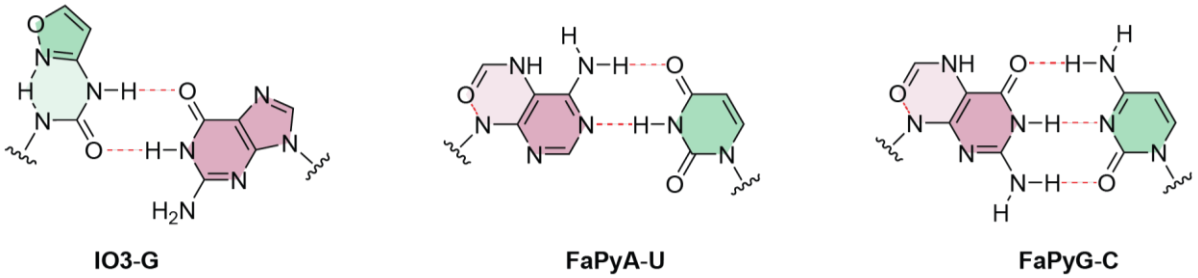


Abbildung 6. Postuliertes Basenpaarungsverhalten der potentiellen Proto-Nukleobasen *N*-Isoxazol-3-yl-Harnstoff **IO3**, 4,6-Diamino-5-formamidopyrimidin **FaPyA** und 2,6-Diamino-4-oxo-5-formamidopyrimidin **FaPyG**.

Die Kreuzhybridisierung mit RNA gilt als die wichtigste Eigenschaft, um die Plausibilität einer potentiellen Proto-RNA-Komponente zu ermitteln. Zu diesem Zweck sollten zunächst die Basenpaarungseigenschaften der Proto-Nukleobasen untersucht werden. Anschließend sollte der stereochemische Einfluss des RNA-Stranges (und des komplementären RNA-Stranges) auf die intramolekularen Reaktionen untersucht werden. Zu einem besseren Verständnis des strukturellen und dynamischen Einflusses der Modifikation auf die RNA-Struktur sollten NMR-Untersuchungen synthetischer Oligonukleotide, die die Proto-Nukleobasen enthalten, in Kooperation mit der Arbeitsgruppe von *Dr. Petra Rovó* durchgeführt werden. Unerlässlich für diese Untersuchungen war die Entwicklung effizienter chemischer Syntheserouten, um die Phosphoramidit-Bausteine und in Referenzmoleküle in ausreichenden Mengenherzustellen. Im Zuge dessen sollte auch ein Vorläufermolekül für das kanonische Nukleosid Uridin identifiziert werden, welches unter präbiotisch plausiblen Bedingungen gebildet werden kann. Dieses Vorläufermolekül sollte anschließend ebenfalls auf die Fähigkeit Basenpaarungen einzugehen untersucht werden.

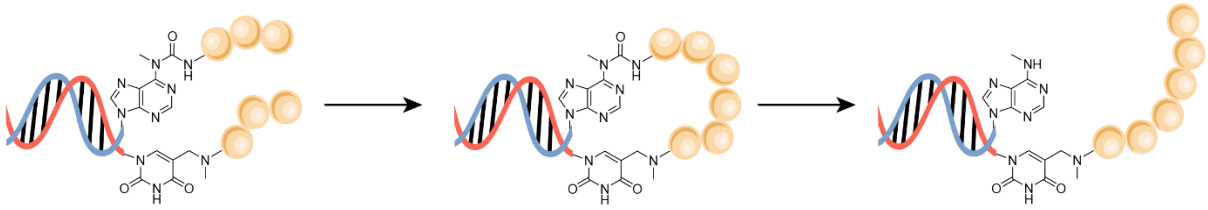


Abbildung 7. Schematische Darstellung eines templatbasierten, primitiven Peptidsynthese-Systems.

Im zweiten Teil dieser Arbeit sollte der Einfluss von Aminosäure-modifizierten Nukleotiden in einer potentiellen RNA-Peptid-Welt untersucht werden. Hierzu sollte ein templatbasiertes, primitives Peptidsynthese-System entwickelt werden (**Abbildung 7**). Um diese Studien durchzuführen, sollten zunächst nicht-kanonische, Aminosäure-modifizierte Phosphoramidit-Bausteine synthetisiert und anschließend in RNA-Oligomere inkorporiert werden. Anschließend sollte die Möglichkeit einer templatgesteuerten Peptidbildung untersucht werden. Ein besonderer Fokus sollte auf die Untersuchung präbiotisch plausibler Mittel zur Carbonsäureaktivierung gelegt werden.

3. Veröffentlichte Ergebnisse

3.1. Ein präbiotisch plausibles Szenario einer RNA-Peptid-Welt

„A prebiotically plausible scenario of an RNA-peptide world"

Felix Müller†, Luis Escobar†, Felix Xu, Ewa Węgrzyn, Milda Nainytė, Tynchtyk Amatov, Chun-Yin Chan, Alexander Pichler, Thomas Carell*

Nature, **2022**, 605, 279-284.

† Beitrag der Autoren zu gleichen Teilen

Für Hintergrundinformationen siehe Anhang I

Prolog

Die RNA-Welt Theorie geht davon aus, dass die RNA am Anfang der chemischen Evolution stand und dass katalytische Peptide in der Evolution viel später entstanden sind. Es stellt sich die Frage, wie die RNA-Welt zu einem Stadium überging, in dem Proteine die Katalysatoren des Lebens wurden. Es wird postuliert, dass statt einer reinen RNA-Welt eine RNA-Peptid-Welt existiert hat, in der neben der RNA auch zufällige, kurze Peptidsequenzen gebildet wurden. Dieses Manuskript zeigt, dass RNA mittels nicht kanonischer RNA-Basen, die heute in tRNAs und rRNAs zu finden sind und als molekulare Fossile der RNA-Welt gelten, eine primitive Form der Peptidsynthese durchführen konnten. Die Fähigkeit Peptide auf RNA zu bilden, ist ein Indiz für eine mögliche Co-Evolution von Nukleinsäuren und Peptiden und kann als ein Hinweis auf ein primitives Translationssystem gedeutet werden.

Autorenbeitrag

Siehe Publikation

Lizenz

Article

A prebiotically plausible scenario of an RNA–peptide world

https://doi.org/10.1038/s41586-022-04676-3

Received: 13 July 2021

Accepted: 22 February 2022

Published online: 11 May 2022

Open access

Check for updates

Felix Müller[1,2], Luis Escobar[1,2], Felix Xu[1], Ewa Węgrzyn[1], Milda Nainytė[1], Tynchtyk Amatov[1], Chun-Yin Chan[1], Alexander Pichler[1] & Thomas Carell[1]✉

The RNA world concept[1] is one of the most fundamental pillars of the origin of life theory[2–4]. It predicts that life evolved from increasingly complex self-replicating RNA molecules[1,2,4]. The question of how this RNA world then advanced to the next stage, in which proteins became the catalysts of life and RNA reduced its function predominantly to information storage, is one of the most mysterious chicken-and-egg conundrums in evolution[3–5]. Here we show that non-canonical RNA bases, which are found today in transfer and ribosomal RNAs[6,7], and which are considered to be relics of the RNA world[8–12], are able to establish peptide synthesis directly on RNA. The discovered chemistry creates complex peptide-decorated RNA chimeric molecules, which suggests the early existence of an RNA–peptide world[13] from which ribosomal peptide synthesis[14] may have emerged[15,16]. The ability to grow peptides on RNA with the help of non-canonical vestige nucleosides offers the possibility of an early co-evolution of covalently connected RNAs and peptides[13,17,18], which then could have dissociated at a higher level of sophistication to create the dualistic nucleic acid–protein world that is the hallmark of all life on Earth.

A central commonality of all cellular life is the translational process, in which ribosomal RNA (rRNA) catalyses peptide formation with the help of transfer RNAs (tRNA), which function as amino acid carrying adapter molecules[14,19,20]. Comparative genomics[21] suggests that ribosomal translation is one of the oldest evolutionary processes[15,16,22,23], which dates back to the hypothetical RNA world[1–4]. The questions of how and when RNA learned to instruct peptide synthesis is one of the grand unsolved challenges in prebiotic evolutionary research[3–5].

The immense complexity of ribosomal translation[14] demands a stepwise evolutionary process[11]. From the perspective of the RNA world, at some point RNA must have gained the ability to instruct and catalyse the synthesis of, initially, just small peptides. This initiated the transition from a pure RNA world[1] into an RNA–peptide world[13]. In this RNA–peptide world, both molecular species could have co-evolved to gain increasing 'translation' and 'replication' efficiency[17].

To gain insight into the initial processes that may have enabled the emergence of an RNA–peptide world[13], we analysed the chemical properties of non-canonical nucleosides[6,7], which can be traced back to the last universal common ancestor and, as such, are considered to be 'living molecular fossils' of an early RNA world[8–12].

This approach, which can be called 'palaeochemistry', enabled us to learn about the chemical possibilities that existed in the RNA world and, therefore, sets the chemical framework for the emergence of life. In contrast to earlier investigations of the origin of translation[24–29], we used naturally occurring non-canonical vestige nucleosides and conditions compatible with aqueous wet–dry cycles[30,31].

Peptide synthesis on RNA

In modern tRNAs (Fig. 1a), the amino acids that give peptides are linked to the CCA 3′ terminus via a labile ester group[32]. Some tRNAs, however, contain additional amino acids in the form of amino acid-modified nucleosides, for example, g^{6}A (ref. [33]), t^{6}A (ref. [34]) and m^{6}t^{6}A (ref. [35]), which are found directly next to the anticodon loop at position 37. Other non-canonical vestige nucleosides often present in the wobble position 34 are nm^{5}U and mnm^{5}U (refs. [36–38]).

Close inspection of their chemical structures (Fig. 1b) suggests that if they are in close proximity (step 1), an RNA-based peptide synthesis may be able to start (step 2), which would create, via a hairpin-type intermediate, a peptide attached by a urea linkage to the nucleobase (m^6)aa^{6}A. Cleavage of the urea[39,40] (step 3) would furnish RNA with a peptide connected to a (m)nm^{5}U (step 4). Subsequently, strand displacement with a new (m^6)aa^{6}A strand may finally enable the next peptide elongation step.

To investigate the potential evolution of an RNA–peptide world, we synthesized two complementary sets of RNA strands, **1a**–**1j** and **2a**–**2c** (Fig. 2). The first set contained various m^{6}aa^{6}A nucleotides[41] at the 5′ end (**1a**–**1j**) as RNA donor strands. The complementary RNA acceptor strands were prepared with an (m)nm^{5}U nucleotide at the 3′ terminus (**2a**–**2c**). Figure 2a shows the reactions between **1a** and **2a**. The analytical data are presented in Fig. 2b. We hybridized **1a** with **2a** and activated the carboxylic acid of **1a** using reagents such as EDC[42]/Sulfo-NHS[43], DMTMM•Cl[43] or methyl isonitrile[44] (pH 6, 25 °C). In all cases we observed high yielding product formation (Fig. 2c).

A kinetic analysis shows that the nature of the amino acid affects the coupling rate (Fig. 2d). For example, G (in **1a**) couples to **2c** with an apparent rate constant (k_{app}) of 0.1 h^{-1}. For the amino acids L (in **1d**), T (in **1e**) and M (in **1h**) a fourfold higher rate constant (≈0.4 h^{-1}) was determined, and the highest rate was measured for F (in **1g**) with k_{app} > 1 h^{-1}. These differences establish a pronounced amino acid selectivity in the coupling reaction, probably as a result of distinct pre-organizations. We next reduced the length of the RNA donor strand to five, and finally to three, nucleotides (Supplementary Information). We detected coupling even with a trimer

[1]Department of Chemistry, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany. [2]These authors contributed equally: Felix Müller, Luis Escobar. ✉e-mail: thomas.carell@lmu.de

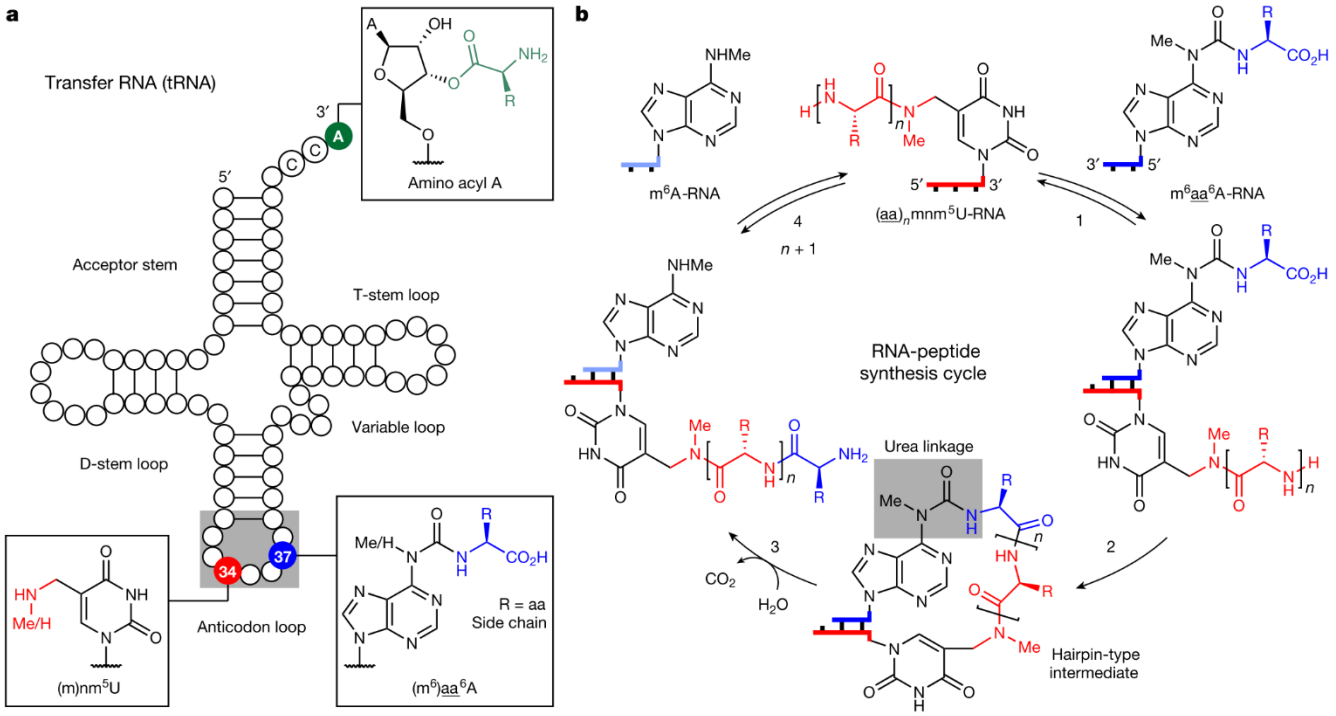


Fig. 1 | Concept of how nucleoside relics of the RNA world enable RNA-based peptide synthesis. a, tRNA structure showing selected ribose and nucleobase modifications. The 3′-amino acid-acylated adenosine is located at the CCA 3′ end in contemporary tRNAs. 5-Methylaminomethyl uridine, mnm^{5}U, is found in the wobble position 34. The amino acid-modified carbamoyl adenosine, (m^6)aa^{6}A (aa, amino acid), is present at position 37 in certain tRNAs. **b**, General RNA–peptide synthesis cycle based on mnm^{5}U and m^{6}aa^{6}A. The structures of oligonucleotides are simplified and only terminal nucleobases are drawn.

RNA donor strand, although it required duplex-enforcing high salt and low temperature conditions (1 M NaCl and 0 °C). The interaction of three nucleotides on the donor with the corresponding triplet on the acceptor seems to be the lower limit for productive coupling. Interestingly, this is the size of the codon–anticodon interaction in contemporary translation[11,18].

We next investigated coupling of the nitrile derivative of **1a** (m^{6}g$_{CN}$^{6}A, **1j**) with the different acceptors **2a**–**2c** under the recently described prebiotically plausible thiol activation conditions[45] (DTT, pH 8, 25 °C). Here also, the coupling products were obtained within a few hours (Fig. 2c). For example, the combination of nm^{5}U **2b** with **1a** gives coupling yields of 64% and 66% using EDC/Sulfo-NHS or DMTMM·Cl, respectively. Coupling of **1a** and **2a**, featuring a secondary amine, afforded **3a** in 16% and 33% yields. The nitrile of **1j** afforded yields of up to 65% after thiol activation coupling.

We next measured the stability of the hairpin-type intermediates. For the hairpin **3a** (Fig. 2a), a melting temperature (T_m) of approximately 87 °C was determined, which in comparison to the starting duplex (approximately 30 °C for **1a·2a**, see Supplementary Information), proves that the peptide formation reaction generated thermally more stable structures. This could have been an advantage during wet–dry cycling under early Earth conditions.

The discovered concept also enabled the synthesis of longer peptides. When we used 3′-vmnm^{5}U-RNA-5′ **2c** as the acceptor, we observed, on reaction with **1a**–**1j**, peptide bond formation with up to 77% yield (Fig. 2c, d and Fig. 3a).

We next studied the cleavage of the urea linkage and found that this reaction was possible at elevated temperatures (90 °C) in water at pH 6 (Fig. 2a, b). After 6 h, the products, m^{6}A-containing RNA **4** and RNA **5a** were formed already with a yield of 15%.

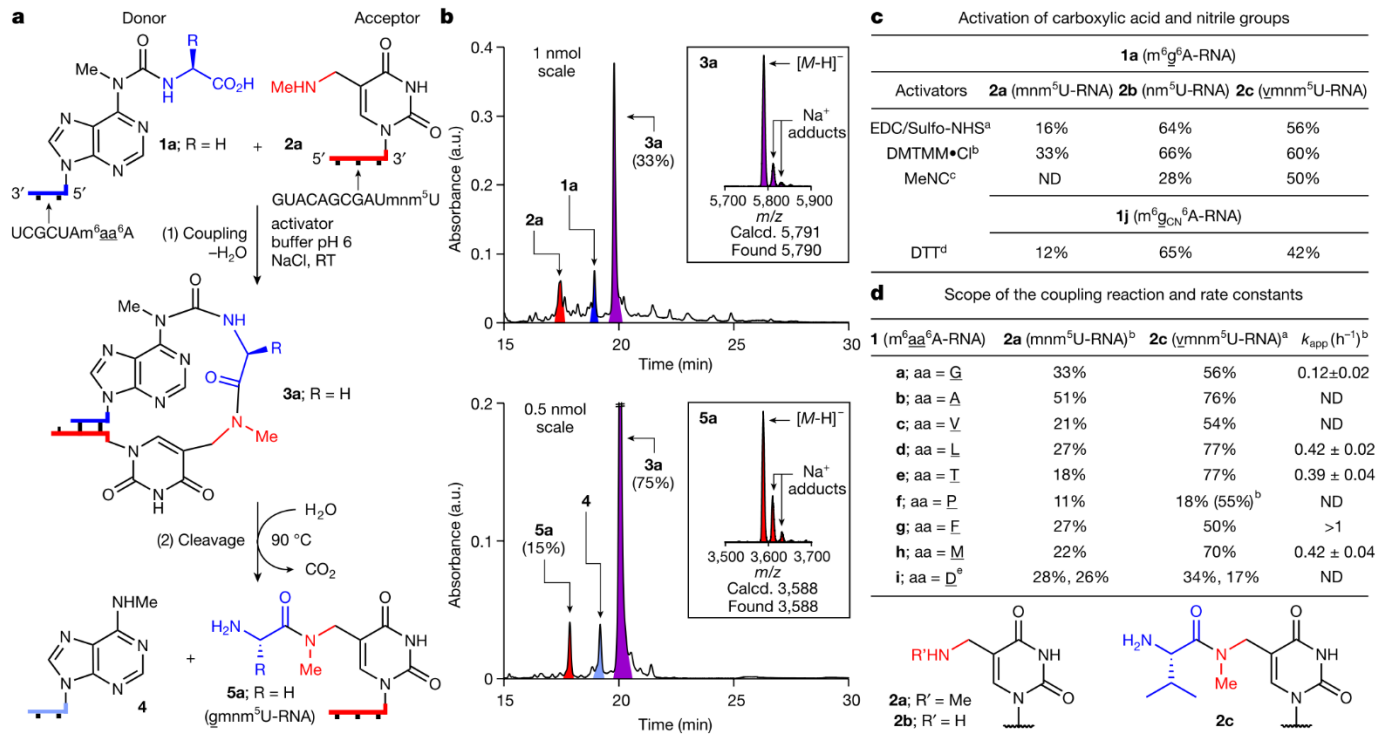
Longer peptide structures on RNA

We next investigated how the length of the generated peptides influences the coupling reaction (Fig. 3 and Extended Data Fig. 1). For this study we used synthetic 3′-peptide-mnm^{5}U-RNA-5′ acceptor strands as starting materials (Supplementary Information). The synthesized acceptor strands were hybridized to the donor strand **1a**. After carboxylic acid activation, rapid formation of elongated hairpin-type intermediates with yields between 40% and 60% was observed (Fig. 3b). We found that the coupling yields did not drop substantially with increasing peptide length, suggesting that other factors, such as the RNA hybridization kinetics, are rate limiting. In all cases, the subsequent urea cleavage (pH 4, 90 °C) affords dipeptide- to hexapeptide-decorated RNAs in 10–15% yield. These modest yields are the result of substantial RNA degradation, driven by the pH and temperature conditions that were used. The decomposition of RNA, however, can be overcome by using 2′-OMe nucleotides (see 'Stepwise growth of peptides on RNA'), which are also vestiges of the early RNA world[46].

During urea cleavage we detected competing formation of hydantoin side products[47], depending on the pH and temperature (Fig. 3a). Under mildly acidic conditions (pH 6, 90 °C), exclusive formation of the hydantoin product, cyclic-**5c**, was observed. Reducing the temperature and a shift to higher acidity (pH 4, 60 °C) led to the preferential formation of the peptide product, **5c** (approximately 7:1 **5c**:cyclic-**5c** ratio).

Fragment coupling on RNA

We investigated whether longer peptides can also be generated by fragment coupling chemistry with RNA donor strands containing an already longer peptide (m^{6}peptide^{6}A). This is essential because an RNA–peptide world, with initially low chemical efficiency, might have been limited to the synthesis of smaller peptides. We found that the required adenosine nucleosides, containing a whole peptide attached to the N^6-position, are available if the peptides that are produced by RNA degradation of the RNA–peptide chimeras, for example, can react with nitrosated N^6-methylurea adenosine (Fig. 4a). When we treated N^6-methylurea adenosine with $NaNO_2$ (5% H_3PO_4) and added the solution to triglycine (pH 9.5), we obtained the peptide-coupled adenosine nucleoside ggg^{6}A



c

Activators	**2a** (mnm^{5}U-RNA)	**2b** (nm^{5}U-RNA)	**2c** (vmnm^{5}U-RNA)
	1a (m^{6}g^{6}A-RNA)		
EDC/Sulfo-NHS[a]	16%	64%	56%
DMTMM•Cl[b]	33%	66%	60%
MeNC[c]	ND	28%	50%
	1j (m^{6}g$_{CN}$^{6}A-RNA)		
DTT[d]	12%	65%	42%

d

1 (m^{6}aa^{6}A-RNA)	**2a** (mnm^{5}U-RNA)[b]	**2c** (vmnm^{5}U-RNA)[a]	k_{app} (h^{-1})[b]
a; aa = G	33%	56%	0.12±0.02
b; aa = A	51%	76%	ND
c; aa = V	21%	54%	ND
d; aa = L	27%	77%	0.42 ± 0.02
e; aa = T	18%	77%	0.39 ± 0.04
f; aa = P	11%	18% (55%)[b]	ND
g; aa = F	27%	50%	>1
h; aa = M	22%	70%	0.42 ± 0.04
i; aa = D[e]	28%, 26%	34%, 17%	ND

Fig. 2 | Peptide synthesis on RNA with terminal (m)nm^{5}U and m^{6}aa^{6}A nucleotides. a, Reaction scheme for **1a** (5′-m^{6}g^{6}A-RNA-3′) and **2a** (3′-mnm^{5}U-RNA-5′) with coupling (1) and cleavage (2). **b**, HPLC chromatograms of the crude reaction mixtures, obtained after coupling of **1a** with **2a** using DMTMM•Cl (see reaction condition b) and cleavage of **3a** (100 mM MES buffer pH 6, 100 mM NaCl, 90 °C, 6 h). HPLC peaks of RNAs are coloured: donor in blue; acceptor in red; hairpin-type intermediate in purple; and cleaved donor strand in pale blue. The insets show MALDI-TOF data (negative mode) of the isolated products **3a** and **5a**. Calcd., calculated. **c**, Coupling results obtained with different activators for **1a** and **1j** with **2a**–**2c**. **d**, Coupling reactions with different donors **1a**–**1i** and acceptors **2a**,**2c**, and apparent rate constants (k_{app}) of selected coupling reactions with **2c**. All coupling reactions were carried out using a concentration of 50 µM for **1a**–**1j** and 50 µM for **2a**–**2c** (100 mM NaCl, 25 °C). [a]50 mM EDC/Sulfo-NHS (100 mM MES buffer pH 6, 24 h). [b]50 mM DMTMM•Cl (100 mM MES buffer pH 6, 24 h). [c]50 mM MeNC (50 mM DCI buffer pH 6, 5 days). [d]50 mM DTT (100 mM borate buffer pH 8, 24 h). [e]The two yields with **1i** (aa, D) describe the reaction of the aspartic acid α-COOH and of the side chain COOH. An assignment was not performed. RT, room temperature; ND, not determined.

in approximately 65% yield. Incorporation of (m^6)ggg^{6}A into RNA and hybridization of this donor strand with a 3′-ggvmnm^{5}U-RNA-5′ acceptor strand furnished, after coupling and urea cleavage, the RNA–peptide chimera 3′-gggggvmnm^{5}U-RNA-5′ (53% coupling, approximately 10% cleavage; Fig. 4b, left). We could also directly transfer longer peptides. When we hybridized the 5′-m^{6}gaggg^{6}A-RNA-3′ donor with the 3′-agggvmnm^{5}U-RNA-5′ acceptor, 3′-gaggggagggvmnm^{5}U-RNA-5′ was obtained as the product (56% coupling, approximately 9% cleavage; Fig. 4b, right). These experiments suggest the possibility of generating highly complex RNA–peptide chimeras with just a small number of reaction steps[48].

Multiple peptide growth on RNA

We next investigated whether peptide growth is possible at different RNA positions simultaneously. To this end, we examined the simultaneous binding of different donor strands to one or two acceptor strands. We hybridized two donor strands (7-mer: 5′-m^{6}g^{6}A-RNA-3′ and 10-mer: 5′-m^{6}v^{6}A-RNA-3′) to a single RNA acceptor strand (21-mer) with a central gmnm^{5}U and a 3′ terminal nm^{5}U (Fig. 5a, left). On activation of the carboxylic acids, a GG-dipeptide was synthesized in the centre of the RNA, whereas a valine amino acid was attached to the 3′ end of the acceptor strand. In a different experiment, we hybridized an RNA donor strand (22-mer), containing both a 3′-m^{6}g^{6}A and a 5′-m^{6}v^{6}A, to two different acceptor RNAs, containing a central vmnm^{5}U (21-mer) and a 3′ terminal vmnm^{5}U (11-mer) (Fig. 5a, right). On activation, we observed formation of a central GV- and a terminal VV-dipeptide.

Effect of base pairing

To investigate the importance of sequence complementarity, we added two RNA donor strands of different lengths (7-mer: 5′-m^{6}g^{6}A-RNA-3′ and 11-mer: 5′-m^{6}v^{6}A-RNA-3′) to an acceptor strand with a vmnm^{5}U at the 3′ end (11-mer: **2c**) (Fig. 5b, left). On the basis of the melting temperatures of the two possible duplexes (approximately 30 °C for the 7-mer·11-mer and 59 °C for the 11-mer·11-mer, see Supplementary Information), only formation of the VV-dipeptide RNA conjugate, derived from the thermodynamically more stable duplex, was observed. Finally, we mixed two RNA donor strands of identical length (7-mer). The first contained a 5′-m^{6}I^{6}A and the second a 5′-m^{6}g^{6}A, together with two mismatches. We added this mixture to an RNA acceptor strand (11-mer: **2c**) with a 3′-vmnm^{5}U nucleotide (Fig. 5b, right). In this experiment, exclusive formation of the LV-dipeptide was found, generated from the fully complementary strands and thus the more stable duplex. Collectively, these results support that full complementarity is needed for efficient peptide synthesis.

Stepwise growth of peptides on RNA

We finally investigated whether one-pot stepwise growth of a peptide on RNA is possible (Fig. 5c). To increase the stability of the RNA towards phosphodiester hydrolysis, as needed for this experiment, we used the RNA acceptor strand **2g**, in which the contemporary canonical bases were replaced by the non-canonical 2′-OMe nucleotides: A_m, C_m, G_m and U_m. The strand **2g** was equipped with an additional 3′-mnm^{5}U

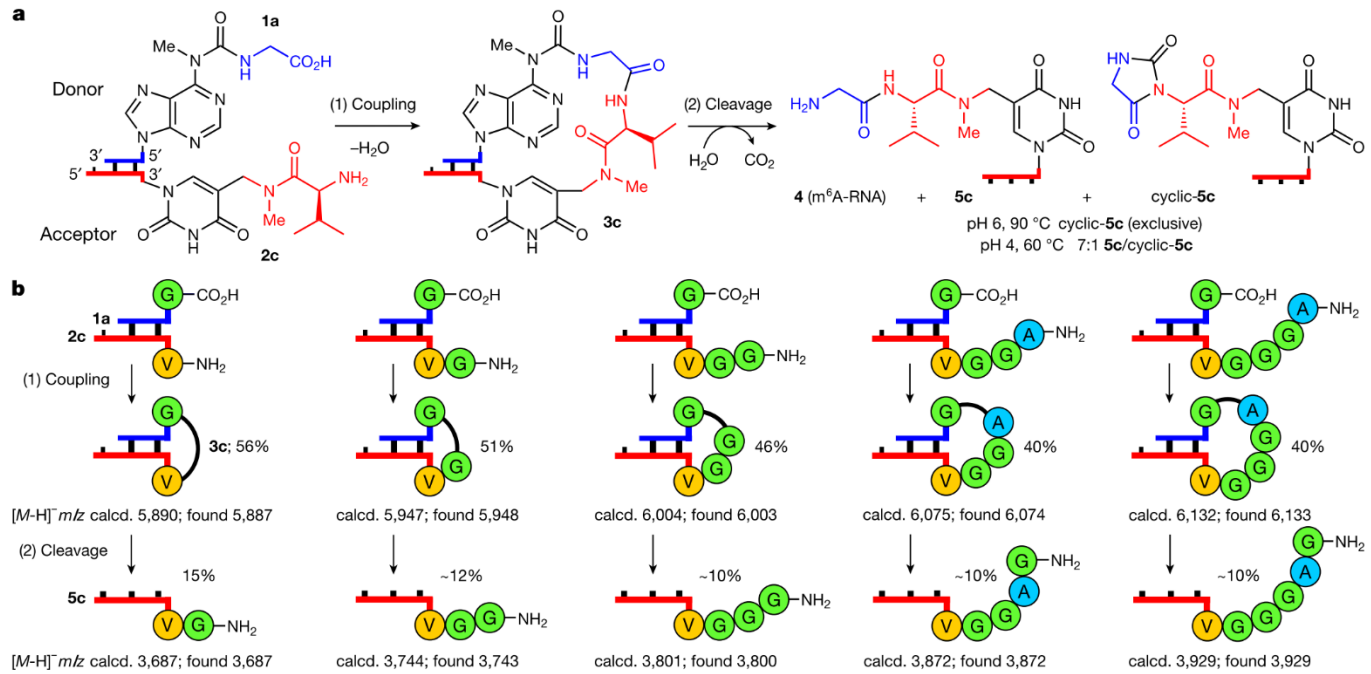


Fig. 3 | Growth of longer peptide structures on RNA. a, Scheme for the reaction of **1a** (5′-m^{6}g^{6}A-RNA-3′) with **2c** (3′-ymnm^{5}U-RNA-5′) including coupling (1) and cleavage (2). **b**, Coupling reactions between **1a** and RNA–peptide acceptor strands using EDC/Sulfo-NHS (see reaction condition a in Fig. 2) and cleavage reactions of the coupled compounds (100 mM acetate buffer pH 4, 100 mM NaCl, 90 °C, 6 h). MALDI-TOF data (negative mode) of the isolated products are given.

nucleotide. For the experiment we used the same amount of donor strand for all coupling steps and performed filtration steps to remove remaining activator. After two couplings, two urea cleavages and two filtrations, we observed, by high-performance liquid chromatography (HPLC) analysis, the presence of the product 3′-ggmnm^{5}U-RNA-5′ **7g** (Fig. 5c, left). The circumvented material consuming isolation steps

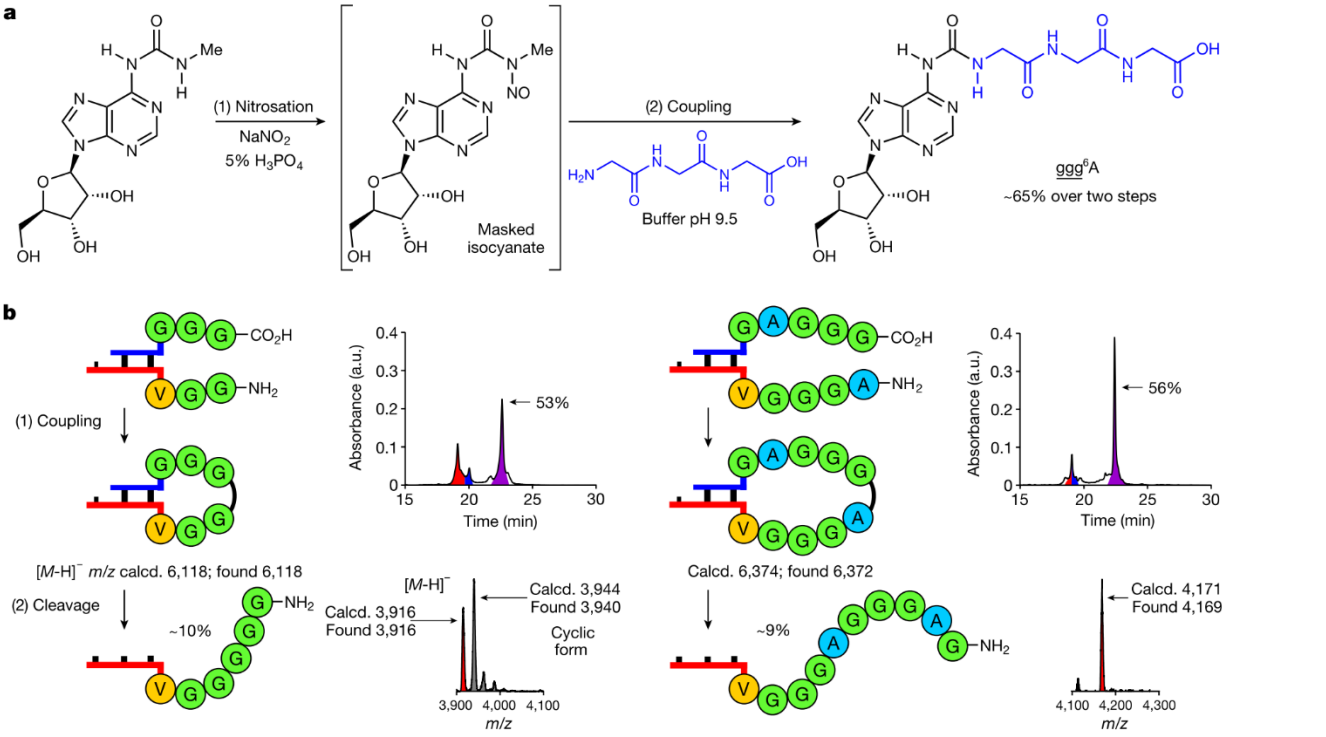


Fig. 4 | Capture of peptides by nitrosated N^6-methylurea adenosine for fragment condensation. a, Prebiotically plausible formation of peptide^{6}A structures, such as ggg^{6}A. **b**, Coupling reactions between RNA-peptide conjugates using EDC/Sulfo-NHS (see reaction condition a in Fig. 2) and cleavage reactions of the coupled compounds (see reaction conditions in Fig. 3). HPLC chromatograms show the crude mixtures of the coupling reactions. The RNA signals are coloured: donor in blue; acceptor in red; and hairpin-type intermediate in purple. MALDI-TOF data (negative mode) are shown for the isolated products, together with the 5′-m^{6}A-RNA-3′ strand **4** and the hydantoin side product (cyclic form) in the case indicated.

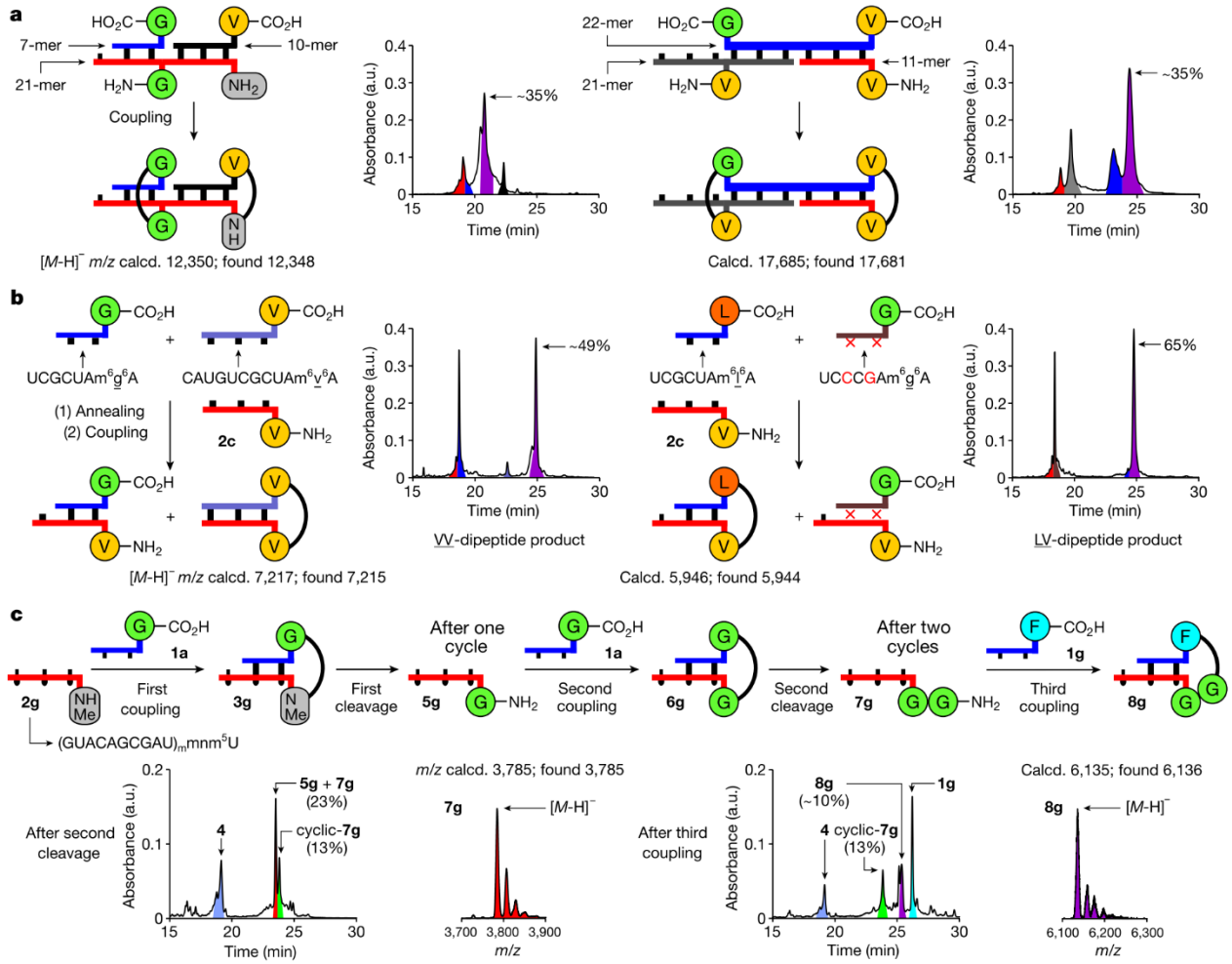


Fig. 5 | Parallel growth of peptides at various positions on RNA, effect of base pairing and RNA–peptide synthesis cycles. a, Coupling of oligonucleotides containing multiple donor or acceptor units (EDC/Sulfo-NHS, see reaction condition a in Fig. 2). **b**, Annealing followed by coupling (EDC/Sulfo-NHS, see reaction condition a in Fig. 2) of an acceptor strand with donor strands of different length (left) or base sequence (right). **c**, Two RNA–peptide synthesis cycles with a third coupling step using a 2′-OMe acceptor strand and performed under one-pot conditions with intermediary filtration to remove the remaining activator (coupling: DMTMM•Cl, see reaction condition b in Fig. 2; cleavage: 100 mM acetate buffer pH 4, 100 mM NaCl, 90 °C, 24 h; MES buffer pH 6 was used in the first cleavage reaction). HPLC chromatograms show the crude mixtures of the coupling and cleavage reactions. Peaks of RNA strands are coloured as in the reaction scheme. MALDI-TOF data (negative mode) of the isolated products are given.

(Extended Data Fig. 2) enabled us to obtain the product in an overall yield of about 18%. A final, third coupling reaction with the 5′-m^{6}f^{6}A donor strand **1g** furnished the FGG-hairpin intermediate **8g** in approximately 10% overall yield (Fig. 5c, right).

We next studied fragment condensation with the 5′-m^{6}ggg^{6}A-RNA-3′ donor strand and the complementary 3′-aggmnm^{5}U-RNA-5′ acceptor strand, consisting only of 2′-OMe nucleotides. Here, coupling with approximately 50% and urea cleavage with approximately 85% generated the product 3′-gggaggmnm^{5}U-RNA-5′, together with some of the hydantoin side product (Supplementary Information). Together these data show that, with the help of 2′-OMe nucleotides, peptides can grow on RNA in a stepwise fashion and via fragment condensation to generate higher complexity.

Discussion

The plausible formation of catalytically competent and self-replicating RNA structures without the aid of proteins is one of the major challenges for the model of the RNA world[1–4]. It is difficult to imagine how an RNA world with complex RNA molecules could have emerged without the help of proteins and it is hard to envision how such an RNA world transitions into the modern dualistic RNA and protein world, in which RNA predominantly encodes information whereas proteins are the key catalysts of life.

We found that non-canonical vestige nucleosides[8–12], which are key components of contemporary RNAs[6,7], are able to equip RNA with the ability to self-decorate with peptides. This creates chimeric structures, in which both chemical entities can co-evolve in a covalently connected form[13], generating gradually more and more sophisticated and complex RNA–peptide structures. Although, in this study, we observe peptide coupling on RNA in good yields, the efficiency will certainly improve if we allow optimization of the structures and sequences of the RNA–peptides by chemical evolution. The simultaneous presence of the chemical functionalities of RNA and amino acids certainly increases the chance of generating catalytically competent structures. The stabilization of RNA by incorporation of 2′-OMe nucleotides significantly improved the urea cleavage yield.

Interestingly, in the coupling step we observed large differences in the rate constants, which suggests that our system has the potential to preferentially generate certain peptides. We also found that peptides can simultaneously grow at multiple sites on RNA on the basis of rules determined by sequence complementarity, which is the indispensable requirement for efficient peptide growth.

All these data together support the idea that non-canonical vestige nucleosides in RNA have the potential to create peptide self-decorating RNAs and hence an RNA–peptide world. The formed RNA–peptide chimeras are comparatively stable, and so it is conceivable that some of these structures learned, at some point, to activate amino acids by adenylation[49] and to transfer them onto the ribose OH groups[50] to capture the reactivity in structures that were large and hydrophobic enough to exclude water. This would then have been the transition from the non-canonical nucleoside-based RNA–peptide world to the ribosome-centred translational process that is a hallmark of all life on Earth today.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04676-3.

1. Gilbert, W. Origin of life: the RNA world. *Nature* **319**, 618 (1986).
2. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393 (1968).
3. Crick, F. H. C., Brenner, S., Klug, A. & Pieczenik, G. A speculation on the origin of protein synthesis. *Orig. Life Evol. Biosph.* **7**, 389–397 (1976).
4. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).
5. Bowman, J. C., Hud, N. V. & Williams, L. D. The ribosome challenge to the RNA world. *J. Mol. Evol.* **80**, 143–161 (2015).
6. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem. Sci* **27**, 344–351 (2002).
7. Carell, T. et al. Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed. Engl.* **51**, 7110–7131 (2012).
8. Wong, J. T.-F. Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. *Orig. Life Evol. Biosph.* **21**, 165–176 (1991).
9. Di Giulio, M. Reflections on the origin of the genetic code: a hypothesis. *J. Theor. Biol.* **191**, 191–196 (1998).
10. 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).
11. Grosjean, H. & Westhof, E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* **44**, 8020–8040 (2016).
12. Beenstock, J. & Sicheri, F. The structural and functional workings of KEOPS. *Nucleic Acids Res.* **49**, 10818–10834 (2021).
13. Di Giulio, M. On the RNA world: evidence in favor of an early ribonucleopeptide world. *J. Mol. Evol.* **45**, 571–578 (1997).
14. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572 (2002).
15. Fox, G. E. Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.* **2**, a003483 (2010).
16. Bowman, J. C., Petrov, A. S., Frenkel-Pinter, M., Penev, P. I. & Williams, L. D. Root of the tree: the significance, evolution, and origins of the ribosome. *Chem. Rev.* **120**, 4848–4878 (2020).
17. Eigen, M. & Schuster, P. A principle of natural self-organization. *Naturwissenschaften* **64**, 541–565 (1977).
18. Szathmáry, E. Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proc. Natl Acad. Sci. USA* **90**, 9916–9920 (1993).
19. Noller, H. F. RNA structure: reading the ribosome. *Science* **309**, 1508–1514 (2005).
20. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9**, 242–253 (2008).
21. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **1**, 127–136 (2003).
22. Woese, C. The universal ancestor. *Proc. Natl Acad. Sci. USA* **95**, 6854–6859 (1998).
23. Becerra, A., Delaye, L., Islas, S. & Lazcano, A. The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annu. Rev. Ecol. Evol. Syst.* **38**, 361–379 (2007).
24. Kuhn, H. Self-organization of molecular systems and evolution of the genetic apparatus. *Angew. Chem. Int. Ed. Engl.* **11**, 798–820 (1972).
25. Kuhn, H. & Waser, J. Molecular self-organization and the origin of life. *Angew. Chem. Int. Ed. Engl.* **20**, 500–520 (1981).
26. Tamura, K. & Schimmel, P. Oligonucleotide-directed peptide synthesis in a ribosome- and ribozyme-free system. *Proc. Natl Acad. Sci. USA* **98**, 1393–1397 (2001).
27. Tamura, K. & Schimmel, P. Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proc. Natl Acad. Sci. USA* **100**, 8666–8669 (2003).
28. Turk, R. M., Chumachenko, N. V. & Yarus, M. Multiple translational products from a five-nucleotide ribozyme. *Proc. Natl Acad. Sci. USA* **107**, 4585–4589 (2010).
29. Jash, B., Tremmel, P., Jovanovic, D. & Richert, C. Single nucleotide translation without ribosomes. *Nat. Chem.* **13**, 751–757 (2021).
30. 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. Engl.* **54**, 9871–9875 (2015).
31. Becker, S. et al. Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Commun.* **9**, 163 (2018).
32. Tetzlaff, C. N. & Richert, C. Synthesis and hydrolytic stability of 5′-aminoacylated oligouridylic acids. *Tetrahedron Lett.* **42**, 5681–5684 (2001).
33. Schweizer, M. P., McGrath, K. & Baczynskyj, L. The isolation and characterization of *N*-[9-(*β*-D-ribofuranosyl)-purin-6-ylcarbamoyl]glycine from yeast transfer RNA. *Biochem. Biophys. Res. Commun.* **40**, 1046–1052 (1970).
34. Perrochia, L. et al. *In vitro* biosynthesis of a universal t^6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.* **41**, 1953–1964 (2012).
35. Kimura-Harada, F., Von Minden, D. L., McCloskey, J. A. & Nishimura, S. *N*-[(9-*β*-D-Ribofuranosylpurin-6-yl)-*N*-methylcarbamoyl]threonine, a modified nucleoside isolated from *Escherichia coli* threonine transfer ribonucleic acid. *Biochemistry* **11**, 3910–3915 (1972).
36. Robertson, M. & Miller, S. Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. *Science* **268**, 702–705 (1995).
37. Murphy, F. V., Ramakrishnan, V., Malkiewicz, A. & Agris, P. F. The role of modifications in codon discrimination by $tRNA^{Lys}_{UUU}$. *Nat. Struct. Mol. Biol.* **11**, 1186–1191 (2004).
38. Kitamura, A. et al. Characterization and structure of the *Aquifex aeolicus* protein DUF752: a bacterial tRNA-methyltransferase (MnmC2) functioning without the usually fused oxidase domain (MnmC1). *J. Biol. Chem.* **287**, 43950–43960 (2012).
39. Hutchby, M. et al. Hindered ureas as masked isocyanates: facile carbamoylation of nucleophiles under neutral conditions. *Angew. Chem. Int. Ed. Engl.* **48**, 8721–8724 (2009).
40. Ohkubo, A. et al. New thermolytic carbamoyl groups for the protection of nucleobases. *Org. Biomol. Chem.* **7**, 687–694 (2009).
41. Nainytė, M. et al. Amino acid modified RNA bases as building blocks of an early Earth RNA-peptide world. *Chem. Eur. J.* **26**, 14856–14860 (2020).
42. Schimpl, A., Lemmon, R. M. & Calvin, M. Cyanamide formation under primitive Earth conditions. *Science* **147**, 149–150 (1965).
43. Gartner, Z. J., Kanan, M. W. & Liu, D. R. Expanding the reaction scope of DNA-templated synthesis. *Angew. Chem. Int. Ed. Engl.* **41**, 1796–1800 (2002).
44. Liu, Z. et al. Harnessing chemical energy for the activation and joining of prebiotic building blocks. *Nat. Chem.* **12**, 1023–1028 (2020).
45. Foden, C. S. et al. Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).
46. Schneider, C. et al. Noncanonical RNA nucleosides as molecular fossils of an early Earth—generation by prebiotic methylations and carbamoylations. *Angew. Chem. Int. Ed. Engl.* **57**, 5943–5946 (2018).
47. Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **41**, 5416–5429 (2012).
48. Bondalapati, S., Jbara, M. & Brik, A. Expanding the chemical toolbox for the synthesis of large and uniquely modified proteins. *Nat. Chem.* **8**, 407–418 (2016).
49. Berg, P. The chemical synthesis of amino acyl adenylates. *J. Biol. Chem.* **233**, 608–611 (1958).
50. Wu, L.-F., Su, M., Liu, Z., Bjork, S. J. & Sutherland, J. D. Interstrand aminoacyl transfer in a tRNA acceptor stem-overhang mimic. *J. Am. Chem. Soc.* **143**, 11836–11842 (2021).

Methods

General method for the peptide coupling reactions

The RNA donor and acceptor strands (1:1 ratio, 5 nmol of each strand) were annealed with NaCl (5 µl from a 1 M aqueous solution) by heating at 95 °C for 4 min, followed by cooling down slowly to room temperature. After that, MES buffer pH 6 (25 µl from a 400 mM aqueous solution) and NaCl (5 µl from a 1 M aqueous solution) were added to the oligonucleotide solution. Finally, carboxylic acid or nitrile activator/s (10 µl of each component from a 500 mM aqueous solution) and water (100 µl of total reaction volume) were added to the solution mixture. The peptide coupling reaction was incubated at 25 °C for 24 h. The crude reaction mixtures were analysed by HPLC and MALDI-TOF mass spectrometry.

General method for the urea cleavage reactions

The hairpin-type intermediate (0.5 nmol) was diluted with MES buffer pH 6 or acetate buffer pH 4 (12.5 µl from a 400 mM aqueous solution), NaCl (5 µl from a 1 M aqueous solution) and water (50 µl of total reacion volume). The urea cleavage reaction was incubated at 60–90 °C at different time intervals. The crude reaction mixtures were analysed by HPLC and MALDI-TOF mass spectrometry.

Data availability

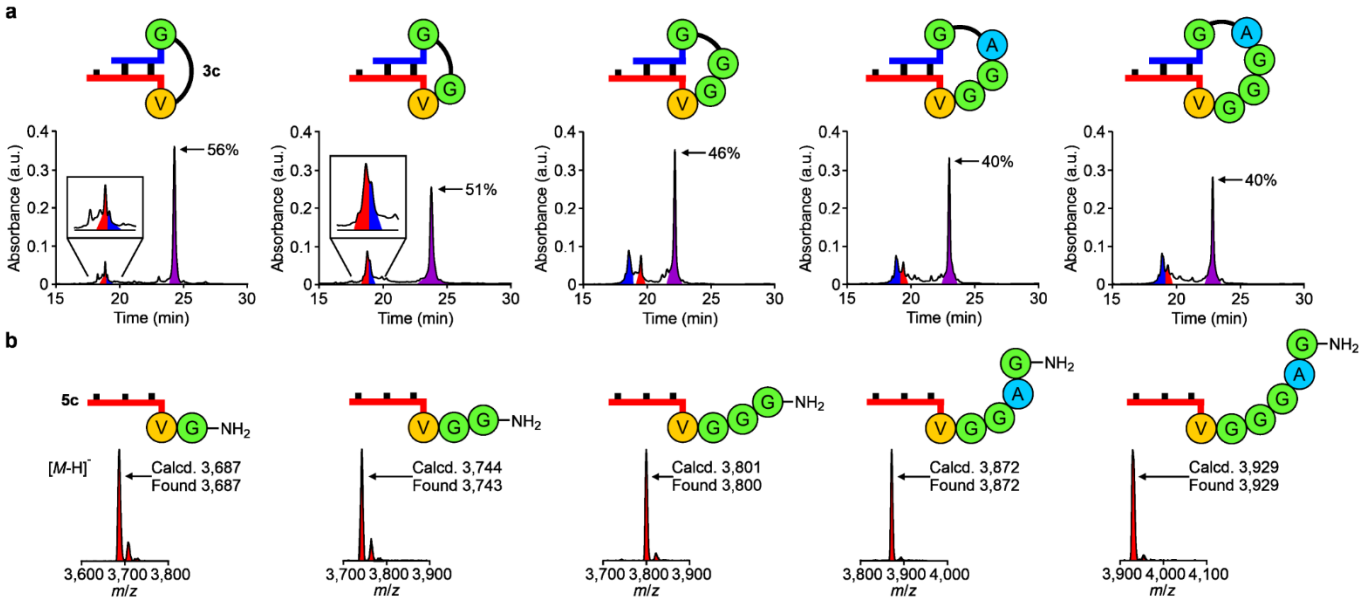
The data that support the findings of this study are available within the paper and its Supplementary Information.

Acknowledgements We thank the Deutsche Forschungsgemeinschaft for supporting this research through the DFG grants: CA275/11-3 (ID: 326039064), CRC1309 (ID: 325871075, A4), CRC1032 (ID: 201269156, A5) and CRC1361 (ID: 393547839, P2). We thank the Volkswagen Foundation for funding this research (grant EvoRib). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 741912 (EPiR) and under the Marie Skłodowska-Curie grant agreement no. 861381 (Nature-ETN). L.E. thanks the Alexander von Humboldt Foundation for a postdoctoral fellowship (ESP 1214218 HFST-P).

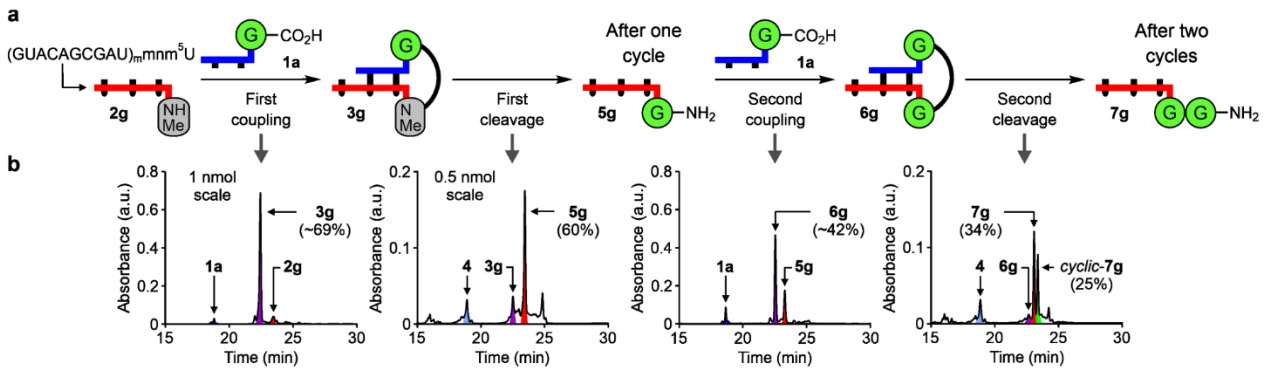
Author contributions F.M., L.E., F.X. and E.W. synthesized the modified phosphoramidites and RNA strands and performed the peptide coupling and urea cleavage experiments. M.N. synthesized RNA donor strands and performed preliminary experiments. T.A. refined and developed mechanistic concepts and performed initial proof-of-principle studies. C.-Y.C. and A.P. synthesized modified phosphoramidites. T.C. conceived the project and directed the research. All authors contributed to the analysis of the results and writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04676-3.
Correspondence and requests for materials should be addressed to Thomas Carell.
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Analytical data of the growth of longer peptides on RNA. a, HPLC chromatograms show the crude mixtures of the coupling reactions (100 mM MES buffer pH 6, 100 mM NaCl, 50 mM EDC/Sulfo-NHS, 25 °C, 24 h) between 5′-m^{6}g^{6}A-RNA-3′ **1a** and RNA-peptide acceptor strands. **b**, MALDI-TOF mass spectra (negative mode) are shown for the isolated products obtained after the cleavage reactions (100 mM acetate buffer pH 4, 100 mM NaCl, 90 °C, 6 h) of the coupled compounds. In the HPLCs, the RNA strands are coloured: donor in blue; acceptor in red and hairpin-type intermediate in purple.



Extended Data Fig. 2 | RNA-peptide synthesis cycles using a 2′-OMe acceptor strand. a, Two RNA-peptide synthesis cycles in which the product of each step was separated and added into the next reaction (coupling conditions: 100 mM MES buffer pH 6, 100 mM NaCl, 50 mM DMTMM•Cl, 25 °C, 24 h; cleavage conditions: 100 mM acetate buffer pH 4, 100 mM NaCl, 90 °C, 24 h). **b**, HPLC chromatograms show the crude mixtures of the coupling and cleavage reactions. In the HPLCs, peaks of RNA strands are coloured as in the reaction scheme. The product 3′-ggmnm^{5}U-RNA-5′ **7g** was obtained in ≈ 6% overall yield.

3.2. Isoxazol-Nukleoside als Bausteine für eine plausible Proto-RNA

„Isoxazole Nucleosides as Building Blocks for a Plausible Proto-RNA"

Felix Xu, Antony Crisp, Thea Schinkel, Romeo C. A. Dubini, Sarah Hübner, Sidney Becker, Florian Schelter, Petra Rovó, Thomas Carell*

Angew. Chem. Int. Ed., **2022**, 61, e202211945.

Für Hintergrundinformationen siehe Anhang II

Prolog

Aufgrund der strukturellen Komplexität des RNA-Moleküls ist es unwahrscheinlich, dass es in einem einzigen Evolutionsschritt gebildet wurde. Es wurde postuliert, dass zunächst Proto-RNAs entstanden sind, die aus einfacheren Molekülen gebildet wurden, welche mit größerer Wahrscheinlichkeit bereits auf der frühen Erde vorhanden waren. In diesem Manuskript wird gezeigt, das Isoxazol-Nukleobasen, die ein zentrales Zwischenprodukt der präbiotischen Pyrmidin-Synthese sind,[99] die Fähigkeit besitzen mit G *Watson-Crick*-Paarung einzugehen. Darüber hinaus können diese Proto-Nukleobasen sich innerhalb des RNA-Stranges zu Cytidin umlagern. Die stereochemische Konfiguration des anomeren Zentrums am neu-geformten Nukleosid, kann durch die Stereochemie der benachbarten Nukleotide gesteuert werden. Diese Ergebnisse zeigen, dass Isoxazol-Riboside potentielle Bausteine einer Proto-RNA sein könnten.

Autorenbeitrag

Für diese Arbeit war ich für die Entwicklung und Durchführung der Synthesen für die Phosphoramidit-Bausteine zuständig. Außerdem führte ich die Entwicklung der RNA-Festphasen Synthese Methoden zur Synthese der Stränge **S1**-**S8**, die Planungen der Experimente, den Verdau der RNA-Stränge mit anschließender massenspektrometrischer Analyse, die Messungen der Schmelzkurven durch. Des Weiteren war ich für die Auswertung der Ergebnisse und die Mitverfassung des Manuskripts zuständig.

Lizenz

Prebiotic Chemistry

How to cite: *Angew. Chem. Int. Ed.* **2022,** *61,* e202211945
International Edition: doi.org/10.1002/anie.202211945
German Edition: doi.org/10.1002/ange.202211945

Isoxazole Nucleosides as Building Blocks for a Plausible Proto-RNA

*Felix Xu, Antony Crisp, Thea Schinkel, Romeo C. A. Dubini, Sarah Hübner, Sidney Becker, Florian Schelter, Petra Rovó, and Thomas Carell**

Abstract: The question of how RNA, as the principal carrier of genetic information evolved is fundamentally important for our understanding of the origin of life. The RNA molecule is far too complex to have formed in one evolutionary step, suggesting that ancestral proto-RNAs (first ancestor of RNA) may have existed, which evolved over time into the RNA of today. Here we show that isoxazole nucleosides, which are quickly formed from hydroxylamine, cyanoacetylene, urea and ribose, are plausible precursors for RNA. The isoxazole nucleoside can rearrange within an RNA-strand to give cytidine, which leads to an increase of pairing stability. If the proto-RNA contains a canonical seed-nucleoside with defined stereochemistry, the seed-nucleoside can control the configuration of the anomeric center that forms during the in-RNA transformation. The results demonstrate that RNA could have emerged from evolutionarily primitive precursor isoxazole ribosides after strand formation.

All concepts that have been developed to explain how RNA oligonucleotides could have formed on the early Earth are based on the prebiotic formation of nucleosides, which oligomerize in an enzyme free reaction.[1–6] These concepts are based on the idea that nucleosides must have initially formed in prebiotic reactions, after which they were phosphorylated[7–10] and subsequently oligomerized (Figure 1a).[2,11–15] This scenario requires complex multi-step prebiotic reactions to generate the nucleosides as a prerequisite for RNA formation.[16] The complexity of this scenario has early on led to the idea that RNA emerged through a process of chemical evolution from a much simpler proto-RNA (first ancestor of RNA) precursor molecule.[17–22] Along this line, it is thinkable that proto-RNA existed that was made up from proto-nucleosides (precursor nucleosides), which rearrange inside the proto-RNA to RNA. It is also thinkable that these proto-nucleosides formed a mixed proto-RNA with a few embedded canonical RNA nucleosides as seed-nucleotides.[23] The proto-RNA nucleosides could then rearrange to finally give RNA strands composed exclusively of canonical nucleotides (Figure 1a). Degradation of this rearranged RNA, which would certainly have

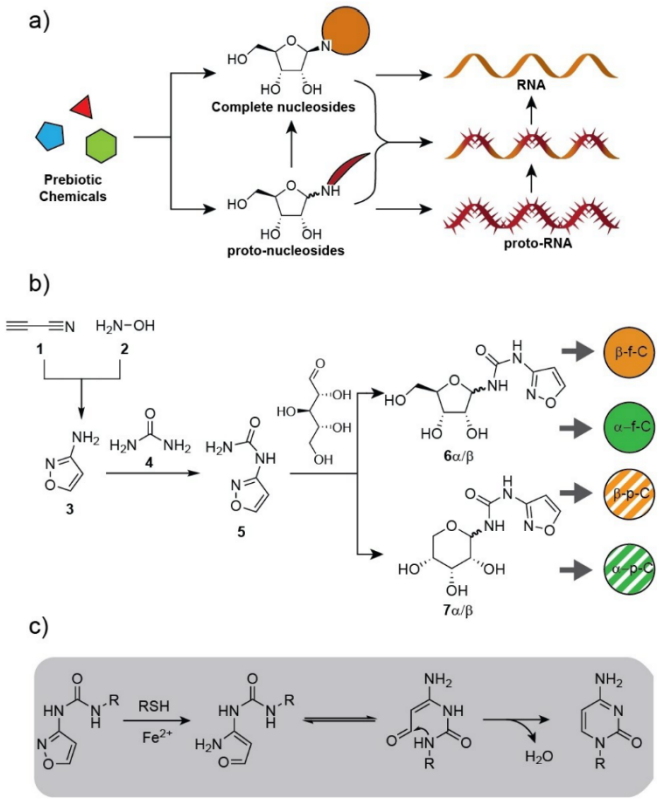


Figure 1. a) Classical prebiotic chemistry provides first the full nucleosides, which are then converted into RNA. Alternatively, proto-nucleosides could have formed a proto-RNA or a mixed RNA-proto-RNA from which RNA was formed. b) Synthesis of the 3-urea isoxazole nucleosides **6** and **7**. Upon insertion into RNA, **6**α/β rearranges directly in RNA to give cytidine. c) Mechanism of the reaction.

[*] MSc. F. Xu, Dr. A. Crisp, MSc. T. Schinkel, MSc. R. C. A. Dubini, MSc. S. Hübner, Dr. S. Becker, MSc. F. Schelter, Dr. P. Rovó, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstr. 5–13, 81377 Munich (Germany)
E-mail: Thomas.carell@lmu.de
Homepage: http://www.carellgroup.de

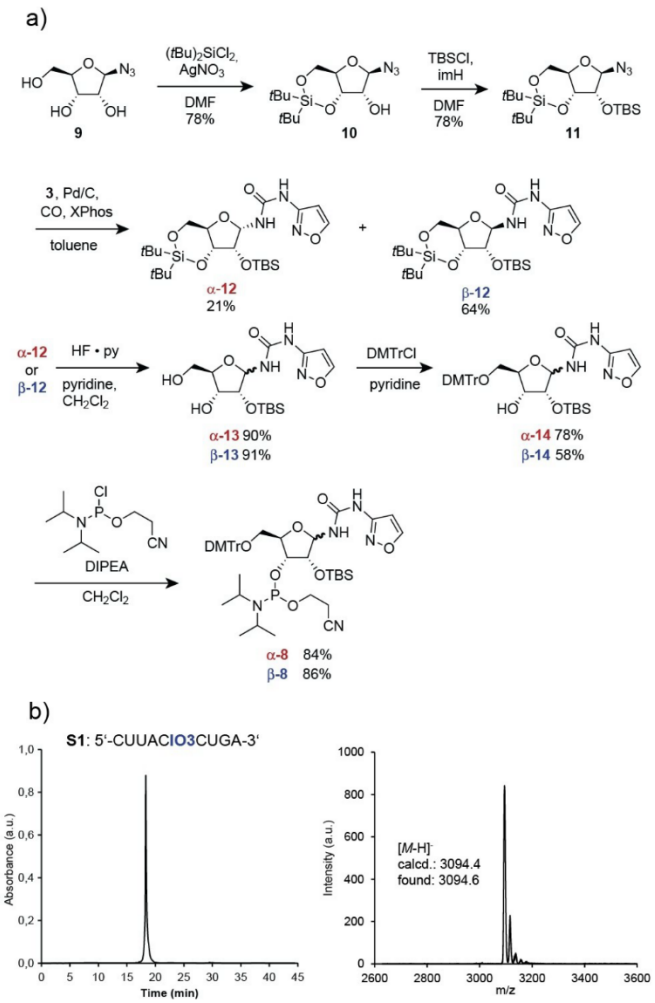
Dr. S. Becker
Current address: Max Planck Institute of Molecular Physiology, Otto-Hahn-Straße 11, 44227 Dortmund (Germany)

Dr. P. Rovó
Current address: Institute of Science and Technology Austria (ISTA),
Am Campus 1, 3400 Klosterneuburg (Austria)

occurred on the early earth, would have liberated the canonical nucleosides, thus allowing them to be reinserted as seed-nucleotides into growing proto-RNA. Here we show that such a concept can indeed give RNA and we provide evidence that the stereo information of the canonical seed-nucleotides can control the stereochemical outcome of the proto-RNA to RNA rearrangement. An obstacle for the concept is that in most precursor RNA nucleosides the glycosidic bonds are too labile, so that stable strand formation cannot occur. This is a particular problem for non-cyclic precursor nucleosides, which typically feature an acidic anomeric H-atom at the glycosidic N-atom. Exceptions from this are, however, urea-based proto-nucleosides, in which the acidic H-atom is stabilized by an internal H-bond.[24]

A recently introduced concept (Figure 1b) of how the pyrimidine nucleoside cytidine could have formed on the early Earth is based on the reaction of cyanoacetylene (**1**) with hydroxylamine (**2**) to give 3-aminoisoxazole (IO3, **3**).[25] **3** was shown to react with urea (**4**) to give *N*-isoxazolyl-urea **5**, which can next react with ribose to furnish the corresponding furanosides (**6**α,β) and pyranosides (**7**α,β). Subsequent reductive N–O bond cleavage with thiols in the presence of catalytic amounts of Fe^{2+} gives cytidines (Figure 1c).[25–28] Here, we report that the initially formed *N*-isoxazolyl-urea **5**, if structurally embedded into (RNA)-proto-RNA structures, can undergo the N–O cleavage induced rearrangement directly in RNA. This "in-RNA" reaction creates a new stereocenter. Excitingly, we observe that the configuration of already present canonical seed-nucleosides controls the stereochemical outcome of the "in-RNA" cyclization reaction, which leads to an amplification of stereochemical information.

To first investigate the possibility of performing the **6**α/β to C rearrangement directly in-RNA, we prepared the IO3 phosphoramidites **α-8** and **β-8** incorporated both building blocks into the different RNA strands shown in (Table 1). The synthesis of **8** is depicted in Scheme 1. It started with 1-azidoribose **9**, which was first 3′-5′-protected with



Scheme 1. a) Synthesis of the IO3 phosphoramidite α,β-**8** and conditions. b) HPL-chromatograms and MALDI-TOF mass spectra of the oligonucleotide **S1** shown as an example, synthesized with the IO3 phosphoramidite **β-8.**

Table 1: Synthesized oligonucleotides with **α,β-8** as needed for this study and their molecular weights determined by MALDI-TOF mass spectrometry.

Sequence name	Sequence (5′→3′)	Calculated $[M-H]^-$	Found (*m/z*)
S1	5′-CUU AC$_{\beta}$**IO3** CUG A-3′	3094.4	3094.6
A-S1	5′-CUU ACA CUG A-3′	3102.4	3102.3
G-S1	5′-CUU ACG CUG A-3′	3118.4	3118.3
C-S1	5′-CUU ACC CUG A-3′	3078.7	3078.3
U-S1	5′-CUU ACU CUG A-3′	3079.4	3079.5
A-R1	5′-UCA GAG UAA G-3′	3205.5	3205.4
G-R1	5′-UCA GGG UAA G-3′	3221.5	3221.2
C-R1	5′-UCA GCG UAA G-3′	3181.5	3181.8
U-R1	5′-UCA GUG UAA G-3′	3182.4	3182.1
S2	5′-$_{\beta}$C$_{\beta}$C**IO3** $_{\beta}$C$_{\beta}$C**IO3** $_{\beta}$C$_{\beta}$C-3′	2409.4	2409.3
S3	5′-$_{\alpha}$C$_{\alpha}$C**IO3** $_{\alpha}$C$_{\alpha}$C**IO3** $_{\alpha}$C$_{\alpha}$C-3′	2409.4	2409.1
S4	5′-$_{\alpha}$C$_{\beta}$C**IO3** $_{\alpha}$C$_{\beta}$C**IO3** $_{\alpha}$C$_{\beta}$C-3′	2409.4	2408.6
S5	5′-**IO3IO3IO3 IO3**-3′	1222.2	1221.9
S6	5′-**IO3IO3IO3 IO3**$_{\beta}$C-3′	1526.2	1526.2
S7	5′-**IO3IO3IO3 IO3**$_{\alpha}$C-3′	1526.2	1526.2
S8	5′-GGU $_{\beta}$**IO3**GA CC3′	2538.4	2535.5

$(tBu)_2SiCl_2$ (DMF, rt, 78 %) to give **10**. Subsequent protection of the 2′-OH group with TBS-Cl (DMF, imidazole, rt, 78 %) provided **11**. This was followed by a one-pot reduction of the azide to the amine (Pd/C, cat. XPhos, toluene, CO, 5 bar, 60 °C) and reaction of the amine with IO3 **3** to generate the ribofuranoside **12** as the expected α/β-mixtures. The two IO3 anomers **α-12** and **β-12** could be separated by flash column chromatography. The nucleosides were next independently 3′-5′-deprotected to **13** (HF-py, CH_2Cl_2 0 °C) and 5′-dimethoxytrityl-protected (DMT-Cl, pyridine, rt) to give **14**. The 5′-DMT-2′-TBS-protected IO3 ribosides **14** were finally converted into the phosphoramidites **α-8** and **β-8**. While we were initially concerned about the stability of the isoxazole-ribosides, we learned during the synthesis that all IO3 nucleosides are astonishingly stable. Similar to the previously reported urea nucleosides,[24] we noticed no deglycosylation reaction during the reaction, which is surprising given that the anomeric center features a rather acidic NH-group.

To investigate whether an incorporation of the IO3 nucleosides into RNA is possible, we used standard solid phase oligonucleotide chemistry and the phosphoramidites **α-8** and **β-8**. The RNA synthesis was indeed possible without large adjustments of the RNA-synthesis and deprotection protocol. We just had to extend the coupling time of **α-8** and **β-8** to 600 s. Deprotection was first achieved with ammonium hydroxide/methylamine (1:1) followed by a second deprotection step with TEA·3HF. Scheme 1b shows the HPLC-chromatogram of the synthesized RNA strand **S1** as an example containing a single β-configured IO3 after purification. The MALDI-TOF mass spectrum proves the integrity of the obtained RNA **S1**. With this experimental proof for the possible and efficient incorporation of the IO3 nucleoside into RNA, we next used the protocol to prepare the RNA strands **S1**–**S8** containing multiple IO3 nucleosides in mixed sequence contexts.

To investigate the "in-RNA"-rearrangement of the IO3 nucleosides to C and to study the stereochemical outcome of the reaction, we first prepared the RNA strand **S2** (Figure 2a) containing two 3-isoxazole nucleosides embedded into a 8mer containing 6 additional cytidines. For the synthesis, we used the phosphoramidite mixture α/β-**8**. To investigate the α/β-ratio after RNA synthesis and HPL-chromatographic purification of the strand, we digested **S2** after the synthesis down to the individual nucleosides with the New England BioLabs Nucleoside digestion mix (Cat. No. NEB M0649S). The obtained mixture of nucleosides was subsequently analyzed by LC-HESI-MS (gray bracket in Figure 2a). Clearly evident is the presence of the β-C nucleosides, as expected. Besides β-C, we detected two additional signals for the two isoxazole nucleosides (α-IO3,β-IO3). Integration of the signals shows an α/β-ratio of 1:4. Next, we treated the isoxazole-containing RNA strand **S2** with dithiothreitol (DTT) and catalytic amounts of Fe^{2+} at 90 °C for 2 h. HPL-chromatographic analysis of the transformation revealed a clean spot-to-spot conversion (Figure 2a). While the original RNA strand **S2** had a retention time of 33 min, this signal fully disappeared in favor of a new HPLC-signal at 23 min, which corresponds to the reduced product strand $\mathbf{S2^{red}}$. Important is the observation that only one signal is detected and not a set of signals as expected for a diastereoisomeric mixture of products. With the purpose of investigating the stereochemical outcome, we again digested the RNA strand $\mathbf{S2^{red}}$ and analyzed

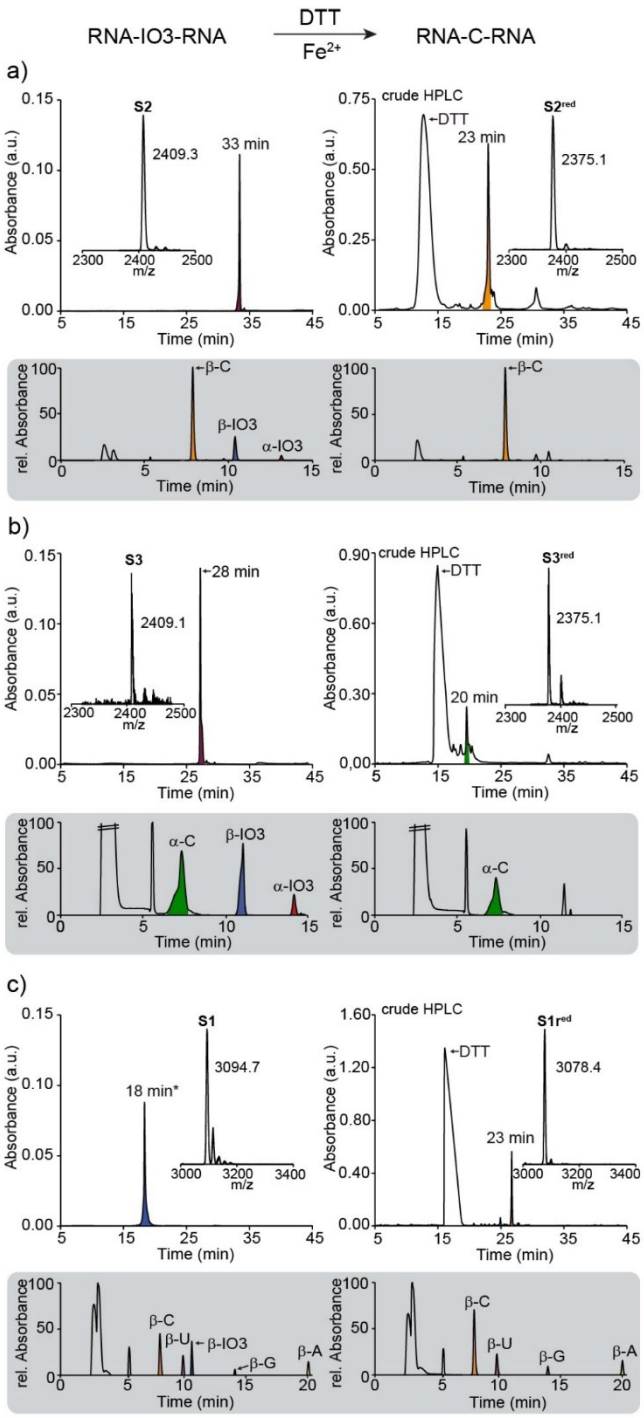


Figure 2. Reaction of the **IO3** containing strands with DTT and Fe^{2+} and HPLC, MALDI, HPLC-MS (at 223 nm; gray brackets) analyses of the N–O cleavage and cyclization reaction (left: starting material; right: product). a) Result of the β-homo-C strand **S2**; b) the α-homo-C strand **S3**; c) the strand **S1**. A = adenosine, C = cytidine, G = guanosine, IO3 = 3-aminoisoxazole, U = uridine. *HPLC gradient: 0–40 % buffer B.

the nucleoside mixture by LC-HESI-MS (Figure 2b). This experiment showed to our surprise only one signal for the compound β-C. Co-injection of a α-C confirmed this experimental outcome. α-C is clearly not formed during the N–O bond cleavage and rearrangement to C in the RNA strand. This result shows that the α/β-mixture of the IO3 nucleoside in the RNA rearranged exclusively to β-C by "amplification" of the stereochemical information provided by the seed-cytidines.

In order to substantiate this stereochemical amplification step, we next prepared the RNA strand **S3**, in which we replaced the flanking β-Cs of **S2** by the corresponding α-Cs. The RNA strand **S3** was again synthesized in excellent quality (Figure 2b). Digestion and LC-HESI-MS analysis before the N–O-bond cleavage shows again the presence of the IO3 nucleosides as an α/β-mixture. When we opened the N–O bonds of **S3** and analyzed the nucleoside mixture after digestion, we observed to our surprise again only a single HPLC signal. This time only α-C was detected, showing that now the "in-RNA"-rearrangement of the α/β-mixture of the IO3 furnished only the α-configured product. Again, the stereo-information of the anomeric center of the seed-α-C nucleosides must have dictated the stereochemical outcome of the cyclization reaction (Figure 2b). We next performed the same reaction with strand **S1** to show that the stereochemical amplification is also achieved with other canonical bases. As expected, we only detected β-configured C in our product strand $\mathbf{S1^{red}}$ (Figure 2c). These data show that the flanking bases obviously transfer stereochemical information, so that the cyclization reaction proceeds with high stereospecificity. We believe that the information transfer is caused by an optimization of the base stacking forces. A stereochemically homogeneous all-α or all-β RNA can establish better base stacking forces in an aqueous solution.

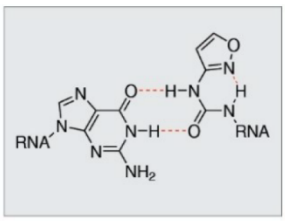
With the purpose of proving that the presence of seed-nucleosides is a prerequisite for the stereochemical amplification, we performed two control experiments. First, we used the strand **S4**, which is a 8mer oligonucleotide containing both α- and β-seed-C in a 1:1 ratio. Proto-RNA **S5** in contrast was devoid of any seed-cytidines. After treatment of these strands with DTT and Fe^{2+}, subsequent LC-HESI-MS analysis showed as expected the formation of a 1:1 mixtures of α/β-Cs (see Supporting Information).[29]

Based on this result, we next designed two 5mers with only one information carrying nucleoside in the form of a single β-C (**S6**) or α-C **(S7)** at the 3′ position. LC-HESI-MS analysis of the digestion products after N–O bond cleavage and subsequent rearrangement showed us now mixtures of α/β-C for both strands. However, a clear stereochemical bias is observed. For the β-C carrying strand $\mathbf{S6^{red}}$ an α/β-ratio of 1:3 is detected and for the α-C carrying strand $\mathbf{S7^{red}}$ an α/β-ratio of 2:1 was obtained. This shows that one stereochemically defined seed-nucleotide is already sufficient to control the stereochemical outcome of the cyclization reaction.

For a chemical compound to function as an instructional proto-nucleoside it is essential that the nucleoside can contribute to the stability of the RNA and that it can encode information.[30,31] Therefore, we investigated if the nucleoside β-IO3 is able to establish informative interactions with one of the other RNA bases. For this study, we synthesized different counter strands for **S1** (**A**/**C**/**G**/**U-R1**) in which we varied the base opposite the IO3 and measured melting points. We also prepared four reference **S1** strand (**A**/**C**/**G**/**U-S1**), in which we replaced the IO3 by one of the other canonical bases (Figure 3a). The duplex with the G:C base pair (**G-S1**:**C-R1**) at the respective position features as expected, the highest melting point of 55 °C. The A:U base pair at this position created a duplex (**A-S1**:**U-R1**) which melted at $T_m = 48$ °C. For the duplex **G-S1**:**U-R1** with a G:U Wobble base pair at the respective position, we measured a melting point of 43 °C. When we investigated the **IO3-S1**:**R1** base pair, we noted that the IO3:G situation created a duplex **IO3-S1**:**G-R1** with a melting point of 45 °C even higher than what was determined for the G:U Wobble base pair. More importantly, the pairing selectivity is remarkable. While the IO3:G containing duplex melts at 45 °C, all other base pairs created duplexes with melting points of 31 °C or even lower, establishing a ΔT_m of minimum 14 °C. This is an unusual discrimination between the productive IO3:G situation and all other base pairing possibilities.

In order to investigate the reason for the constructive interaction of the IO3:G base pair, we designed two palindromic 8mer RNA strands **S8** and $\mathbf{Can_8}$ (see Supporting Information), which upon annealing result in formation of either G:U or IO3:G wobble base pairs at the central two positions in each duplex, respectively (GGU**X**GACC). These strands were prepared in mg-quantities and the formed duplexes were analyzed by NMR spectroscopy (Figure S7). The high chemical shift similarity observed across the two samples, both in homo- and heteronuclear experiments, showed us that both palindromes form double-strands with an overall A-like conformation which is typical for RNA duplexes (Figure S7–S11, Table S1). 2D ^{1}H–^{1}H NOESY was next employed to elucidate the conformation of the IO3 base (Figure 3b, see Supporting Information for extended methods). The most probable orientation was determined by analysis of the cross peaks in the imino region. A considerable upfield shift of the G-5H1 proton can be observed (from 12.1 to 11.0 ppm), indicating the presence of a weak hydrogen bond of G-5 to IO3–4 (Figure S11). The intense cross peak between G-5H1 and IO3-4H3 (purple line in Figure 3b), which could be detected even at a short mixing time of 40 ms, shows that the two bases face each other. This observation implies that IO3-4H3 points in the direction of the second strand and hence away from the sugar phosphate backbone. A cross peak between protons G-5H1 and IO3-4H8 confirms that the isoxazole-ring of IO3–4 point towards G-5. A cross peak with the sugar proton U-3H2′ (orange line in Figure 3b) furthermore suggests that the IO3-4H1 protrudes towards the sugar phosphate backbone, similarly to H8 protons in adenine and guanine or the H5 protons of cytosine or uracil. Finally, the strong cross peaks between IO3-4H7 and the aromatic protons of U-3 (green lines in Figure 3b) and the weaker cross peaks of IO3-4H8 to these protons indicate that IO3-4 and U-3 are stacked, arguing that the aromatic ring is oriented in the direction of the major groove.

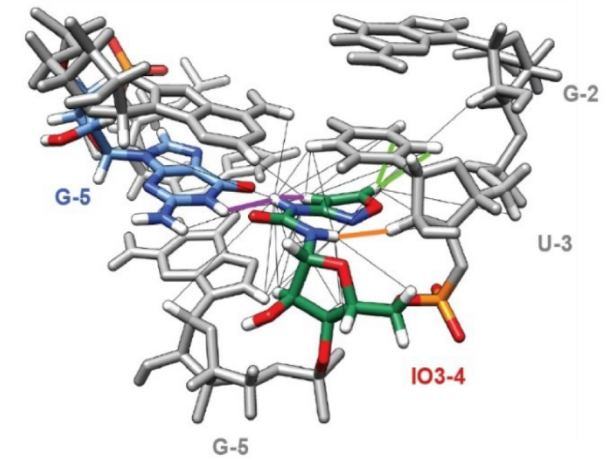
a)



Oligo 1: 5'-CUUACYCUGA-3'
Oligo 2: 3'-GAAUGZGACU-5'

Oligo 1 (Y)	Oligo 2 (Z)	T_m (°C)
G-S1	C-R1	55
A-S1	U-R1	48
G-S1	U-R1	43
IO3-S1	G-R1	45
IO3-S1	A-R1	31
IO3-S1	C-R1	25
IO3-S1	U-R1	27

b)



c)

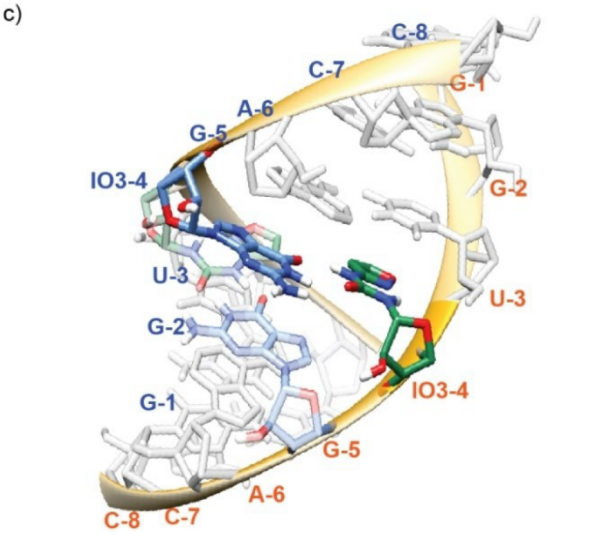


Figure 3. a) Chemical structure and base pairing properties of a IO3:G base pair and summary of T_m analyses for canonical oligonucleotides and oligonucleotides containing IO3. Solutions were buffered with 10 mM sodium phosphate (pH 7) and 150 mM NaCl. b) Predicted orientation of the IO3 with all the NOE contacts (black lines) and the most important contacts for the structure determination highlighted (IO3-4H3:G5H1 (purple), IO3-4H1:U3H2′ (orange), IO3-4H7:U3H5 and IO3-4H7:U3H6 (green)). c) Structural model of **S8**, showing the non-canonical base pairing between G-5 (blue) and IO3–4 (green).

This leads to an orientation such that IO3-4H1 can form an intramolecular hydrogen bond with IO3-4O4. Such an arrangement has the carbonyl group pointing towards the guanine base, allowing, in line with the melting point studies, a productive base-pairing interaction with G-5H1. Overall, this analysis establishes that IO3 adopts a flat conformation in which the NH-CO-unit of the base pairs points towards the G counter base to form constructive H-bonds. It is the urea substructure, which establishes H-bonds with the G counter base. The fact that only two H-bonds can form, is consistent with the lower melting point. The 2D ^{1}H-^{1}H NOESY data were further used to perform structure calculations (Figure 3c). The low-energy structural model obtained for the modified 8mer shows that the IO3:G base-pair is sterically perfectly embedded within the double-stranded structure. Interesting is the fact that the internal H-bond between the isoxazole N and the urea N–H is strongly stabilizing the structure. The IO3-nucleoside forms 2 hydrogen bond with the G counter base, which upon rearrangement to C increases to 3, which goes in hand with a further stabilization of the base pair. This could have been the evolutionary driving force.

In summary, we show that isoxazole-urea nucleosides that are formed under prebiotically plausible conditions from just cyanoacetylene, hydroxylamine, urea and ribose[25] can form a base pair with G that has a stability comparable to a U:G Wobble base pair, based on two instructive H-bonds (Figure 3a). More importantly, the IO3:G base pair forms rather selectively because all other combinations (IO3:A/C/U) are substantially weaker. We show that the N–O bond of the isoxazole nucleoside can be rapidly cleaved within a proto-RNA strand with thiols and catalytic amounts of Fe^{2+}. This cleavage induces a cyclization-elimination cascade (Figure 1c) that gives β- or α-cytidine. Importantly, the presence of α- or β-seed-nucleosides defines the stereochemical outcome of the rearrangement reaction.

We do not know whether such IO3-nucleosides were once components of an early proto-RNA world, but it is generally assumed that RNA was preceded by simpler precursor structures.[16,30] Here, we show that isoxazole nucleosides are excellent candidates for such precursor nucleosides and proto-RNAs. Because the proto-nucleosides can rearrange to C also just in solution, it is reasonable to assume that IO-nucleosides and the canonical α-/β-nucleosides might have co-existed to form proto-RNA strands containing a few canonical seed cytidines. These nucleosides may then have controlled the stereochemical outcome of the rearrangement that converted the G:IO3 base pair into the better pairing G:C base pair. Because α- and β-Cs form with the same probability, our data provide no solution of the question of how homo-chirality was established.[32] The data shows, however, that stereochemical information could have been amplified by "in-RNA reactions" of precursor nucleosides to canonical nucleosides and this can stimulate thought about the origin of homo-chirality.

Acknowledgements

We thank Stefan Wiedemann for the synthesis of reference compounds and Pia Heinrichs for assistance in the NMR measurements of the oligonucleotides. We also thank Dr. Luis Escobar and Jonas Feldmann for valued discussions. This work was supported by the German Research Founda-

tion (DFG) for financial support via CRC1309 (Project ID 325871075, A04), CRC1361 (Project ID 893547839, P02) and CRC1032 (Project ID 201269156, A5). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 741912 (EpiR). We are grateful for additional funding from the Volkswagen Foundation (EvoRib). Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: Isoxazoles · Origin of Life · Prebiotic Chemistry · Proto-RNA · RNA

[1] W. Gilbert, *Nature* **1986**, *319*, 618–618.
[2] O. Leslie E, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.
[3] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381–393.
[4] D. M. Fialho, T. P. Roche, N. V. Hud, *Chem. Rev.* **2020**, *120*, 4806–4830.
[5] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**, *120*, 4766–4805.
[6] M. W. Powner, J. D. Sutherland, J. W. Szostak, *Synlett* **2011**, 1956–1964.
[7] R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 64–66.
[8] R. Reimann, G. Zubay, *Orig. Life Evol. Biosph.* **1999**, *29*, 229–247.
[9] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei, S. A. Benner, *Angew. Chem. Int. Ed.* **2016**, *55*, 15816–15820; *Angew. Chem.* **2016**, *128*, 16048–16052.
[10] 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; *Angew. Chem.* **2016**, *128*, 13443–13447.
[11] J. Morávek, *Tetrahedron Lett.* **1967**, *8*, 1707–1710.
[12] W. Huang, J. P. Ferris, *J. Am. Chem. Soc.* **2006**, *128*, 8914–8919.
[13] D. Deamer, *Chem. Soc. Rev.* **2012**, *41*, 5375–5379.
[14] C. Gibard, S. Bhowmik, M. Karki, E.-K. Kim, R. Krishnamurthy, *Nat. Chem.* **2018**, *10*, 212–217.
[15] C. Briones, M. Stich, S. C. Manrubia, *RNA* **2009**, *15*, 743–749.
[16] G. F. Joyce, *Nature* **2002**, *418*, 214–221.
[17] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Chem. Biol.* **2013**, *20*, 466–474.
[18] N. V. Hud, *Nat. Commun.* **2018**, *9*, 5171.
[19] A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821–12844.
[20] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4398–4402.
[21] A. E. Engelhart, N. V. Hud, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002196.
[22] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat. Commun.* **2016**, *7*, 11328.
[23] J. P. Dworkin, A. Lazcano, S. L. Miller, *J. Theor. Biol.* **2003**, *222*, 127–134.
[24] H. Okamura, A. Crisp, S. Hübner, S. Becker, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2019**, *58*, 18691–18696; *Angew. Chem.* **2019**, *131*, 18864–18869.
[25] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76.
[26] G. Wächtershäuser, *Microbiol. Rev.* **1988**, *52*, 452–484.
[27] E. Camprubi, S. F. Jordan, R. Vasiliadou, N. Lane, *IUBMB Life* **2017**, *69*, 373–381.
[28] M. Kijima, Y. Nambu, T. Endo, *J. Org. Chem.* **1985**, *50*, 1140–1142.
[29] Reaction of **S6** with an equimuolar amount of a complementary strand furnished a α/β-ratio of 1:1 (see Figure S7).
[30] F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367–379.
[31] S. Halder, D. Bhattacharyya, *Prog. Biophys. Mol. Biol.* **2013**, *113*, 264–283.
[32] M. Bolli, R. Micura, A. Eschenmoser, *Chem. Biol.* **1997**, *4*, 309–320.

Manuscript received: August 12, 2022
Accepted manuscript online: September 5, 2022
Version of record online: October 11, 2022

3.3. Eine Isoxazol-basierte Proto-RNA

„An Aminoisoxazole-Based Proto-RNA"

Felix Xu†, Stefan Wiedemann†, Jonas Feldmann, Sidney Becker, Thomas Carell*

ChemistryEurope **2023**, e202300057.

† Beitrag der Autoren zu gleichen Teilen

Für Hintergrundinformationen siehe Anhang III

Prolog

Die RNA-Welt-Hypothese besagt, dass das Leben mit der Entwicklung der replizierenden und katalytisch aktiven RNA begann. Allerdings wurde postuliert, dass der RNA-Welt eine Zeit vorrausging, in der einfachere Strukturen als Hauptträger der genetischen Information fungierten. In dieser Arbeit wurde gezeigt, dass die Konstitutionsisomere 3-Aminoisoxazol und 5-Aminoisoxazol potentielle Proto-Nukleobasen sind. Beide Verbindungen können unter präbiotisch plausiblen Bedingungen ausgehend von den gleichen Startmaterialien gebildet werden. Darüber hinaus können die Proto-Nukleobasen, wenn sie in RNA inkorporiert wurden, durch eine intramolekulare Reaktionskaskade selektiv zu den kanonischen Pyrimidin-Nukleobasen C und U reagieren. Proto-RNAs, die auf Aminoisoxazol-Strukturen basieren, können als ein mögliches Modell für den Übergang von der Proto-RNA-Welt zu der RNA-Welt dienen, das kein Austausch der Basen-Komponente erfordert.

Autorenbeitrag

Für diese Arbeit war ich für die Entwicklung und Durchführung der Synthesen der Phosphoramidite α-/β-**13** zuständig. Außerdem war ich für die Synthese der Oligonukleotid-Stränge, für die Planung der Experimente, Verdau der RNA-Stränge und für die Auswertung der Ergebnisse zuständig. Weiterhin half ich bei der Erstellung des Manuskripts. *Stefan Wiedemann* entwickelte und optimierte die präbiotischen Synthesen, synthetisierte die Massenspektrometrischen Standards und entwickelte die Analytik Methoden zur Auswertung der präbiotischen Reaktionen.

Lizenz

ChemistryEurope
Excellence in Chemistry Research Worldwide

Research Article
doi.org/10.1002/ceur.202300057

www.journal.chemistry-europe.org

An Aminoisoxazole-Based Proto-RNA

Felix Xu[+],[a] Stefan Wiedemann[+],[a] Jonas Feldmann,[a, b] Sidney Becker,[a, c] and Thomas Carell*[a]

The RNA world hypothesis predicts that life started with the development of replicating and catalytically active RNA, which evolved in a process of molecular evolution to increasingly complex chemical structures. RNA is, however, so complex that it has most likely formed from a precursor (proto-RNA) that was more easily accessible in a prebiotic world. Recently, 3-aminoisoxazoles (IO3) were identified as building blocks that can form under prebiotic conditions and can rearrange to give the nucleoside cytidine (C). The present study shows that the constitutional isomer 5-aminoisoxazole (IO5) can undergo the same reaction to give uridine (U). Both compounds (IO3 and IO5), if embedded in RNA, react selectively to C and U, which are the main pyrimidine nucleosides of the genetic system. Importantly, the stereochemical outcome of the IO5 reaction in RNA depends on the neighboring bases. If they are β-configured RNA nucleosides, the reaction proceeds with high selectivity to give exclusively the β-configured U RNA base (anomeric control).

Introduction

RNA plays a central role in our concepts of how life has started because the molecule is capable of catalyzing reactions, and has the potential for self-replication.[1–7] These two characteristics likely served as the foundation for the development of increasingly complex molecules and chemical systems through chemical evolution, ultimately paving the way for the emergence of life as we know it.[1,8] The difficulty with the concept is that the prebiotically plausible reaction pathways that have been postulated to give the canonical RNA nucleosides, as the key building blocks for RNA, are complex and it is hard to conceive how RNA could have formed in the first place.[9] It is consequently a major scientific goal to find molecular proto-RNA structures that could have preceded RNA and to investigate how such structures could have evolved into RNA.[10–15]

We recently introduced the concept of urea-RNA, in which ribofuranose with a triuret, connected to the C1-position of ribose, acts as a proto-RNA. In these nucleosides, the triuret unit was able to establish productive H-bonds with the counter base guanin.[16] In addition, we showed that 3-aminoisoxazoles (IO3), connected via a urea linkage to the C1-atom of ribose are highly efficient precursors of cytidine (C). The needed *N*-isoxazol-3-yl-urea was shown to be readily accessible under prebiotically plausible conditions.[17] The compound embedded as a riboside in RNA was able to perform the cascade reaction to C (Figure 1A).[18] Although this chemistry provides access to C, the second important pyrimidine nucleoside of the genetic code U is in this scenario only available via the hydrolysis of C. This, however, is a slow reaction particularly if the C is embedded in RNA.[19,20]

In order to solve this problem and to find a common prebiotically plausible route to both pyrimidines C and U, we investigated the starting point of the IO-based chemistry, namely the reaction of cyanoacetylene (**1**) with hydroxylamine (**2**) in greater detail. As previously shown, 3-aminoisoxazole could have formed by reacting cyanoacetylene (**1**) with hydroxylamine (**2**) under strongly basic (pH ≥ 14) conditions or

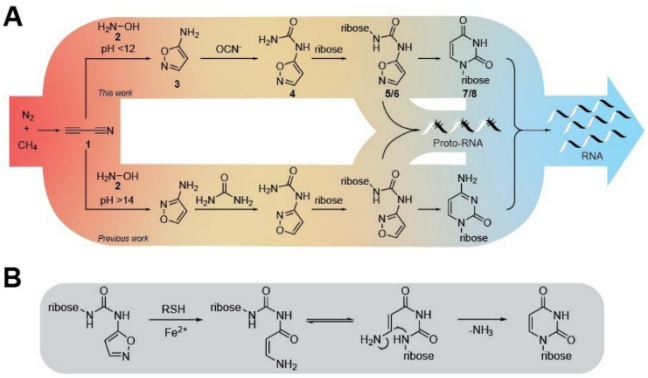


Figure 1. A) Prebiotically plausible formation of cytidine (C) and uridine (U) following the reaction of cyanoacetylene (**1**) with hydroxylamine (**2**) by reaction with the O-nucleophile of hydroxylamine (bottom)[17] and by reaction with the N-nucleophile of hydroxylamine (top). B) Putative cascade reaction that allows IO5 to react to uridine.

[a] F. Xu,[+] Dr. S. Wiedemann,[+] Dr. J. Feldmann, Dr. S. Becker, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-Universität (LMU) München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@lmu.de

[b] Dr. J. Feldmann
Current address:
Department of Chemistry
University of Cambridge
Lensfield Road, Cambridge CB2 1EW (UK)

[c] Dr. S. Becker
Current address:
Department of Chemical Biology
Max Planck Institute of Molecular Physiology
Otto-Hahn-Strasse 11, 44227 Dortmund (Germany)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under https://doi.org/10.1002/ceur.202300057

Research Article
doi.org/10.1002/ceur.202300057

with hydroxylurea under slightly basic conditions (pH~10).[17] The reaction of **1** (considered to be a prebiotically plausible starting material[21–24]) with **2**[17,25] under strongly basic conditions, favours the reaction of the O-atom of **2** with **1**, which then starts the chemical trajectory to C via 3-aminoisoxazole (Figure 1A). If, however, the reaction is performed at ambient pH-values, we reasoned that it might allow the N-atom of **2** to react first which could open a pathway to U (Figure 1A) by a cascade reaction shown in Figure 1B.

The beauty of this concept lies in the possibility that both canonical pyrimidine nucleosides C and U could be traced back to the same starting materials, with just the pH-value of the first reaction dictating whether C or U is formed. This would provide conceptually ideal conditions for the prebiotic origin of the genetic code with a focus on the pyrimidine nucleosides.

Results and Discussion

Formation of uridine from 5-aminoisoxazole

In order to investigate the concept, we reacted cyanoacetylene (**1**) with hydroxylamine (**2**) at pH-values below 14, particularly at slightly basic pH values of around pH = 9–10. Indeed, under these conditions we observed exclusively the nucleophilic attack of the amino group of **2** with **1**. The reaction was performed in water (1 eq. $NaHCO_3$, 25 °C, 24 h). Upon dry down precipitation of the pure 5-aminoisoxazole (**3**) may occur, because it is a solid material in contrast to 3-aminoisoxazole, which was found to be a liquid. Washing of the remaining material with water provided NMR-pure **3** (Figure 2A). This reaction is easily conceivable in a connected pond model with one wet-dry cycle and rain showers,[26,27] which would have allowed the washing away of residual **1** and **2** along with potential side products (Figure 2B). Next, we treated the precipitated material **3** with cyanate,[28–30] which might have flooded into the pond from a second pond. We now detected a clean and exclusive formation of *N*-isoxazol-5-yl-urea (**4**), which again precipitated upon drying down. Another rain shower may have again washed away residual cyanate. Inflow of an aqueous solution of ribose[31–33] from a third pond into the reaction pond followed by another dry-down step leads to the reaction of **4** with ribose. When we performed the reaction, we indeed detected the clean formation of all four expected reaction products namely the two α- and β-ribopyranosides (α-/β-**5**) and the two α- and β-ribofuranosides (α-/β-**6**) as a mixture . Under our conditions the α-furanoside α-**6** was formed as the dominant reaction product with a yield of 39%. When we started the envisioned cascade cyclization reaction (Figure 1B) by addition of a thiol (DTT) and catalytic amounts of Fe^{2+} to the obtained α- and β-**6**, we indeed detected efficient reductive N–O-bond cleavage followed by a clean cyclization[34] to give the U-ribopyranosides (α-/β-**7**) and the U-ribofuranosides (α-/β-**8**, β-**8** = U). To our surprise, we noted that now the β-furanoside was by far the dominating product. The yield of the reaction is with 31% for all ribosides and with 25% for the β-furanoside U (β-**7**) astonishingly high, given the simplicity of the reaction sequence. The high yield of the β-furanoside U together with the remarkable stereoselectivity of the reaction make us believe that the IO-pathway to C and U provides a good model of how

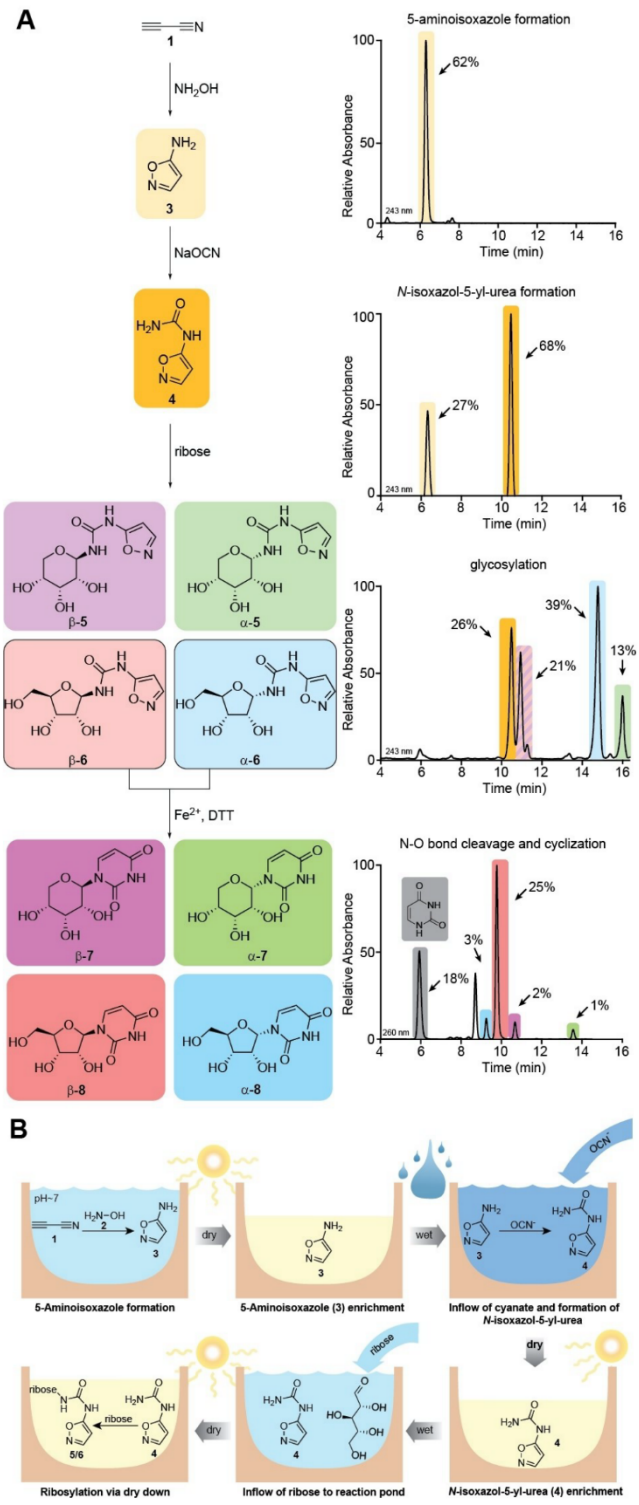


Figure 2. A) The reaction sequence that generates the pyrimidine U via the IO pathway with the corresponding HPL-chromatographic analyses of the reaction outcome showing clean conversions. B) Depiction of the dry-down connected pond model for the putative prebiotic formation of the IO5-ribosides.

Research Article
doi.org/10.1002/ceur.202300057

the pyrimidine nucleoside could have formed under Early Earth's condition. As hypothesized, in this scenario, it is the pH-value of the initial step that determines whether the pathway delivers the C- or the U-nucleoside.

Formation of C and U from isoxazoles directly in RNA

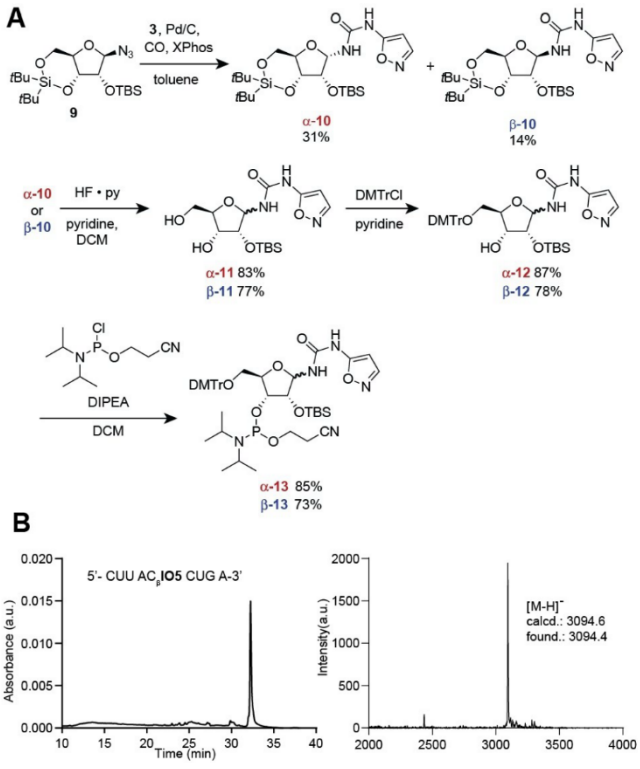
We recently showed that the nucleoside C can form directly in RNA from embedded IO3 ribofuranosides, which allowed us to show that IO3-precursor RNAs are an extant model for a proto-RNA that could have preceded canonical RNA.[18] In order to investigate if this direct "in-RNA" transformation is also possible with IO5 to generate U and to study if C and U can form simultaneously from IO3 and IO5-containing proto-RNA, we prepared the IO3-phosphoramidite building blocks for solid phase RNA synthesis. This was achieved following a synthetic pathway outlined in Scheme 1A. The starting point of the reaction was the well-known 3′-5′-protected riboside-azide **9**, which we reacted with 5-aminoisoxazole (**3**) in a Pd-catalyzed CO insertion reaction to directly give the IO5 riboside **10** as a mixture of the α- and the β-anomers. The anomers were subsequently separated by flash column chromatography. Next, we proceeded separately with the α- and the β-isomers. Deprotection of the 5′-3′ silyl protecting group with HF in pyridine gave **11**, followed by dimethoxy-tritylation of the 5′OH group to **12**. Subsequent conversion of **12** into the phosphoramidite **13** furnished the needed building block for solid phase oligonucleotide chemistry. Because we performed the reactions with the two anomers in parallel, we obtained both phosphoramidites β-**13** and α-**13**.

We subsequently used the building blocks β- and α-**13** to perform and optimize the RNA synthesis. We found that both building blocks (0.1 M in THF) are best incorporated into RNA using an elongated coupling time (600 s) and shortened detritylation time (40 s). Deprotection was first achieved with ammonium hydroxide (55 °C, 5 h) followed by a second deprotection step with TEA·3HF (65 °C, 2.5 h). The result of the synthesis of an RNA strand having β-**13** incorporatetd is depicted in Scheme 1B. The HPL-chromatogram of the purified strand 5′-CUUAC-$_{\beta}$5IO-CUGA-3′ and the corresponding MALDI-TOF mass spectrum are depicted in Scheme 1B. The data proves the structural integrity of the RNA strand with the embedded IO5 nucleoside.

To study the "in-RNA" N–O-bond cleavage and cyclization of IO5 to U and to learn about the stereochemical outcome of the reaction, we prepared the RNA strand 5′-CUAAC-$_{\beta}$IO5-CUGA-3′ using β-**13** (Figure 3A). This oligonucleotide was digested to the nucleoside level and the obtained nucleoside composition was analyzed by HPLC-MS which revealed the presence of all expected nucleosides β-C, β-U, β-G, β-A and of the additional unit β-IO5. Next, we treated the oligonucleotide with DTT and catalytic amounts of Fe^{2+}. This led to the formation of a new oligonucleotide with a slightly changed retention time. Digestion of this oligonucleotide provided the expected nucleoside mixture, but now without the β-IO5 nucleoside, arguing that the conversion of β-IO5 to uridine had happened as expected in the RNA strand.

We next investigated if two β-IO5 nucleosides would also be converted to U and in addition we wanted to get better proof for the conversion of β-IO5 to U and about the stereoselectivity. To answer these questions, we prepared the U-and IO5-only RNA strand using a mixture of α-**13** and β-**13** and we isolated the RNA strand with exactly one β-IO5 and one α-IO5 (Figure 3B). Digestion of this RNA strand provided as expected one signal for β-U and one each for the α- and β-IO5 nucleoside. Upon treatment with DTT and Fe^{2+}, a new strand with a shifted retention time was again formed in a clean reaction. Upon digestion, this strand gave only one signal for β-U, showing that both compounds α- and β-IO5 had reacted to the corresponding β-configured U nucleoside. The stereochemical information of the RNA strand obviously directs the outcome of the reaction, like we have previously shown for the IO3 nucleoside. This leads to a conversion of also the α-configured IO5 to β-configured U.

Finally, we wanted to examine if a proto-RNA strand containing both IO3 and IO5 nucleosides would react upon treatment with DTT to give C and U simultaneously. For this experiment, we prepared the oligonucleotide 5′-CC-$_{\beta}$IO3-CU-$_{\beta}$IO5-UU-3′ using the corresponding β-configured phosphoramidites (Figure 3C). Digestion of the strand generated signals for β-C, β-U and as expected for β-IO3 and β-IO5. Again, treatment of the strand with DTT and Fe^{2+} cleanly generated a new strand with a different retention time. Upon digestion only two signals were obtained: One for β-C and one for β-U,



Scheme 1. A) Synthesis of the IO5 phosphoramidite and conditions B) HPL-chromatogram and MALDI-TOF mass spectra of an oligonucleotide synthesized with the IO5 phosphoramidite β-**13** shown as an example.

Research Article
doi.org/10.1002/ceur.202300057

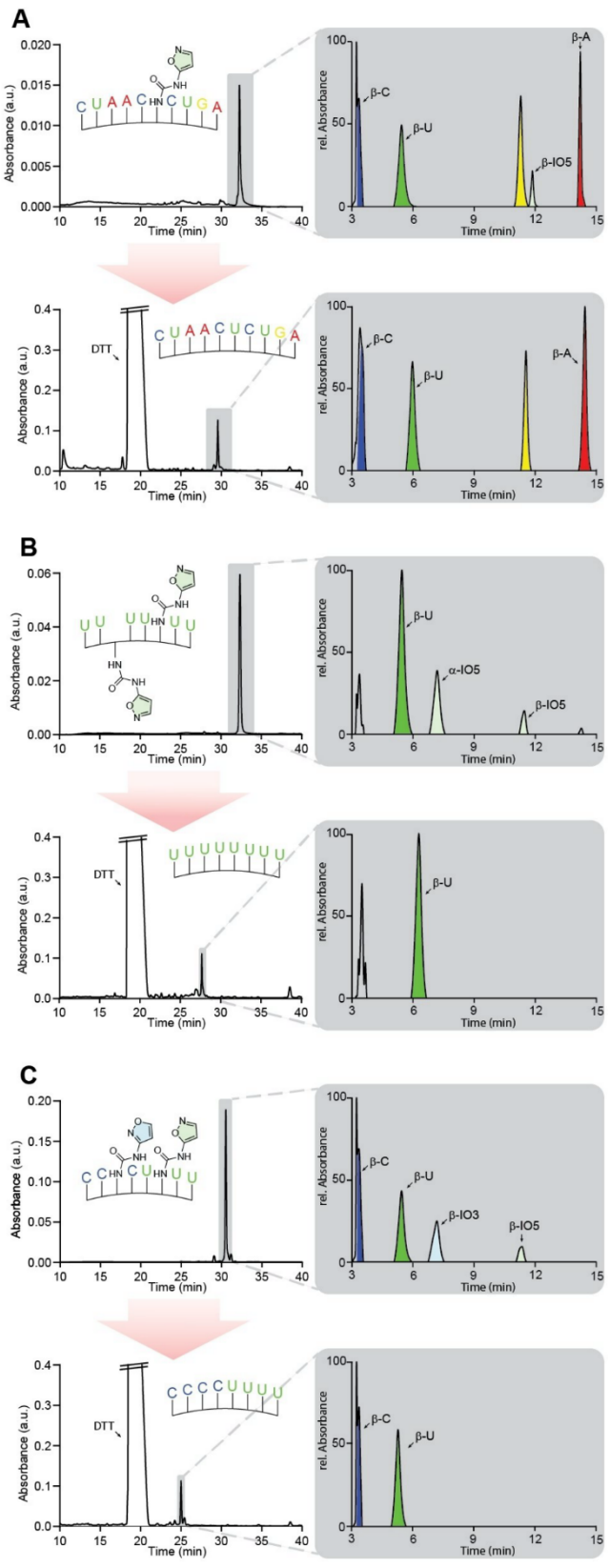


Figure 3. HPLC, MALDI, HPLC-MS (at 243 nm; grey brackets) analyses of the IO5 containing Strands before the reaction with DTT, Fe^{2+} (top) and after N–O cleavage and cyclization reaction (bottom). A) 5′-CUAAC-$_{\beta}$IO5-CUGA-3′ B) 5′-UU-$_{\alpha}$IO5-UU-$_{\beta}$IO5-GA-3′ C) 5′-CC-$_{\beta}$IO3-CU-$_{\beta}$IO5-UU-3′. A = adenosine, C = cytidine, G = guanosine, IO3 = 3-aminoisoxazole, IO5 = 5-aminoisoxazole, U = uridine.

showing that in the strand the precursor nucleosides β-IO3 and β-IO5 rearrange simultaneously to give C and U.

Conclusions

RNA is the central molecule within the current concepts of how life could have emerged on the Early Earth because of the molecule's potential to encode information and to catalyze reactions. It therefore effectively combines genotype with phenotype. The RNA molecule, however, is so complex that it most likely formed from a simpler precursor proto-RNA molecule that already had potentially information encoding properties. Current concepts of how RNA may have evolved are based on the formation of canonical nucleosides that got phosphorylated at some point and then oligomerized to yield RNA strands. This, however, requires the prebiotic formation of the canonical nucleosides.

We showed recently that nucleosides such as 3-amino-isoxazole (IO3) are potential precursors for the pyrimidine base C. Cytosine is formed upon simple reductive N–O bond cleavage followed by a cyclization reaction (Figure 1). Here, we show that the constitutional isomer, the 5-aminoisoxazole (IO5) can serve as a precursor for the pyrimidine nucleoside U. Both, the 3- and the 5-aminoisoxazoles form by reacting cyanoacetylene (**1**) with hydroxylamine (**2**). The formation of the different constitutional isomers depends exclusively on the pH-conditions of the reaction. It is prebiotically highly attractive to create the two needed pyrimidine nucleosides from the same chemical precursors (**1** and **2**) by just a change in reaction conditions. In addition, we show that IO5, if embedded in RNA, performs the formation of the nucleoside (U) directly in the RNA strand. This finding allowed us to insert both IO3 and IO5 into a proto-RNA for the simultaneous reaction of the two IO-precursors to C and U by just exposing the strand to thiols and catalytic amounts of Fe^{2+}.

Our data show that amino- and subsequently urea-isoxazoles are highly promising prebiotic precursors for the pyrimidine nucleosides. We show that these compounds can lead to the formation of a proto-RNA and through a simple N–O bond cleavage, followed by a cyclization cascade directly within the proto-RNA, canonical RNA can emerge.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for supporting this research through the DFG grants: CA275/11-3 (ID: 326039064), CRC1309 (ID: 325871075, A04), CRC1032 (ID: 201269156, A05) and CRC1361 (ID: 393547839, P02). We thank the Volkswagen Foundation for funding this research (grant EvoRib). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No. 741912 (EPiR) and from the BMBF Clusters4Future program CNATM.

Research Article
doi.org/10.1002/ceur.202300057

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

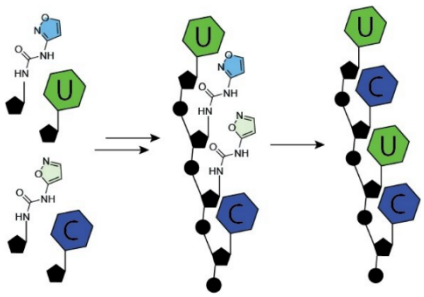
Keywords: nucleosides · origin of life · prebiotic chemistry · proto-RNA · RNA

[1] W. Gilbert, *Nature* **1986**, *319*, 618–618.
[2] F. H. Westheimer, *Nature* **1986**, *319*, 534–536.
[3] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.
[4] A. J. Zaug, T. R. Cech, *Science* **1986**, *231*, 470–475.
[5] D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc. Natl. Acad. Sci. USA* **1967**, *58*, 217–224.
[6] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229–1232.
[7] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.
[8] G. F. Joyce, *Angew. Chem. Int. Ed.* **2007**, *46*, 6420–6436.
[9] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**, *120*, 4766–4805.
[10] G. F. Joyce, *Nature* **2002**, *418*, 214–221.
[11] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Chem. Biol.* **2013**, *20*, 466–474.
[12] N. V. Hud, *Nat. Commun.* **2018**, *9*, 5171.
[13] A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821–12844.
[14] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *PNAS* **1987**, *84*, 4398–4402.
[15] A. E. Engelhart, N. V. Hud, *Cold Spring Harbor Perspect. Biol.* **2010**, *2*.
[16] H. Okamura, A. Crisp, S. Hübner, S. Becker, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2019**, *58*, 18691–18696.
[17] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76–82.
[18] F. Xu, A. Crisp, T. Schinkel, R. C. A. Dubini, S. Hübner, S. Becker, F. Schelter, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2022**, *61*, e202211945.
[19] E. R. Garrett, J. Tsau, *J. Pharm. Sci.* **1972**, *61*, 1052–1061.
[20] H. Lönnberg, R. Käppi, *Nucleic Acids Res.* **1985**, *13*, 2451–2456.
[21] P. Thaddeus, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 1681–1687.
[22] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *Science* **1966**, *154*, 784–785.
[23] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **2015**, *7*, 301–307.
[24] P. B. Rimmer, O. Shorttle, *Life* **2019**, *9*.
[25] K. B. Muchowska, S. J. Varma, J. Moran, *Nature* **2019**, *569*, 104–107.
[26] B. Damer, D. Deamer, *Life* **2015**, *5*, 872–887.
[27] A. Y. Mulkidjanian, A. Y. Bychkov, D. V. Dibrova, M. Y. Galperin, E. V. Koonin, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E821–E830.
[28] Y. Yamagata, *Origins Life Evol. Biospheres* **1999**, *29*, 511–520.
[29] J. P. Ferris, E. A. Williams, D. E. Nicodem, J. S. Hubbard, G. E. Voecks, *Nature* **1974**, *249*, 437–439.
[30] P. A. Gerakines, M. H. Moore, R. L. Hudson, *Icarus* **2004**, *170*, 202–213.
[31] H. J. Cleaves, *Precambrian Res.* **2008**, *164*, 111–118.
[32] A. W. Schwartz, R. M. de Graaf, *J. Mol. Evol.* **1993**, *36*, 101–106.
[33] C. Reid, L. E. Orgel, *Nature* **1967**, *216*, 455–455.
[34] M. Kijima, Y. Nambu, T. Endo, *J. Org. Chem.* **1985**, *50*, 1140–1142.

Version of record online: ■■, ■■

RESEARCH ARTICLE

Prebiotically plausible isoxazole nucleosides are shown to be excellent candidates for proto-nucleosides and proto-RNAs. Formation 3- and 5-aminoisoxazole can be traced back to the same starting materials. After incorporation into RNA the proto-nucleosides selectively rearrange to cytidine and uridine. The stereochemical outcome of this reaction at the anomeric center is controlled by the neighbouring bases.



*F. Xu, Dr. S. Wiedemann, Dr. J. Feldmann, Dr. S. Becker, Prof. Dr. T. Carell**

4. Unveröffentlichte Ergebnisse

4.1. Prolog

Es wurde lange Zeit angenommen, dass RNA aufgrund ihrer Fähigkeiten, sowohl Information zu speichern als auch die eigene Replikation zu katalysieren, eine zentrale Rolle in der Entstehung des Lebens spielte.[58] Die Suche nach *de-novo* Syntheserouten ausgehend von präbiotisch plausiblen Startmaterialien hat sich jedoch als schwierig erwiesen,[66] weswegen die Theorie aufgestellt wurde, dass das Molekül RNA selber auch ein Evolutionsprodukt ist und sich im Laufe der Abiogenese aus proto-RNAs entwickelt hat.[175] Die bereits berichteten präbiotischen Syntheserouten zu den RNA-Bausteinen (**Abschnitt 1.5.1**) konnten genutzt werden, um Strukturen zu identifizieren, die mit einer höheren Wahrscheinlichkeit auf der frühen Erde verfügbar waren. Es konnte bereits in den hier beschriebenen Publikationen gezeigt werden, dass die Pyrimidin-Vorläufermoleküle *N*-Isoxazol-Harnstoffe Teil einer Proto-RNA gewesen sein könnten. Die *N*-Isoxazol-Harnstoffe könnten anschließend über eine intramolekulare Reaktion in die kanonischen Pyrimidin-Basen umgewandelt worden sein. Im Leben, wie wir es heute kennen, wird der Mechanismus der Informationsübertragung und der Informationsspeicherung durch die vier verschiedenen Nukleobasen der RNA und DNA (A, G, C, T/U) sichergestellt. Diese vier Nukleobasen sind in der Lage 64 (4^3) unterschiedliche Codons zu formen, die im Translationsprozess zu den 20 kanonischen Aminosäuren übersetzt werden. Es ist schwer vorstellbar, dass dieser Prozess und die Vielzahl der in der Natur vorkommenden Nukleinsäure Strukturen mit nur zwei verschiedenen Nukleobasen oder einer Klasse an Nukleobasen (nur Purine oder nur Pyrimidine) erhalten werden konnte.[175]

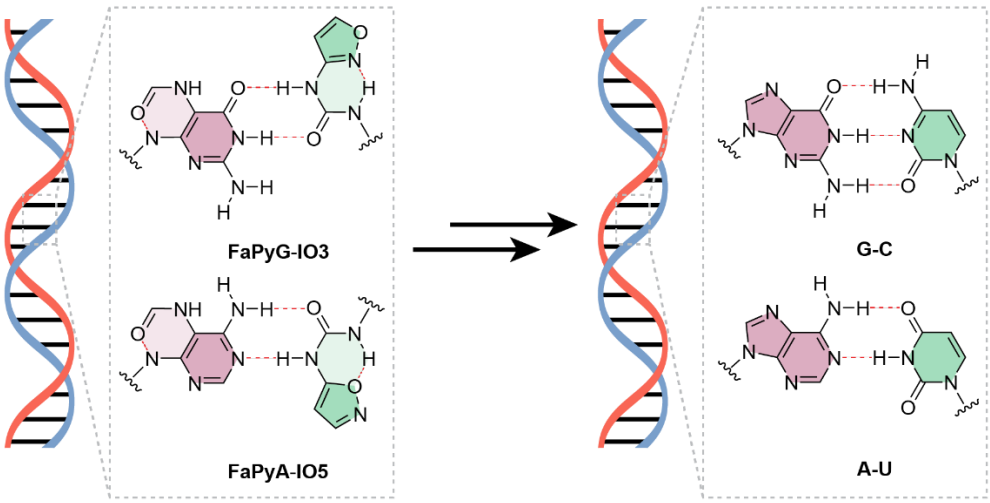
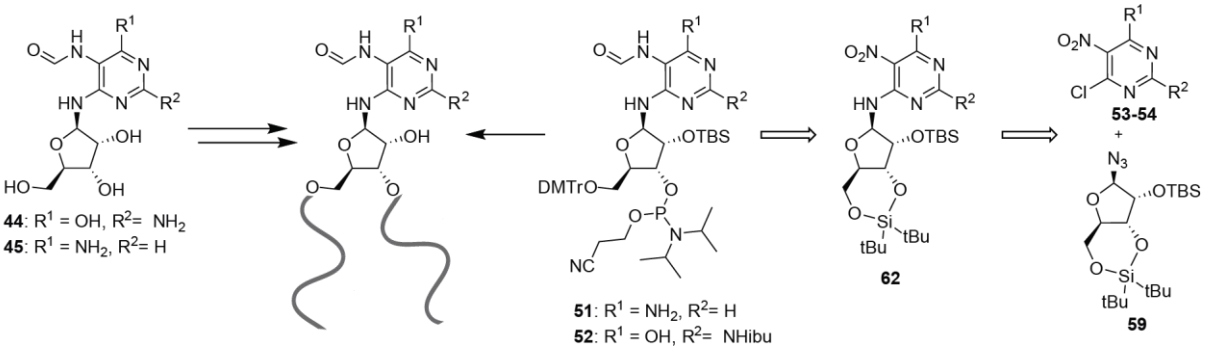


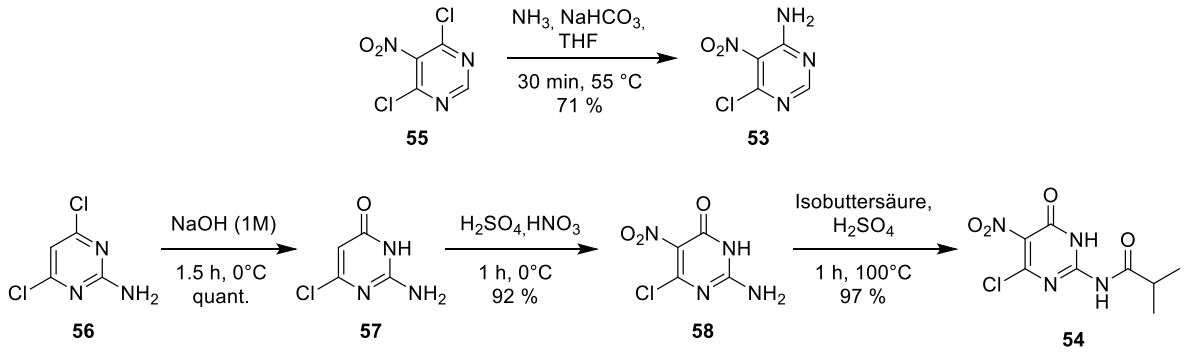
Abbildung 8. Postulierte *Proto-Watson-Crick*-Basenpaare, die mit einer Steigerung der Basenpaarungs-Stabilität, zu den kanonischen *Watson-Crick*-Basenpaare reagieren.

Diese Tatsache regte uns an ein Purin-Vorläufermolekül zu identifizieren, welches ebenfalls in der Lage ist Basenpaarungen einzugehen und somit eine plausible Komponente einer potentiellen Proto-RNA sein kann. Wir postulierten, dass Formamidopyrimidine (FaPys) in der Lage sind mit Pyrimidin-Nukleobasen zu paaren und ähnlich wie *N*-Isoxazol-Harnstoffe (IO) durch eine einfache intramolekulare Reaktion zu den kanonischen Purin-Nukleobasen umgewandelt werden können. Darüber hinaus besteht die Möglichkeit, dass FaPys auch mit *N*-Isoxazol-Harnstoffen Basenpaare bilden könnten. Dies würde uns den Weg zu einem primitiven Polymer eröffnen, welches über *Proto-Watson-Crick*-Basenpaare Information speichern könnte (**Abbildung 8**). Dieses System könnte im Laufe der chemischen Evolution durch bestimmte äußere Reaktionsbedingungen und Selektionsdrücke (z.B. stabilere Basenpaarung) ohne Komponentenaustausch in die heutige RNA umgewandelt worden sein. Um dieses Model zu überprüfen, mussten zunächst die proto-Nukleoside **44** und **45** als Phosphoramdit-Bausteine **51-52** synthetisiert und anschließend in RNA inkorporiert werden (**Schema 7**).



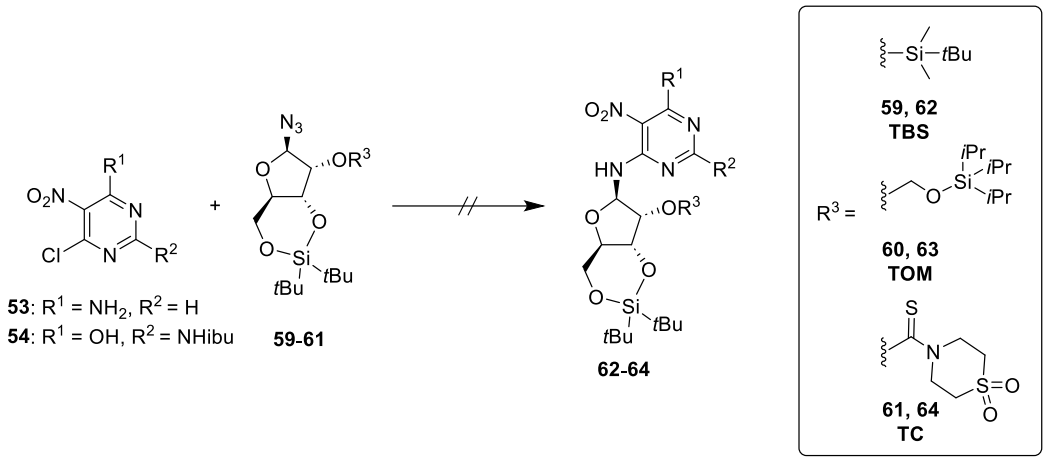
Schema 7. Synthesestrategie von Phosphoramidit-Bausteinen **51** und **52** und Inkorporation in RNA.

4.2. Ergebnisse und Diskussion



Schema 8. Synthese der Nitropyrimidine **53** und **54**.

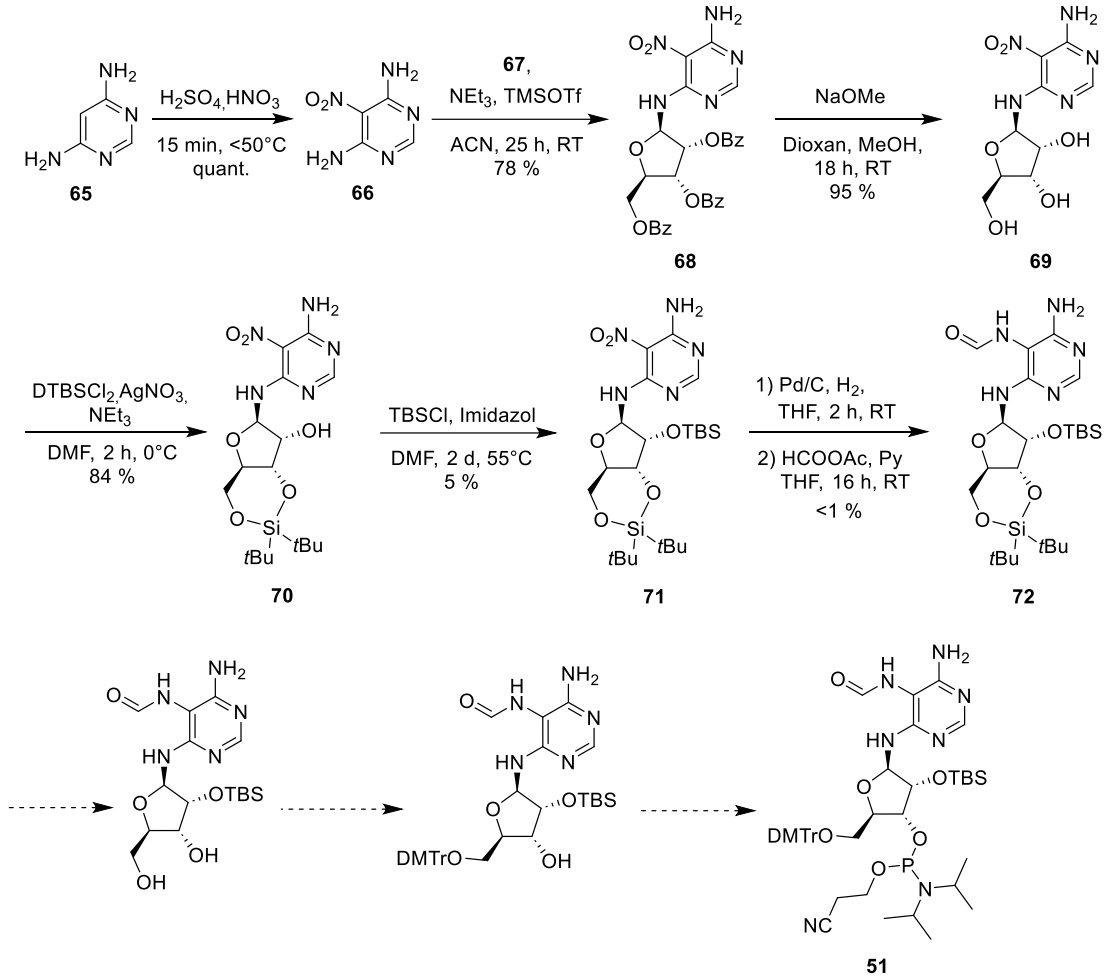
Für die erste Strategie zur Synthese der Phosphoramidite **51**-**52** sollten bereits etablierte Verfahren genutzt werden, die bereits zuvor in der Arbeitsgruppe von *Prof. Thomas Carell* für die Synthese der Deoxyribose-FaPy-Derivate verwendet wurden (**Schema 7**).[203, 204] Hierfür wurden zunächst die Nitro-Pyrimidine **53** und **54**, nach modifizierten Vorschriften aus der Literatur, synthetisiert (**Schema 8**).[205-208] Das kommerziell erhältliche 4,6-Dichlor-5-nitro-pyrimidin **55** wurde über eine Substitutionsreaktion mit Ammoniak (7 M in MeOH) zu 4-Amino-6-chloro-5-nitropyrimidin **53** umgesetzt. Die Einführung einer Schutzgruppe für die exozyklische Aminofunktion stellte sich als schwierig heraus, da die Nitrogruppe den bereits elektronarmen Pyrimidinring zusätzlich deaktiviert. Es wurde an diesem Schritt der Synthese bewusst auf eine Schützung des N^4-Amins verzichtet, mit der Hoffnung, dass das Amin nach der Reduktion der Nitrogruppe eine höhere Reaktivität aufweist oder dass aufgrund der geringen Reaktivität eine Schutzgruppe für die RNA-Synthese nicht benötigt wird. Der FaPyG-Baustein **54** wurde ausgehend von 2-Amino-4,5-dichloropyrimidin **56** über drei Schritte gewonnen. **56** wurde zunächst in Natriumhydroxid erhitzt, um **57** zu erhalten, welches wiederum zu **58** nitriert wurde. Im letzten Schritt wurde das N^2-Amin mit Isobuttersäure geschützt, um **54** zu erhalten.



Schema 9. Erfolglose Synthese der Nukleoside **62-64**.

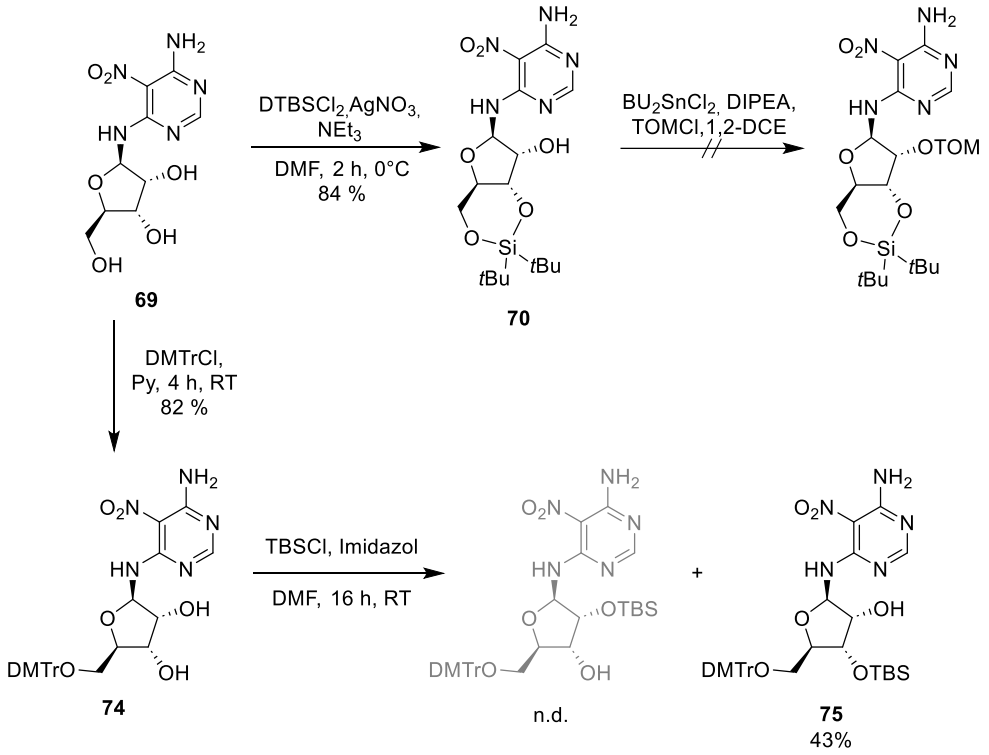
Im nächsten Schritt sollte die bereits in der Aminoisoxazol-Phosphoramidit-Synthese (**Abschnitt 3**) verwendete Azidoribose **59**[209, 210] mit den Nitro-Pyrimidinen gekoppelt werden. Hierzu wurde zunächst die Azid-Gruppe zum Amin reduziert und über eine nukleophile aromatische Substitution mit den Nitro-Pyrimidinen umgesetzt. Allerdings konnte unter den Standardbedingungen die Bildung des Produktes **62** nicht beobachtet werden (**Schema 9**). Es wurde vermutet, dass die sterische Hinderung der TBS-Gruppe an der 2'-Position die Reaktion beeinflusst. Infolgedessen wurden die sterisch weniger anspruchsvollen Schutzgruppen TOM ((Triisopropylsiloxy)methyl) **60** und TC (1,1-Dioxo-λ6-thiomorpholin-4-carbothioat) **61** verwendet. Auch in Reaktionen mit diesen Azidoribose-Derivaten konnte die Bildung der Produkte **63**-**64** weder über LCMS noch über NMR beobachtet werden.

Angesichts der Schwierigkeiten wurde eine alternative Syntheseroute entwickelt, in der erst das Nitro-Pyrimidin Nukleosid hergestellt wird, welches anschließend silyliert werden sollte (**Schema 10**). Hierzu wurde im ersten Schritt 6-Diaminopyrimidin **65** zu 4,6-Diamino-5-nitropyrimidin **66** nitriert und anschließend über eine *Vorbrüggen*-Glykosylierung mit 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-*β*-D-ribofuranose **67** selektiv zu **68** umgesetzt.[211, 212] Als nächstes wurden die Benzoyl-Schutzgruppen mit Natriummethanolat abgespalten, um das Nukleosid **69** (Klitozin) zu erhalten, welches anschließend an der 3'- und 5'-Position mit Di-*tert*-butyldichlorsilan silyliert wurde. Schließlich konnte die TBS-Schutzgruppe selektiv an der 2'-Position eingeführt werden, allerdings konnte **71** bei der Umsetzung mit TBSCl lediglich in einer Ausbeute von 5 % erhalten werden. In der folgenden *One-Pot*-Reaktion zur Reduktion der Nitrogruppe und anschließender Formylierung mit Ameisensäure-Essigsäureanhydrid konnte das FaPy-Produkt **72** nur in Spuren durch LCMS-Messungen nachgewiesen werden.



Schema 10. Synthese des geschützten Nukleosids **72** und geplante Syntheseroute zum Phosphoramidit **51**.

Aufgrund der geringen Ausbeuten in diesen beiden Schritten wurde nach Alternativen gesucht. Es wurde vermutet, dass die Reaktivität der 2'-Hydroxygruppe durch den sterischen Anspruch des Nitropyrimidins und durch Wasserstoffbrücken-Wechselwirkungen negativ beeinflusst wird. Um dies zu umgehen, wurden zwei unterschiedliche Ansätze untersucht (**Schema 11**). Erstens wurde die Verwendung der flexibleren TOM-Schutzgruppe zur Schützung der freien 2'-Hydroxygruppe von **70** ausprobiert und zweitens wurde ein direkterer Ansatz erprobt, in dem zuerst die 5'-Hydroxygruppe von **69** selektiv mit DMTrCl geschützt wird und im nächsten Schritt nicht-regioselektiv silyliert wird. Allerdings konnte durch den ersten Ansatz keine Produktformation beobachtet werden, während im zweiten Ansatz lediglich das 3'-geschützte Produkt **75** isoliert werden konnte. Aufgrund der Vielzahl an Schwierigkeiten, die mit der Synthese von **51** und **52** verbunden waren, wurde deshalb eine indirekte Methode untersucht, um die FaPy-Nukleobasen in RNA-Stränge inkorporieren zu können.



Schema 11. Alternative Syntheseversuche, um die geringe Reaktivität der 2'-Hydroxygruppe von **70** zu umgehen.

Es wurde bereits postuliert, dass die Stabilität von organischen Verbindungen, die zu Polymeren zusammengesetzt sind, generell die Stabilität der gleichen Verbindung in Monomerer-Form übersteigt. Somit könnten auf der frühen Erde insbesondere Makromoleküle, die über thermodynamisch ungünstige, Stabilitäts-steigernde Reaktionen entstehen, angereichert worden sein.[213] Im gleichen Zuge wurde vorgeschlagen, dass racemisches Ribose-5-phosphat[214, 215] (D und L) ebenfalls in der Lage war zu einem nicht-genetischen Polymer zu oligomerisieren.[213, 216] Die Zucker, die durch Phosphodiester-Bindungen verbunden waren, unterlagen spontanen Umesterungsreaktionen mit Selektion auf Stabilität.[213] Benachbarte präbiotische Ribosephosphate könnten dann Phosphodiesterbindungen gebildet haben und zu einem Zucker-Phosphat-Grundgerüst polymerisiert worden sein.[217] Intermolekulare Strukturen könnten durch den gelegentlichen Einbau von komplementären Nukleotiden oder durch die Beladung mit Nukleobasen weiter stabilisiert worden sein. Durch diesen Prozess konnten präbiotische Polymere und schließlich komplementäre Oligonukleotide angereichert worden sein.[213] Das von *Yakhin* entwickelte Modell war neben der ungeklärten Herkunft der Ribose, vor allem wegen des *Glykosylierungs-Problems* (**Abschnitt 1.5.1**) umstritten.[217, 218] Wir postulieren hier, dass Formamidopyrimidine und *N*-Isoxazol-Harnstoffe eine Lösung für das *Glykosylierungs-Problem* sein könnten.

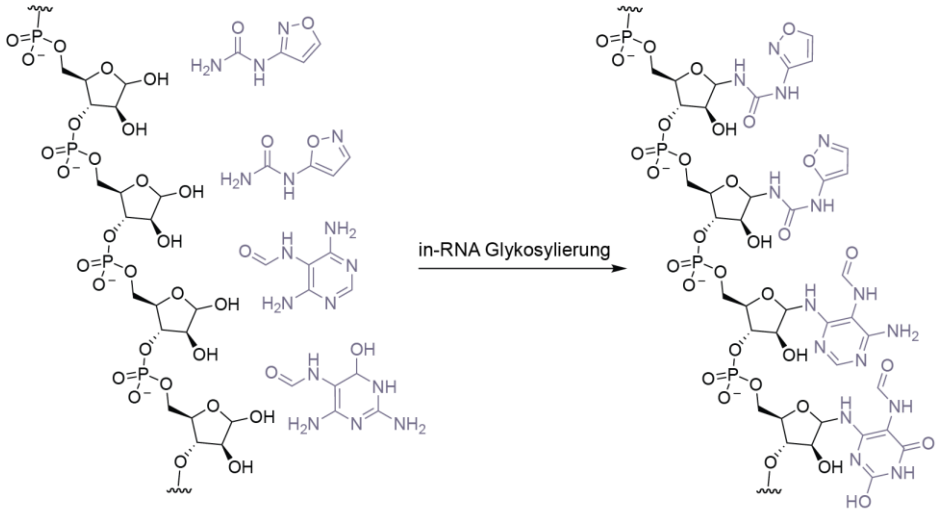
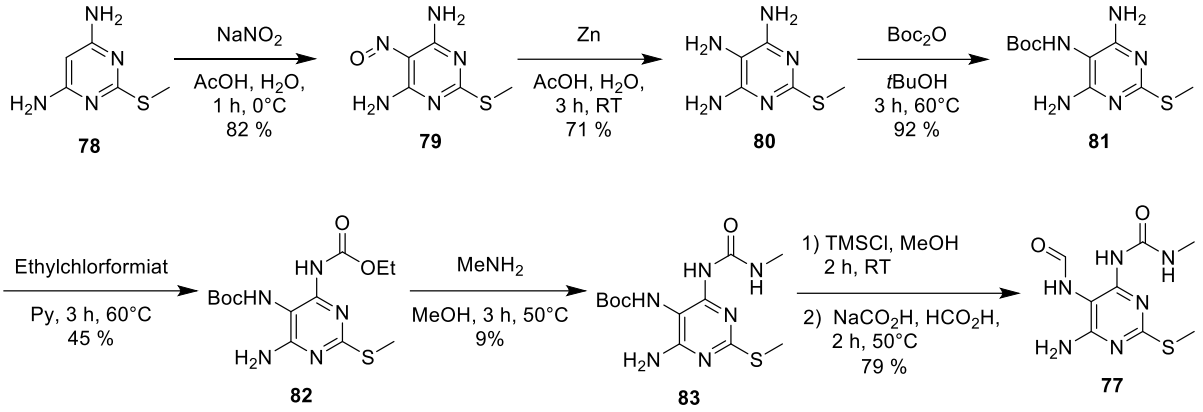


Abbildung 9. Vorgeschlagenes Konzept zur Beladung von abasischen Stellen mit Proto-Nukleobasen.

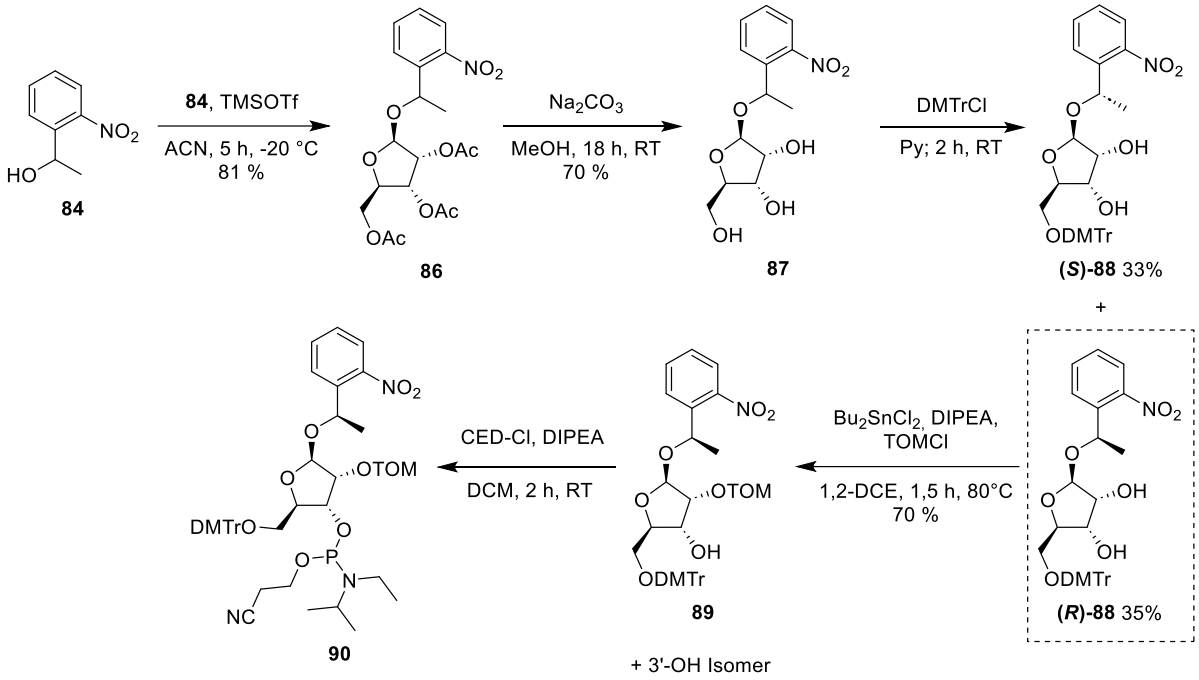
Um dies zu Untersuchen und gleichzeitig die oben beschriebenen synthetischen Herausforderungen zu umgehen, sollten RNA-Stränge mit abasischen Stellen hergestellt werden, die anschließend mit den Proto-Nukleobasen beladen werden sollten (in-RNA Glykosylierung, **Abbildung 9**). Hierzu wurden zunächst die Vorläufer-Nukleobasen *N*-Isoxazol-3-yl-Harnstoff **27** (IO3), *N*-Isoxazol-5-yl-Harnstoff **76** (IO5), 4,6-diamino-5-formamidopyrimidin **43** (FaPyA) und 2,6-diamino-4-oxo-5-formamidopyrimidin **42** (FaPyG) nach bekannten Synthesevorschriften hergestellt.[98, 99, 210] Zusätzlich wurde die *N*-Harnstoff modifizierte Vorläufer-Nukleobase **77** nach leicht modifizierter Vorschrift von *Antony Crisp* hergestellt. Diese Proto-Nukleobase ist ein plausibler Vorläufer von N^6-Methylcarbamoyl-Adenosin[219] und der Modifikationen g^6A und t^6A.[220] **77** wurde ausgehend von dem kommerziell erhältlichen 4,6-Diamino-2-(methylthio)pyrimidin **78** hergestellt, welches im ersten Schritt mit Natriumnitrit zu **79** umgesetzt wurde (**Schema 12**). Darauf folgte eine Reduktion der Nitroso-Gruppe zu **80** und die Boc-Schützung der neu entstandenen Aminofunktion zu **81** mit einer Ausbeute von 92 %. Als nächstes wurde **81** mit Ethylchloroformiat zum Carbamat **82** umgesetzt, das anschließend mit ethanolischem Methylamin zum *N*-Methylharnstoff **83** reagiert. Schließlich wurde die Vorläufer-Nukleobase **76** in zwei Schritten durch die Abspaltung der Boc-Gruppe mit TMS-Cl und anschließender Formylierung mit Natriumformiat in Ameisensäure mit einer Ausbeute von 79 % erhalten.



Schema 12. Synthese der *N*-Harnstoff modifizierten Vorläufer-Nukleobase **77** in Anlehnung an *Antony Crisp*.[220]

Die abasischen Stellen sollten mit Hilfe eines von *Leumann et al.* beschriebenen Phosphoramidits in die Oligonukleotide eingeführt werden.[221] Die Strategie involvierte die photolabile Schutzgruppe 1-(2-Nitropenyl)ethyl (NPE), die nach der RNA-Festphasensynthese abgespalten werden sollte, um das abasische Oligonukleotid zu erhalten. Diese wurde ausgehend von racemischem 1-(2-Nitrophenyl)ethanol **84** über eine *Hilbert-Johnson*-Reaktion an der 1-Position der *β*-D-Ribofuranose-1,2,3,5-tetraacetat **85** installiert. Im nächsten Schritt wurden die Acetyl-Gruppen abgespalten und anschließend selektiv an der 5‘-Hydroxygruppe mit DMTrCl geschützt. In diesem Schritt konnten die Diastereoisomere **(*S*)-88** und **(*R*)-88** chromatographisch aufgetrennt werden. Für die nachfolgenden Schritte wurde nur das **(*R*)-88**[1] Isomer verwendet, welches als nächstes mit TOM-Cl zu **89** umgesetzt wurde. Das erwünschte Regioisomer konnte durch wiederholte Säulenchromatographie aufgereinigt werden. Im letzten Schritt wurde das Phosphoramidit **90** durch eine Reaktion mit CED-Cl mit einer Ausbeute von 85 % erhalten. **90** wurde anschließend mittels RNA-Festphasensynthese erfolgreich in die Oligonukleotide **S1^{NPE}-S2^{NPE}** eingebaut.

[1] Das chirale Zentrum hat keinen Einfluss auf die Abspaltung der NPE-Schutzgruppe nach der RNA-Synthese. In beiden Fällen verläuft diese quantitativ. Zur Vereinfachung der Analysen wurde nur das **(*R*)-87** Isomer für die folgenden Schritte verwendet.



Schema 13. Synthese des abasischen Phosphoramidits **89** nach Vorschrift von *Leumann et al.*[221]

Als nächstes musste die NPE-Schutzgruppe abgespalten werden. Hierzu wurden die Oligonukleotide in ddH_2O gelöst (100 µM) und bei Raumtemperatur mit einer *Leica KL 1500 LCD-Kaltlichtquelle* ausgestattet mit einer 150 W Halogenlampe (>300 nm) bestrahlt. Zur Bestimmung der Entschützungskinetik wurden Proben nach 1 min, 5 min, 15 min, 30 min und 60 min entnommen und über RP-HPLC analysiert (**Abbildung 10**). Anhand dieser Ergebnisse wurde für die Entschützung eine Bestrahlungszeit von 2 h gewählt, um die Stränge **S1-S2** zu erhalten. Überraschenderweise konnte während dieser Reaktion keine Fragmentierung der abasischen Stelle, durch β-Eliminierung oder der Bildung des Cyclophosphat-Intermediats beobachtet werden. Generell waren die Stränge **S1^{NPE}-S2^{NPE}** und **S1-S2** erstaunlich stabil und konnten mehr als sechs Monate bei -20 °C, pH 6-8 ohne Zersetzung gelagert werden. Trotz dieser Erkenntnisse wurden **S1^{NPE}-S2^{NPE}** für die nachfolgenden Experimente jeweils frisch entschützt und aufgereinigt, um **S1-S2** zu erhalten.

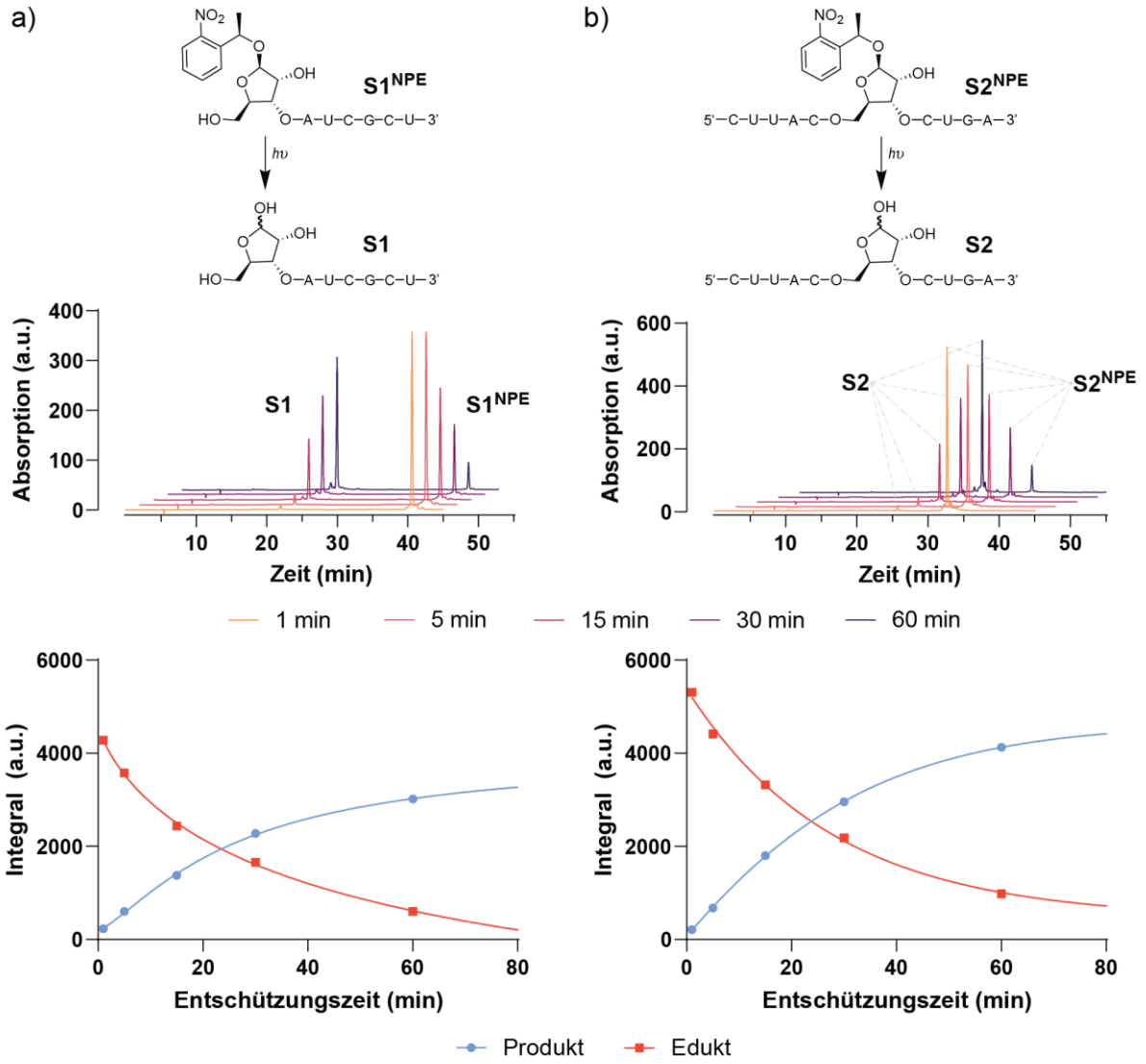


Abbildung 10. Entschützung der photolabilen Schutzgruppe NPE: HPLC-Chromatogramme (0-30% Puffer B in 45 min) der Reaktion und exponentieller Fit von Edukt und Produkt-Integrale der HPLC-Chromatogrammen nach 1 min, 5 min, 15 min, 30 min und 60 min für a.) **S1** und b) **S2**.

Nach der erfolgreichen Synthese der Oligonukleotide **S1-S2** konnten vorläufige Experimente zur Beladung der abasischen Stelle mit den Proto-Nukleobasen (in-RNA Glykosylierung) durchgeführt werden. Hierzu wurden die Oligonukleotide in einer Lösung aus 50 mM Natriumborat Puffer (pH 8), 50 mM NaCl und 250 mM der Proto-Nukleobasen gelöst. Die Reaktionsansätze wurden über eine *dry-down*-Methode bei 30 °C in einem *Thermoshaker* inkubiert. Auch unter diesen Reaktionsbedingungen waren die Oligonukleotide **S1**-**S2** überraschend stabil und zeigten auch nach 30 Tagen keine komplette Fragmentierung des Eduktes. Proben für MALDI- und RP-HPLC-Messungen wurden in regelmäßigen Zeitabständen entnommen. Dabei konnte die Bildung des Produktes **S1**$^{\mathbf{FaPyA}}$ bereits nach 3 Tagen mit einer Ausbeute von 5 % (geschätzt über Integration der HPLC-

Signale) beobachtet werden. Für die Kopplungen der Proto-Nukleobasen *N*-Isoxazol-3-yl-Harnstoff **27** (IO3), *N*-Isoxazol-5-yl-Harnstoff **75** (IO5), 2,6-diamino-4-oxo-5-formamidopyrimidin **42** (FaPyG) und **76** konnten bisher noch keine geeigneten Bedingungen gefunden werden. Hierzu finden aktuell laufende Studien im Arbeitskreis *Carell* statt.

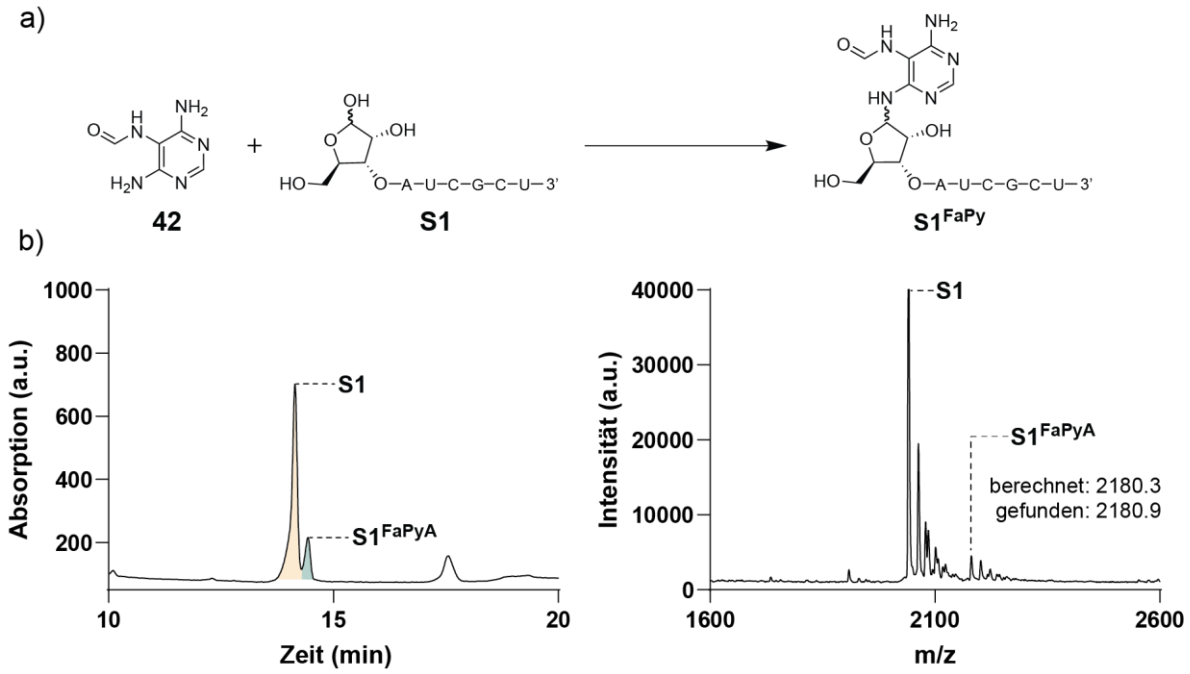


Abbildung 11. HPLC-Chromatogramm (0-40% Puffer B in 45 min) und MALDI-TOF Spektrum der Beladungsreaktion von **S1** mit **43** nach 3 Tagen.

Zusammenfassend konnte gezeigt werden, dass die abasische RNA unter präbiotisch plausiblen Bedingungen mit dem Proto-Nukleosid FaPyA beladen werden kann. Dieses Ergebnis könnte einen entscheidenden Beweis für das präbiotische Phosphodiester-Modell von *Yakhin* liefern.[213] Neben FaPyA **43** könnten die Proto-Nukleobasen **25**, **42**, **75** und **76** ebenfalls Teil eines solchen Systems gewesen sein, um die intermolekularen Bindungen der Makromoleküle durch Wasserstoffbrücken-Bindungen zu stabilisieren. Um dies und die Plausibilität der FaPy-Nukleobasen als Komponente einer potentiellen Proto-RNA zu überprüfen können in künftigen Studien die Basenpaarungs- und physikochemischen-Eigenschaften von FaPy-beladener RNA untersucht werden.

5. Experimenteller Teil

5.1. Methoden

Chemische Synthese

Alle Reaktionen wurden, sofern nicht anders angegeben, mit Anwendung von Standard-Schlenktechnik unter Stickstoff- oder Argonatmosphäre durchgeführt. Alle Glasgeräte wurden vor Gebrauch unter Hochvakuum mit einer Heißluftpistole bei 550 °C ausgeheizt und ausgekühlt. Für Reaktionen mit Flusssäure wurden Polypropylen Reaktionsgefäße verwendet.

Chemikalien und Lösungsmittel

Die verwendeten Chemikalien wurden von den Firmen *Sigma Aldrich*, *Carbosynth*, *ABCR*, *Acros Organics*, *Merck* und *TCI Europe* bezogen. Diese wurden, wenn nicht anders erwähnt, ohne weitere Aufreinigung verwendet. Für Synthesen wurden über Molsieb getrocknete Lösungsmittel der Firma *Sigma-Aldrich* verwendet, welche nur unter Schutzgasatmosphäre entnommen wurden. Zur Extraktion und Säulenchromatographie wurden, wenn nicht anders erwähnt, Lösungsmittel technischer Güte verwendet, die vorher *in vacuo* destilliert wurden.

Dünnschichtchromatographie und Säulenchromatographie

Für Reaktionskontrollen wurden Aluminium-Fertigplatten Kieselgel F_{254} von *Merck* verwendet. Die Visualisierung erfolgte durch Bestrahlung mit UV-Licht (λ_{254nm}, $_{366nm}$) und/oder durch Anfärben mit Cer-Ammoniummolybdat-Lösung, Kaliumpermanganat-Lösung oder *p*-Anisaldehyd-Lösung. Die präparative Aufreinigung der Rohprodukte wurde durch Normalphasen-Säulenchromatographie unter Verwendung der *Flash* Methode mit Stickstoff-Druck erreicht. Als stationäre Phase diente Kieselgel von *Machery-Nagel* oder *Merck Millipore* (Korngröße 40-63 µm oder 14-40 µm). Als mobile Phase wurden Lösungsmittel von technischer Güte verwendet, die vor der Verwendung *in vacuo* destilliert wurden. Fraktionen wurden je nach Ansatzgröße in Reagenzgläsern oder 50-300 mL Gefäßen gesammelt und über DC oder LC-MS analysiert

Rotationsverdampfer

Zur *in vacuo* Destillation der Lösemittel wurde ein *Laborota 4000* der Firma *Heidolph* oder ein *Rotavapor R-300* der Firma *Büchi* verwendet. Sofern nicht abweichend angegeben, wurden Lösemittel bei 45 °C entfernt.

Bidestilliertes Wasser

Für HPLC-Aufreinigungen und für RNA-Experimente wurde Wasser verwendet, das zuvor über eine MilliQ-Anlage von *Satorius* aufgereinigt wurde.

Lyophilisation

Zum Entfernen der Lösemittel mittels Gefriertrocknung wurde eine Lyophille *2-4 LD plus* der Firma *Christ* verwendet. Die Lösemittel wurden dabei zuvor mit flüssigem Stickstoff eingefroren. Benzen-haltige Proben wurden an der Schlenk-Line lyophilisiert.

Schmelzpunktanalyse

Für Schmelzpunktbestimmung kristalliner Feststoffe wurde ein *Büchi Melting Point B-540* Schmelzpunktmessgerät verwendet.

pH-Meter

Ein pH-Meter *MP 220* der Firma *Metter Toledo* wurde für pH-Messungen verwendet. Dieses wurde vor jeder Messung mit einer pH = 4.0 und pH = 7.0 Standardlösung kalibriert.

Infrarotspektroskopie (IR)

Infrarotspektren wurden an einem *FT-IR Spektrum BXII* Spektrometer mit einem *Smith Dura SampIIR II Diamant-ATR* Sensor oder einem *Shimadzu IRSpirit* Spektrometer mit einem *QATR*-S Sensor aufgenommen. Feststoffe wurden dabei entweder direkt auf die ATR-Einheit gegeben und mit einem Stempel auf der ATR-Einheit verdichtet oder zuvor in wenig Aceton gelöst. Ein Tropfen der Lösung wurde auf die ATR-Einheit aufgetragen und das Lösemittel verdampft. Öle und flüssige Substanzen wurden als Filme aufgetragen. Absorptionsbanden wurden in Wellenzahlen (cm^{-1}) angegeben. Der Messbereich lag zwischen 4000 cm^{-1} und 650 cm^{-1}. Die Intensität der Signale wurden mit s (stark), m (medium), w (schwach) angegeben.

Kernresonanzspektroskopie (NMR)

NMR-Spektren zur Produkt- und Reaktionskontrolle wurden an einem *Bruker Ascend 400* oder einem *Bruker Ascend 500* aufgenommen. Eindimensionale ^{1}H-, ^{13}C-, ^{31}P-NMR-Spektren und zweidimensionale ^{1}H- und ^{13}C-NMR-Spektren wurden von der Analytik Abteilung für Kernresonanzspektroskopie des Departments Chemie der Ludwig-Maximilians-Universität München mit Hilfe eines *Varian Inova 400*, *Bruker ARX 600* oder *Bruker Avance III HD 800* angefertigt. Als interner Standard wurden deuterierte Lösungsmittel der Firmen *Sigma-Aldrich* und *Eurisotop* verwendet. Die chemischen Verschiebungen δ für ^{1}H und ^{13}C-Spektren wurden in parts per million (ppm) angegeben und Kopplungskonstanten *J* sind jeweils in

Hertz (Hz) angegeben. Die Multiplizitäten der Signale wurden als s (Singulett), d (Dublett), t (Triplett), q (Quartett) oder m (Multiplett) und deren Kombinationen angegeben.

Massenspektrometrie (LC-MS, HR-ESI, HR-EI, Orbitrap, QExactive, MALDI)

LC-MS Messungen zur Reaktionskontrolle wurden an einem *MSQ Plus* Spektrometer von *Thermo Scientific* mit einem *Dionex Ultimate 3000 LC-System* durchgeführt. Hochauflösende Massenspektren (HRMS) wurden von der Analytik-Abteilung der Ludwig-Maximilians-Universität München, Department Chemie (Haus F) gemessen. ESI-Massenspektren wurden an einem *Finnigan LTQ FTICR* aufgenommen und EI-Massenspektren wurden an einem *Joel JMS-700 spectrometer* aufgenommen. Für die Analyse von Oligonukleotiden wurde ein *Bruker Autoflex* II-Gerät mit einem *time-of-flight* Detektor verwendet. Die Proben wurden auf einem 0.025 µm *VSWP Filter* (Millipore) mit ddH_2O entsalzt und anschließend mit einer HPA-Matrix (3-Hydroxypicolinsäure in ddH_2O mit 10 mg/mL Diammoniumcitrat) co-kristallisiert. Präbiotische Reaktionskontrollen wurden über ESI-LC-MS zunächst durch ein *Dionex Ultimate 3000 HPLC*-System der Firma *Thermo Scientific* chromatographisch aufgetrennt und anschließend an einer LTQ Orbitrap XL von *Thermo Finnigan* analysiert. Chromatographische Auftrennung wurde an einer C18-Säule (*Interchim Uptisphere120 3HDO*) von *Interchim* mit einer Flussrate von 0.15 mL/min und einer Säulentemperatur von 30 °C durchgeführt. Zur Elution wurde ein 2 mM $HCOONH_4$ Puffersystem (Puffer A: 2 mM $HCOONH_4$ in H_2O (pH 5.5) und Puffer B: 2 mM $HCOONH_4$ in H_2O/ACN 20/80 (pH 5.5)) mit einem Gradienten von 0-4 % in 35 min verwendet. Die Elution wurde mit einem *Dionex Ultimate 3000 Diode Array* Detektor der Firma *Thermo Scientific* überwacht. RNA-Verdau Experimente wurden mit Hilfe eines *Thermo Scientific QExactive HF* Massenpektrometers gekoppelt an einem *Thermo Scientific Vanquish* System analysiert. Chromatographische Auftrennung wurde an einer eine C18-Säule (*Interchim Uptisphere120 3HDO*) von *Interchim* mit einer Flussrate von 0.2 mL/min und einer Säulentemperatur von 30 °C durchgeführt. Zur Elution wurde ein 2 mM $HCOONH_4$ Puffersystem (Puffer A: 2 mM $HCOONH_4$ in H_2O (pH 5.5) und Puffer B: 2 mM $HCOONH_4$ in H_2O/ACN 20/80 (pH 5.5)) mit einem Gradienten von 0-15 % in 18 min verwendet. Die Elution wurde bei 223, 243 and 260 nm über einen *Thermo Scientific Vanquish Diode Array* Detektor überwacht.

High Performance Liquid Chromatography (HPLC)

Analytische HPLC-Untersuchungen wurden an einem *1260 Infinity II 800 bar* System der Firma *Agilent Infinitylab* mit einem *G7165A* Detektor und einer Flussrate von 1.0 mL/min durchgeführt. Für die Auftrennung wurde eine C18-Säule (*EC 250/4 Nucleodur 100-3 C18ec*) von *Machery-Nagel* verwendet. Zur semi-präparativen Aufreinigung wurde ein *1260 Infinity II Manual Preparative 400 bar* System von der Firma *Agilent Infinitylab* mit einem *G7114A*

Detektor und einer Flussrate von 5.0 mL/min verwendet. Als stationäre Phase wurde eine C18-Säule (*VP 250/10 Nucleodur 100-5 C18ec*) von *Machery-Nagel* verwendet. Für die Aufreinigung und Analyse von Oligonukleotiden wurde ein 100 mM NEt_3/AcOH Puffersystem verwendet (Puffer A: 100 mM NEt_3/AcOH in H_2O (pH 7) und Puffer B: 100 mM $HCOONH_4$ in H_2O/ACN 20/80 (pH 7)). Für die Aufreinigung und Analyse von niedermolekularen Verbindungen wurde ein 2 mM $HCOONH_4$ Puffersystem (Puffer A: 2 mM $HCOONH_4$ in H_2O (pH 5.5) und Puffer B: 2 mM $HCOONH_4$ in H_2O/ACN 20/80 (pH 5.5)) verwendet.

Oligonukleotid Synthese

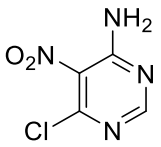
Oligonukleotide wurden im 1 µmol - 2 µmol Maßstab im DMT-OFF-Modus an einem *394 DNA/RNA Synthesizer* von *Applied Biosystems* durchgeführt. Die verwendeten Phosphoramidite der kanonischen Nukleoside Bz-A-CE, Ac-C-CE, Dmf-G-CE und U-CE wurden von *LinkTech* oder *Sigma-Aldrich* bezogen. Als Festphasenmaterial wurde *RNA SynBase™ CPG 1000/110* und *High Load Glen UnySupport™* verwendet. Die Phosphoramidite wurden in Acetonitril (≥ 99.8 %, 100 mM) gelöst. Dichloressigsäure (DCA) wurde als Entschützungsreagenz in Dichlormethan, BTT, ETT oder *Activator 42®* als Aktivator in ACN, Ac_2O als Capping-Reagenz in Pyridin/THF und I2 als Oxidationsmittel in Pyridin/H_2O verwendet. Für die Entschützung der RNA-Oligonukleotide wurde die Kartusche nach der Synthese zunächst unter Hochvakuum getrocknet, anschließend wurde das CPG-Trägermaterial in ein Mikroreaktionsgefäß (2 mL Eppendorf) überführt. die Festphase wurde in einer wässrigen 1:1 Mischung (0.5 mL) aus 30 % NH_4OH und 40 % $MeNH_2$ suspendiert. Die Suspension wurde bei 65 ºC erhitzt (25 min für SynBase™ CPG 1000/110 und 60 min für High Load Glen UnySupport™). Anschließend wurde der Überstand gesammelt, und die Festphase wurde mit Wasser (2×0.3 ml) gewaschen. Die Überstände wurden vereinigt, unter vermindertem Druck an einer *SpeedVac Plus* von *Thermo Life Sciences* eingeengt und durch Lyophilisation getrocknet. Der Rückstand wurde in DMSO (100 µL) resuspendiert und mit HF•Triethylamin (98 %, 125 µL) versetzt. Die Lösung wurde bei 65 °C, für 2.5 h, bei 750 rpm inkubiert. Anschließend wurde das Gemisch auf 0 °C gekühlt, NaOAc (3 M in ddH_2O, 25 µL) wurde hinzugegeben und vermischt. *n*-Butanol (1 mL) wurde hinzugeben und die RNA wurde 2 h lang bei -80 °C gefällt. Anschließend wurde das Gemisch bei 4 °C für 1 h zentrifugiert, der Überstand entfernt und der farblose Rückstand mittels Lyophilisation getrocknet. Die Absorption der synthetisierten Oligonukleotide in wässriger Lösung wurde mit einem *IMPLEN NanoPhotometer®* N60/N50 bei 260 nm gemessen. Der Extinktionskoeffizient des ONs wurde mit dem *OligoAnalyzer Version 3.0* von *Integrated DNA Technologies* berechnet. Für Oligonukleotide, die nicht-kanonische Basen enthalten, wurde angenommen, dass die Extinktionskoeffizienten identisch zu den kanonischen Basen sind.

UV-Vis-Spektroskopie

Für UV-Vis-spektroskopische Messungen wurde ein JASCO V650 Photospektrometer verwendet. Als Blindprobe zur Basislinienbestimmung vor den Messungen wurde das jeweilige Lösungsmittel der Probe verwendet.

5.2. Synthesevorschriften und Analytik

4-Amino-6-chloro-5-nitropyrimidin (53)

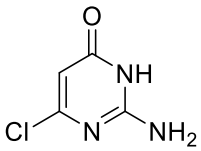


53

Zu einer Lösung von 4,6-Dichlor-5-nitro-pyrimidin **55** (6.00 g, 30.9 g, 1.00 eq.) in trockenem THF (45 mL) wurde $NaHCO_3$ (2.60 g, 30.9 g, 1.00 eq.) hinzugegeben. Das Reaktionsgemisch wurde bei 55 °C für 30 min erhitzt, bevor NH_3 (7 M in Methanol, 4.90 mL, 34.0 mmol, 1.10 eq.) zugegeben wurde. Das resultierende Gemisch wurde bei 55 °C für weitere 18 h erhitzt. Das entstandene Präzipitat wurde abfiltriert und mit THF gewaschen, um **53** (3.82 g, 21.9 mmol, 71 %) als gelblichen Feststoff zu erhalten.

R_f = 0.33 (*i*Hex:EtOAc = 2:1); **^{1}H NMR** (400 MHz, DMSO-d_6): δ(ppm) = 8.37 (s, 1H); **^{13}C NMR** (101 MHz, DMSO-d_6): δ(ppm) = 159.9, 155.0, 151.6, 114.9; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3306 (w), 3142 (w), 1959 (w), 1651 (w), 1579 (m), 1519 (s), 1482 (s), 1396 (m), 1341 (s), 1233 (s), 1042 (s), 982 (w), 963 (s), 862 (s), 786 (s), 777 (s), 750 (m), 714 (w), 667 (w); **HRMS** (ESI): berechnet für $C_4H_4ClN_4O_2^+$ $[M+H]^+$: 175.0017, gefunden: 175.0019.

2-Amino-6-chlor-3H-pyrimidin-4-on (57)

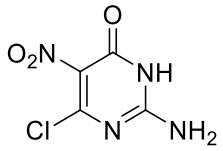


57

2-Amino-4,6-dichloropyrimidin **56** (10.0 g, 61.0 mmol, 1.0 eq.) wurde in einer NaOH-Lösung (1 M, 140 mL) für 1.5 h unter Rückfluss erhitzt. Anschließend wurde das Reaktionsgemisch auf 0 °C gekühlt, bevor das Produkt durch Zugabe von Essigsäure (15 mL) präzipitiert wurde. Der Feststoff wurde abfiltriert und mit Wasser gewaschen, um **57** (9.15 g, 62.9 mmol, quant.) als gelbliche Kristalle zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 11.58 (s, 1H), 7.16 (s, 2H), 5.55 (s, 2H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) = 163.2, 159.3, 156.1, 99.2; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3455 (w), 3269 (w), 3110 (m), 2917 (m), 2768 (m), 1644 (s), 1608 (s), 1559 (s), 1484 (s), 1398 (m), 1370 (s), 1230 (m), 1158 (m), 1117 (w), 1026 (w), 982 (s), 907 (m), 864 (m), 795 (s), 756 (m), 719 (m), 635 (s), 585 (s), 562 (s), 420 (s); **HRMS** (ESI): berechnet für $C_4H_3ClN_3O^-$ [M-H]$^-$: 143.99646, gefunden: 143.99703.

2-Amino-6-chlor-5-nitro-3H-pyrimidin-4-on (58)

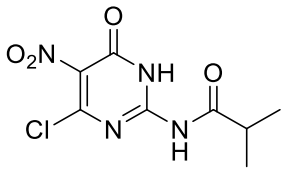


58

2-Amino-6-chlor-3H-pyrimidin-4-on **57** (10.0 g, 68.7 mmol, 1.0 eq.) wurde portionsweise zu einer Lösung aus Schwefelsäure (konz., 46 mL) und Salpetersäure (konz., 19 mL) bei 0 °C gegeben und für 1 h gerührt. Anschließend wurde die Reaktionslösung in ein Eis-Wasser Gemisch (150 mL) geschüttet. Das Gemisch wurde über Nacht bei 5 °C gelagert, anschließend wurde der Feststoff abfiltriert, mit eiskaltem Wasser gewaschen und getrocknet, um **58** (12.0 g, 62.9 mmol, 92 %) als gelblichen Feststoff zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 12.16 (s, 1H), 8.56 (s, 1H), 7.12 (s, 1H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) = 155.1, 154.2, 153.2, 126.7; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3305 (m), 3176 (s), 1638 (s), 1558 (m), 1446 (m), 1365 (s), 1336 (m), 1267 (s), 1125 (s), 1045 (s), 854 (m), 798 (m), 776 (s), 724 (m), 690 (s), 664 (s), 615 (s), 527 (s); **HRMS** (ESI): berechnet $C_4H_2ClN_4^-$ [M-H]$^-$: 188.98207, gefunden: 188.98154.

N-(6-Chlor-3,4-dihydro-5-nitro-4-oxo-2-pyrimidinyl)isobutyramid (54)



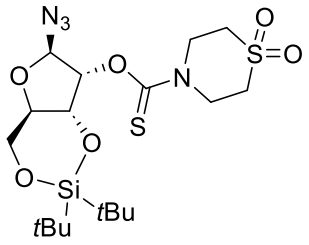
54

2-Amino-6-chlor-5-nitro-3H-pyrimidin-4-on **58** (2.92 g, 15.3 mmol, 1.00 eq.) wurde in Isobuttersäure (48.0 mL, 291 mmol, 19.0 eq.) suspendiert, zwei Tropfen konz. Schwefelsäure wurden zugegeben und das Reaktionsgemisch wurde bei 100 °C für 1 h gerührt. Anschließend wurde das Gemisch auf 50 °C gekühlt, Methanol (100 mL) wurde vorsichtig hinzugeben und für weitere 3 h bei Raumtemperatur gerührt. Das Methanol wurde

in vacuo entfernt und der Rückstand bei 5 °C für 48 h gelagert. Das präzipitierte Produkt wurde abfiltriert, mit kaltem Methanol gewaschen und getrocknet, um **54** (3.85 g, 14.8 mmol, 97 %) als gelbe Kristalle zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ(ppm) = 12.41 (s, 1H), 2.74 (sep, 1H, ^3J = 6.9 Hz), 1.32 (s, 1H), 1.11 (d, 6H, ^3J = 6.9 Hz, 6H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ(ppm) = 180.7, 152.9, 151.4, 150.9, 131.8, 35.1, 18.6; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3308 (w), 3150 (m), 1661 (s), 1595 (s), 1522 (s), 1459 (s), 1387 (m), 1351 (m), 1313 (m), 1249 (m), 1223 (m), 1177 (m), 1132 (s), 1042 (s), 953 (m), 917 (s), 872 (m), 832 (m), 782 (s), 749 (s), 724 (s), 687 (s), 625 (s), 591 (s), 537 (s), 500 (s), 425 (s); **HRMS** (ESI): berechnet für $C_8H_8ClN_4O_4^-$ [M-H]$^-$: 259.02341, gefunden: 259.02393.

1-Azido-2-*O*-(1,1-dioxo-1λ^6-thiomorpholine-4-carbothioat)-3,5-*O*-di-*tert*butylsilyl-*β*-D-ribofuranose (61)



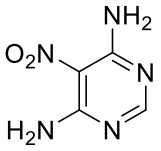
61

6-Azido-2,2-di-*tert*-butyltetrahydro-4H-furo[3,2-d][1,3,2]dioxasilin-7-ol (2.88 g, 9.13 mmol, 1.00 eq.) wurde in DCM gelöst (14 mL). Zu der Lösung wurde 1,1'-Thiocarbonyldiimidazol (1.79 g, 10.0 mmol, 1.10 eq.) und eine katalytische Menge DMAP (56.4 mg, 503 µmol, 0.06 eq.) gegeben. Die Reaktionslösung wurde zunächst 3 h bei Raumtemperatur und anschließend 1 h bei 50 °C gerührt. Anschließend wurde 1,1-Dioxothiomorpholine (1.36 g, 10.0 mmol, 1.10 eq.) in DCM (15 mL) hinzugegeben und das Gemisch bei Raumtemperatur über Nacht gerührt. Das Lösungsmittel wurde *in vacuo* entfernt, der Rückstand aus ACN umkristallisiert und mit kaltem ACN gewaschen. Das Rohprodukt wurde anschließend chromatographisch (Silikagel, *i*Hex:EtOAc = 5:1 → 1:1 → 0:1) aufgereinigt, um **61** (2.25 g, 4.57 mmol, 50 %) als einen farblosen Feststoff zu erhalten.

***R*$_f$** = 0.23 (*i*Hex:EtOAc = 5:1); **^{1}H NMR** (400 MHz, $CDCl_3$): δ (ppm) = 5.49-5.47 (d, ^3J = 4.5 Hz, 2H), 4.86 (dd, ^3J = 12.2 Hz, ^3J = 7.1 Hz, 1H), 4.51 (d, ^3J = 14.4 Hz, 1H), 4.45 (dd, ^3J = 8.6 Hz, ^3J = 4.4 Hz, 1H), 4.32 (td, ^3J = 10.3 Hz, ^3J = 9.1 Hz, ^3J = 5.3 Hz, 1H), 4.23 (dd, ^3J = 9.3 Hz, ^3J = 4.6 Hz, 1H), 4.14-3.92 (m, 3H), 3.30-2.90 (m, 4H), 1.07 (s, 9H), 0.98 (s, 9H); **^{13}C NMR** (101 MHz, $CDCl_3$): δ(ppm) = 187.0, 93.7, 82.5, 76.1, 74.7, 68.5, 51.9, 51.5, 48.5, 43.7, 27.5, 27.2, 22.9, 20.3; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2925 (w), 2856 (w), 2123 (w), 1492 (m), 1477 (w), 1463

(w), 1449 (w), 1438 (w), 1408 (w), 1385 (w), 1359 (w), 1319 (w), 1301 (w), 1276 (s), 1250 (w), 1230 (s), 1190 (s), 1161 (w), 1124 (s), 1091 (m), 1061 (s), 1049 (s), 1041 (s), 1014 (m), 983 (w), 949 (w), 935 (w), 923 (s), 885 (w), 865 (m), 847 (w), 824 (s), 812 (s), 796 (w), 744 (s), 722 (w), 713 (w), 705 (w), 667 (w); **HRMS** (ESI): berechnet für $C_{18}H_{33}N_4O_6S_2Si^+$ $[M+H]^+$: 493.1605, gefunden: 493.1606.

4,6-Diamino-5-nitropyrimidin (66)

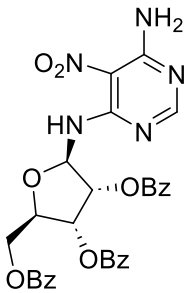


66

6-Diaminopyrimidine **65** (1.50 g, 13.6 mmol, 1.00 eq.) wurde portionsweise zu einer eiskalten Lösung von Schwefelsäure (konz., 10 mL) gegeben. Zu der Reaktionslösung wurde Salpetersäure (konz., 8 mL) tropfenweise hinzugegeben, während die Temperatur unter 50 °C gehalten wurde. Die Suspension wurde für 15 min bei Raumtemperatur gerührt und anschließend in ein Eis-Wasser Gemisch (150 mL) gegeben. Anschließend wurde Ammoniumhydroxid zugegeben bis der pH-Wert 9-10 betrug. Das Gemisch wurde über Nacht bei 5 °C gelagert, anschließend wurde der Feststoff abfiltriert, mit eiskaltem Wasser gewaschen und getrocknet, um **66** (2.11 g, 13.6 mmol, quant.) als einen weißen Feststoff erhalten.

^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm) = 8.48 (s, 2H), 8.42 (s, 2H), 7.87 (s, 1H); **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3450 (m), 3260 (w), 2973 (m), 2630 (w), 2035 (w), 1614 (s), 1551 (s), 1533 (s), 1513 (s), 1395 (m), 1371 (s), 1340 (w), 1251 (s), 1177 (s), 1128 (w), 1012 (s), 905 (w), 813 (s), 794 (s), 736 (w), 705 (w), 668 (w); **HRMS** (EI): berechnet für $C_4H_5N_5O_2^+$ $[M]^+$: 155.0443, gefunden: 155.0438.

6-Amino-5-nitro-4-[(2,3,5-tri-*O*-benzoyl-*β*-D-ribofuranosyl)amino]pyrimidin (68)

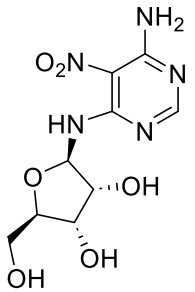


68

4,6-Diamino-5-nitropyrimidine **66** (1.00 g, 6.45 mmol, 1.00 eq.) wurde bei 0 °C in trockenem Acetonitril gelöst (17 mL). Anschließend wurde Triethylamin (0.90 mL, 6.49 mmol, 1.01 eq.) und TMSOTf (2.30 mL, 12.9 mmol, 2.00 eq.) tropfenweise hinzugegeben. Nachdem 1 h bei Raumtemperatur gerührt wurde, wurde 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-*β*-D-ribofuranose **67** (3.25 g, 6.45 mmol, 1.00 eq.) hinzugegeben. Das Reaktionsgemisch wurde langsam auf Raumtemperatur erwärmt und anschließend 25 h gerührt. Anschließend wurde die Reaktion mit einer eiskalten Mischung aus Acetonitril/Ammoniumhydroxid (10/1, 44 mL) gequencht und mit Wasser und Ethylacetat versetzt. Die wässrige Phase wurde mit Ethylacetat (3 × 100 mL) extrahiert und die vereinten organischen Phasen wurden mit einer ges. NaCl-Lösung (50 mL) gewaschen, über $MgSO_4$ getrocknet, filtriert und unter verminderten Druck eingeengt. Das Rohprodukt wurde anschließend über Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 1:1) aufgereinigt, um **68** (3.03 g, 5.06 mmol, 78 %) als einen hellgelben Feststoff zu erhalten.

R_f = 0.41 (*i*Hex:EtOAc = 1:1); **^{1}H NMR** (600 MHz, DMSO-d_6): δ (ppm) = 9.62 (d, ^3J = 8.2 Hz, 1H), 8.61 (br s, 1H), 8.55 (brs, 1H), 8.05 (s, 1H), 8.03-7.38 (m, 15H), 6.36 (dd, ^3J = 8.3 Hz, ^3J = 5.0 Hz, 1H), 6.00 (dd, ^3J = 6.0 Hz, ^3J = 4.9 Hz, 1H), 5.93-5.89 (t, 1H), 4.67-4.60 (m, 2H), 4.60-4.54 (m, 1H); **^{13}C NMR** (151 MHz, DMSO-d_6): δ (ppm) = 165.5, 164.7, 159.1, 158.5, 156.5, 133.9, 133.8, 133.6, 129.3, 129.2, 128.8, 128.7, 128.6, 112.5, 83.9, 78.0, 74.0, 71.0, 63.9; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3449 (w), 3330 (w), 3060 (w), 1720 (s), 1600 (m), 1575 (s), 1514 (m), 1450 (w), 1353 (w), 1314 (w), 1244 (s), 1176 (w), 1092 (s), 1068 (s), 1024 (s), 1000 (m), 893 (w), 825 (w), 798 (m), 705 (s), 685 (s), 667 (m); **HRMS** (ESI): berechnet für $C_{30}H_{26}N_5O_9^+$ $[M+H]^+$: 600.1725, gefunden: 600.1722.

6-Amino-5-nitro-4-(*β*-D-ribofuranosylamino)pyrimidine (69)



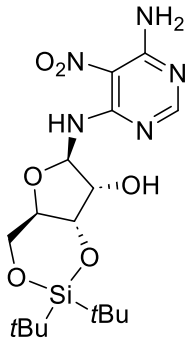
69

68 (1.61 g, 2.69 mmol, 1.00 eq.) wurde in Dioxan (13 mL) und Methanol (64 mL) gelöst. Eine frisch hergestellte Natriummethanolat-Lösung (1 M in Methanol, 0.4 mL, 539 µmol, 0.20 eq.) wurde tropfenweise bei 0 °C hinzugegeben. Das Reaktionsgemisch wurde langsam auf Raumtemperatur erwärmt und über Nacht gerührt. Anschließend wurde der pH-Wert durch

Zugabe von Dowex® 50W-X 8 (H^+-Form) auf 6-7 eingestellt und das Lösungsmittel *in vacuo* entfernt. Der Rückstand wurde über Säulenchromatographie (Silikagel, DCM:MeOH = 85:15) aufgereinigt, um **69** (735 mg, 2.56 mmol, 95%) als einen farblosen Feststoff zu erhalten.

R_f = 0.29 (DCM:MeOH = 85:15); **^{1}H NMR** (600 MHz, DMSO-d_6): δ (ppm) = 9.29 (d, ^3J = 7.7 Hz, 1H), 8.57 (s, 2H), 8.00 (s, 1H), 5.78 (dd, ^3J = 7.7 Hz, ^3J = 3.6 Hz, 1H), 5.21 (d, ^3J = 5.4 Hz, 1H), 5.08 (t, ^3J = 5.0 Hz, 1H), 4.93 (d, ^3J = 5.9 Hz, 1H), 4.07 (q, ^3J = 5.3 Hz, 1H), 3.92 (td, ^3J = 5.1 Hz, ^3J = 3.5 Hz, 1H), 3.79 (dt, ^3J = 5.7 Hz, ^3J = 3.0 Hz, 1H), 3.54 (ddd, ^3J = 11.6 Hz, ^3J = 4.8 Hz, ^3J = 3.3 Hz, 1H), 3.45 (ddd, ^3J = 11.7 Hz, ^3J = 5.3 Hz, ^3J = 2.8 Hz, 1H); **^{13}C NMR** (151 MHz, DMSO-d_6): δ (ppm) = 159.3, 158.6, 155.8, 111.8, 86.1, 83.9, 74.9, 69.9, 60.3; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3438 (w), 3320 (w), 3123 (w), 2207 (w), 1976 (w), 1644 (m), 1601 (s), 1514 (s), 1347 (w), 1301 (m), 1238 (s), 1151 (m), 1111 (m), 1022 (s), 898 (w), 868 (m), 823 (m), 794 (s), 742 (w), 661 (s), **HRMS** (ESI): berechnet für $C_9H_{14}N_5O_6^+$ $[M+H]^+$:288.0939, gefunden: 288.0940.

6-Amino-5-nitro-4-[(3,5-*O*-di-*tert*-butylsilyl-*β*-D-ribofuranosyl)amino]pyrimidin (70)



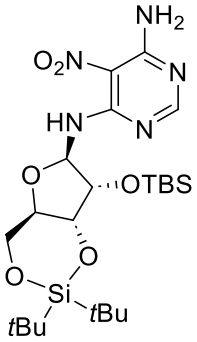
70

Zu einer Lösung von **69** (1.02 g, 3.55 mmol, 1.00 eq.) in DMF (50 mL) wurden Di-*tert*-butyldichlorsilan (0.90 mL, 4.26 mmol, 1.20 eq.) und $AgNO_3$ (1.45 g, 8.52 mmol, 2.40 eq.) zugegeben. Das Reaktionsgemisch wurde auf 0 °C gekühlt und anschließend Triethylamin (1.20 mL, 8.66 mmol, 2.44 eq.) tropenweise hinzugegeben. Das resultierende Gemisch wurde für 2 h bei 0 °C gerührt. Das Präzipitat wurde abfiltriert und das Filtrat *in vacuo* eingeengt. Der Rückstand wurde über Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 1:1 → 1:2) aufgereinigt, um **70** (1.28 g, 2.99 mmol, 84 %) als einen farblosen Feststoff zu erhalten.

R_f = 0.46 (*i*Hex:EtOAc = 1:2); **^{1}H NMR** (800 MHz, $CDCl_3$): δ (ppm) = 10.23 (d, ^3J = 8.0 Hz, 1H), 8.47 (br s, 1H), 8.14 (s, 1H), 6.29 (dd, ^3J = 8.0 Hz, ^3J = 4.4 Hz, 1H), 6.16 (s, 1H), 4.43 (t, ^3J = 4.3 Hz, 1H), 4.41-4.36 (m, 1H), 4.15-3.98 (m, 2H), 3.99-3.90 (m, 1H), 1.08 (s, 9H), 1.04

(s, 9H); **^{13}C NMR** (201 MHz, $CDCl_3$): δ (ppm) = 160.1, 159.0, 157.1, 82.0, 77.6, 73.1, 69.1, 67.8, 27.5, 27.4, 22.9, 20.5; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3292 (w), 2932 (w), 2892 (w), 2857 (w), 2360 (w), 2186 (w), 1727 (w), 1664 (w), 1621 (m), 1583 (s), 1515 (m), 1470 (w), 1386 (w), 1360 (w), 1273 (w), 1241 (m), 1203 (w), 1146 (m), 1104 (m), 1038 (s), 1011 (w), 971 (w), 938 (s), 901 (w), 879 (w), 855 (w), 825 (s), 794 (s), 745 (s); **HRMS** (ESI): berechnet für $C_{17}H_{30}N_5O_6Si^+$ $[M+H]^+$: 428.1960, gefunden: 428.1960.

6-Amino-5-nitro-4-[(2-*tert*-butyldimethylchlorsilyl-3,5-*O*-di-*tert*-butylsilyl-*β*-D-ribofuranosyl)amino]pyrimidin (71)



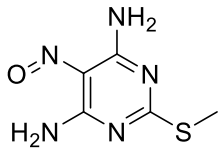
71

70 (498 mg, 1.16 mmol, 1.00 eq.) wurde in DMF (13 mL) gelöst. Zu der Lösung wurden Imidazol (403 mg, 5.86 mmol, 5.05 eq.) und *tert*-Butyldimethylchlorsilan (213 mg, 1.41 mmol, 1.22 eq.) gegeben. Das resultierende Gemisch wurde für 23 h bei Raumtemperatur gerührt. Anschließend wurde weiteres *tert*-Butyldimethylchlorsilan (213 mg, 1.41 mmol, 1.22 eq.) hinzugegeben und das Reaktionsgemisch bei 55 °C für 48 h erhitzt. Das Gemisch wurde mit Wasser und Ethylacetat verdünnt. Die wässrige Phase wurde mit Ethylacetat (3 × 100 mL) extrahiert. Die vereinten organischen Phasen wurden mit einer ges. NaCl-Lösung (50 mL) gewaschen, über $MgSO_4$ getrocknet und unter verminderten Druck eingeengt. Der Rückstand wurde über Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 4:1 → 2:1 → 1:1) aufgereinigt, um **71** (30.2 mg, 55.7 µmol, 5 %) als einen farblosen Feststoff zu erhalten.

R_f = 0.61 (*i*Hex:EtOAc = 5:1); **^{1}H NMR** (800 MHz, $CDCl_3$): δ (ppm) = 9.02 (d, ^3J = 7.0 Hz, 1H), 8.46 (s, 1H), 8.08 (s, 1H), 6.09 (s, 1H), 5.84 (d, ^3J = 7.0 Hz, 1H), 4.44 (dd, ^3J = 9.2 Hz, ^3J = 5.1 Hz, 1H), 4.25 (d, ^3J = 4.4 Hz, 1H), 4.12 (td, ^3J = 10.1 Hz, ^3J = 5.1 Hz, 1H), 3.95-3.87 (m, 2H), 1.11-1.04 (m, 9H), 1.02 (s, 9H), 0.94 (s, 9H), 0.16 (d, ^3J = 16.7 Hz, 6H); **^{13}C NMR** (201 MHz, $CDCl_3$): δ (ppm) = 160.0, 159.0, 155.9, 89.9, 76.3, 74.0, 68.7, 27.4, 27.2, 26.0, 22.9, 20.5, 18.5, -4.1, -4.8; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3424 w), 2924 (m), 2854 (m), 2190 (w), 2179 (w), 2168 (w), 2160 (w), 2017 (w), 1712 (s), 1646 (w), 1580 (s), 1520 (w), 1472 (w), 1358 (s), 1290 (w),1249 (m), 1220 (s), 1141 (m), 1110 (s), 1058 (s), 1008 (s), 937 (w), 890 (m), 846 (s),

827 (s), 798 (m), 791 (m), 778 (s), 756 (m),732 (w), 694 (w), 672 (w); **HRMS** (ESI): berechnet für $C_{23}H_{44}N_5O_6Si_2^+$ $[M+H]^+$: 542.2825, gefunden: 542.2824.

4,6-Diamino-2-(methylthio)-5-nitrosopyrimidin (78)

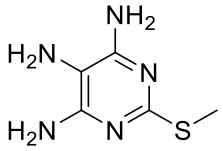


78

Zu einer Lösung aus 4,6-Diamino-2-(methylthio)pyrimidin **78** (22.0 g, 141 mmol, 1.00 eq.) in Essigsäure (90 mL) und Wasser (910 mL) wurde Natriumnitrit (21.4 g, 319 mmol, 2.20 eq.) portionsweise bei 0 °C gegeben. Die Reaktionslösung wurde zunächst bei 0 °C für 1 h und anschließend 4 h bei Raumtemperatur gerührt. Das Präzipitat wurde abfiltriert, mit Wasser gewaschen und unter Hochvakuum getrocknet, um **78** (21.3 g, 115 mmol, 82 %) als ein blaues Pulver zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 10.17 (s, 1H), 8.98 (s, 1H), 8.40 (s, 1H), 8.00 (s, 1H), 2.46 (s, 3H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) = 178.6, 164.3, 145.8, 138.9, 13.6; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2360 (m), 1527 (s), 1320 (m), 1284 (s), 1199 (s), 964 (m), 955 (s), 784 (s), 668 (w), 639 (s), 554 (s), 495 (s), 413 (w); **HRMS** (EI): berechnet für $C_5H_7N_5OS^+$ $[M]^+$: 185.03713, gefunden: 185.03653.

4,5,6-Triamino-2-methylthiopyrimidin (79)

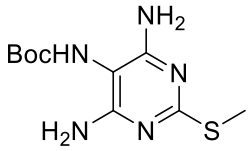


79

78 (13.0 g, 70 mmol, 1.00 eq.) wurde in Wasser (80 mL) und Essigsäure (40 mL) gelöst. Zinkpulver (18.4 g, 281 mmol, 4.00 eq.) wurde bei 0 °C hinzugegeben. Das Reaktionsgemisch wurde langsam über 3 h auf Raumtemperatur erwärmt und anschließend über Celite abfiltriert. Der pH-Wert des Filtrats wurde durch Zugabe von Ammoniumhydroxid auf pH 9 gebracht. Daraufhin wurde das Filtrat bei 0 °C über Nacht gelagert und das gebildete Präzipitat mit kaltem Wasser gewaschen, um **79** (8.49 g, 49.6 mmol, 71 %) als einen farblosen kristallinen Feststoff zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 5.68 (s, 4H), 3.52 (s, 2H), 2.31 (s, 3H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) = 157.0, 152.8, 102.8, 13.2; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2970 (w), 1738 (s), 1303 (m), 1228 (s), 1217 (s), 1023 (s), 993 (s), 824 (m), 763 (s), 441 (s), 435 (s), 418 (s), 406 (s); **HRMS** (ESI): berechnet für $C_5H_{10}N_5S^+$ [M+H]$^+$: 172.06514, gefunden: 172.06515.

tert-Butyl-(4,6-diamino-2-(methylthio)pyrimidin-5-yl)carbamat (80)

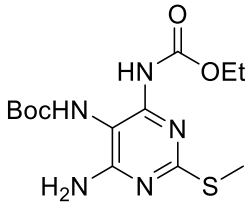


80

Zu einer Lösung aus **79** (8.40 g, 49.0 mmol, 1.00 eq.) in *tert*-Butanol (240 mL) wurde Di-*tert*-butyldicarbonat (11.6 mL, 54.0 mmol, 1.10 eq.) tropfenweise hinzugegeben. Anschließend wurde das Reaktionsgemisch für 3 h bei 60 °C gerührt und auf Raumtemperatur abgekühlt. Das Präzipitat wurde abfiltriert. Das Filtrat wurde auf 100 mL eingeengt, zentrifugiert und der Überstand entfernt. Die vereinigten Präzipitate wurde mit kaltem *tert*-Butanol gewaschen und unter Hochvakuum getrocknet, um **80** (12.3 g, 45.3 mmol, 92 %) in Form eines farblosen Feststoffes zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 7.56 (s, 1H), 5.84 (s, 4H), 2.34 (s, 3H), 1.42 (s, 9H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) =164.8, 159.8, 153.9, 92.5, 78.5, 28.2, 13.3; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3510 (s), 3446 (w), 3336 (m), 3266 (s), 2996 (w), 2981 (w), 1704 (s), 1603 (s), 1556 (s), 1375 (s), 1386 (w), 1366 (m), 1327 (m), 1303 (m), 1266 (m), 1240 (s), 1153 (s), 1053 (m), 1018 (m), 952 (m), 899 (m), 820 (m), 784 (m), 751 (w), 616 (m), 560 (m), 497 (w), 456 (m), 426 (s); **HRMS** (EI): berechnet für $C_{10}H_{17}N_5O_2S^+$ [M]$^+$: 271.11030, gefunden: 271.11005.

tert-Butyl-ethyl-(6-amino-2-(methylthio)pyrimidin-4,5-yl)carbamat (81)

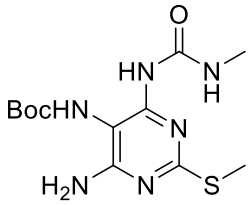


81

80 (12.3 g, 45.3 mmol, 1.00 eq.) wurde in Pyridin (100 mL) gelöst und Ethylchloroformiat (6.47 mL, 68.0 mmol, 1.50 eq.) bei 0 °C tropfenweise hinzugegeben. Die Reaktionslösung wurde für 3 h auf 60 °C erhitzt. Anschließend wurde das Lösungsmittel unter verminderten Druck entfernt und der Rückstand mittels Säulenchromatographie (Silikagel, DCM:MeOH = 99:1) aufgereinigt, um **81** (7.03 g, 20.5 mmol, 45 %) in Form eines farblosen Feststoffes zu erhalten.

***R*f** = 0.30 (DCM:MeOH = 99:2); **^{1}H NMR** (500 MHz, $CDCl_3$): δ (ppm) = 7.35 (s, 1H), 7.06 (s, 1H), 5.27 (s, 2H), 4.24 (q, *J* = 7.2 Hz, 2H), 2.43 (s, 3H), 1.44 (s, 9H), 1.30 (t, *J* = 7.2 Hz, 3H); **^{13}C NMR** (126 MHz, $CDCl_3$): δ (ppm) = 168.2, 161.5, 153.8, 151.1, 149.8, 128.3, 123.7, 102.5, 81.0, 62.5, 28.2, 14.3, 14.2; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2359 (w), 1717 (m), 1602 (m), 1366 (s), 1333 (m), 1307 (m), 1237 (s), 1161 (m), 1095 (m), 1033 (s), 909 (m), 882 (m), 826 (m), 773 (m), 730 (s), 701 (m), 668 (m), 582 (m), 413 (m); **HRMS** (EI): berechnet für $C_{13}H_{21}N_5O_4S^+$ $[M]^+$: 343.13143, gefunden: 343.13068.

tert-Butyl-(4-amino-2-(methylthio)-6-(3-methylureido)pyrimidin-5-yl)carbamate (82)

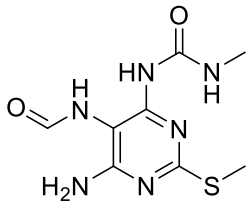


82

81 (1.00 g, 2.91 mmol, 1.00 eq.) wurde in einem Druckröhrchen in Methylamin (33 % in MeOH, 10 mL) suspendiert. Das Reaktionsgemisch wurde für 3 h bei 50 °C erhitzt. Anschließend wurde das Lösungsmittel unter reduziertem Druck entfernt und der Rückstand mittels Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 5:1 → 4:1) aufgereinigt, um **82** (84.0 mg, 256 µmol, 9 %) in Form eines farblosen Feststoffes zu erhalten.

***R*f** = 0.33 (*i*Hex:EtOAc = 4:1); **^{1}H NMR** (500 MHz, DMSO-d_6): δ (ppm) = 8.77 (q, *J* = 4.8 Hz, 1H), 7.78 (s, 1H), 7.69 (s, 1H), 6.67 (m, 2H), 2.75 (d, *J* = 4.7 Hz, 3H), 2.42 (s, 3H), 1.42 (s, 9H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) = 165.8, 160.7, 154.2, 153.3, 94.7, 79.3, 28.0, 26.1, 13.6; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3433 (w), 2970 (w), 2930 (w), 1740 (s), 1720 (s), 1671 (m), 1625 (m), 1595 (m), 1533 (s), 1487 (s), 1416 (m), 1392 (m), 1366 (s), 1339 (m), 1290 (m), 1231 (s), 1217 (s), 1152 (s), 1080 (m), 1051 (m), 991 (m), 962 (m) 905 (m), 834 (w), 801 (m), 783 (w), 765 (s), 694 (m), 616 (s), 527 (m), 504 (m), 458 (m), 406 (m); **HRMS** (EI): berechnet für $C_{12}H_{20}N_6O_3S^+$ $[M]^+$: 328.13176, gefunden: 328.13089.

1-(5,6-Diamino-2-(methylthio)pyrimidin-4-yl)-3-methylharnstoff (76)

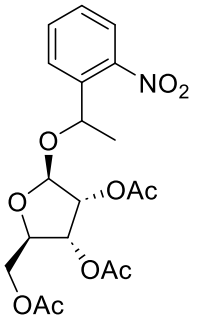


76

Zu einer Lösung aus **82** (50 mg, 152 µmol, 1.00 eq.) in Methanol (4 mL) wurde Trimethylsilylchlorid (3 mL, 23.6 mmol, 155 eq.) hinzugegeben. Die resultierende Lösung wurde für 2 h bei Raumtemperatur gerührt und anschließend das Lösungsmittel unter reduziertem Druck entfernt. Der Rückstand wurde in DCM (4 mL) resuspendiert und das Gemisch zentrifugiert. Der Überstand wurde entfernt, der Rückstand unter Hochvakuum getrocknet und in Ameisensäure (200 µL) gelöst. Natriumformiat (12.5 mg, 184 µmol, 1.20 eq.) wurde hinzugegeben und das Gemisch für 2 h bei 50 °C erhitzt. Das Lösungsmittel wurde unter reduziertem Druck entfernt und der Rückstand in Wasser (1 mL) suspendiert. Das Gemisch wurde zentrifugiert, der Überstand wurde entfernt und der Rückstand unter Hochvakuum getrocknet, um **76** (31.1 mg, 121 µmol, 79 %) in Form eines farblosen Feststoffes zu erhalten.

^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 8.93 (q, *J* = 4.7 Hz, 1H), 8.76 (s, 1H), 8.19 (s, 1H), 8.07 (s, 1H), 6.75-6.70 (m, 2H), 2.75 (d, *J* = 4.6 Hz, 3H), 2.42 (s, 3H); **^{13}C NMR** (126 MHz, DMSO-d_6): δ (ppm) = 166.6, 161.7, 160.3, 154.4, 153.5, 92.9, 26.0, 13.6; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3390 (w), 3275 (w), 3161 (w), 2970 (w), 1739 (m), 1708 (m), 1672 (s), 1651 (s), 1600 (m), 1495 (s), 1437 (m), 1342 (s), 1296 (m), 1264 (m), 1230 (m), 1217 (m), 1161 (m), 1139 (m), 1039 (m), 1073 (m), 997 (m), 899, (w), 870 (w), 799 (m), 758 (s), 622 (s), 584 (s), 527 (s), 467 (m), 448 (m), 423 (m); **HRMS** (EI): berechnet für $C_8H_{12}N_6O_2S^+$ $[M]^+$: 256.07424, gefunden: 256.07361.

1'-[-1-(2-Nitrophenyl)ethyl]-2',3',5'-tri-*O*-acetyl-ribofuranose (85)

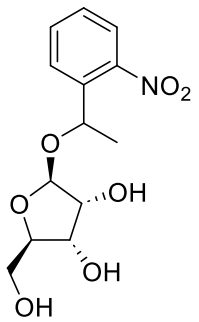


85

β-D-Ribofuranose-1,2,3,5-tetraacetat **84** (11.6 g, 36.5 mmol, 1.22 eq.) wurde unter Stickstoff Atmosphäre zu einer Lösung aus 1-(2-Ntrophenyl)ethanol (5.00g, 29.9 mmol, 1.00 eq.) in Acetonitril (225 mL) gegeben. Das resultierende Gemisch wurde auf -20 °C gekühlt und Trimethylsilyltrifluormethansulfonat (1.90 mL, 10.5 mmol, 0.35 eq.) tropfenweise über 1 h hinzugegeben. Nach 4 h wurde das Reaktionsgemisch mit Ethylacetat (300 mL) verdünnt und mit einer ges. $NaHCO_3$-Lösung (2 × 100 mL) gewaschen. Die wässrige Phase wurde anschließend mit Ethylacetat (2 × 100 mL) extrahiert und die vereinten organischen Phasen wurden über Na_2SO_4 getrocknet und unter verminderten Druck eingeengt. Der Rückstand wurde über Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 3:1 → 2:1 → 1:1 → 1:2) aufgereinigt, um **85** (10.5 g, 24.6 mmol, 82 %) als ein gelbliches Öl zu erhalten. Das Produkt wurde als Diastereomerengemisch charakterisiert.

R_f = 0.21 (*i*Hex:EtOAc = 3:1); **^{1}H NMR** (500 MHz, $CDCl_3$): δ (ppm) = 7.98-7.92 (m, 2H), 7.70-7.76 (m, 2H), 7.69-7.61 (m, 2H), 7.42 (m, 2H), 5.42 (q, *J* = 6.4 Hz, 1H), 5.38-5.35 (m, 2H), 5.30 (dd, *J* = 4.8 Hz, *J* = 1.1 Hz, 1H), 5.25 (dd, *J* = 4.8 Hz, *J* = 1.1 Hz, 1H), 5.20 (dd, *J* = 7.1 Hz, *J* = 4.8 Hz, 1H), 5.17 (d, *J* = 1.1 Hz, 1H), 4,73 (brs, 1H), 4.35-4.27 (m, 2H), 4.23-4.16 (m, 1H), 4.02 (dd, *J* = 11.8 Hz, *J* = 3.7 Hz, 1H), 3.63 (dd, *J* = 11.8 Hz, *J* = 6.6 Hz, 1H), 2.14 (s, 3H), 2.11 (s, 3H), 2.05 (s, 3H), 2.04 (s, 3H), 2.02 (s, 3H), 1.92 (s, 3H), 1.55 (d, *J* = 6.0 Hz, 3H), 1.53 (d, *J* = 6.0 Hz, 3H); **^{13}C NMR** (126 MHz, $CDCl_3$): δ (ppm) = 170.8, 170.6, 169.9, 169.8, 169.8, 169.78, 148.6, 147.4, 139.6, 138.4, 133.9, 133.6, 128.6, 128.5 128.2, 128.1, 124.5, 124.2, 104.1, 103.1, 78.6, 78.4, 75.0, 74.9, 71.6, 71.5, 71.2, 70.1, 64.7, 64.4, 24.0, 22.6, 21.0, 20.7, 20.7, 20.6, 20.6, 20.6; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2360 (m), 1743 (s), 1526 (m), 1368 (m), 1214 (s), 1041 (m), 962 (m), 898 (m), 855 (w), 789 (w), 749 (w), 707 (w), 668 (w), 602 (w), 429 (w), 424 (w), 418 (w), 413 (w), 402 (w); **HRMS** (ESI): berechnet für $C_{19}H_{23}NNaO_{10}^+$ $[M+Na]^+$: 448.12197, gefunden: 448.12151.

1'-[-1-(2-Nitrophenyl)ethyl]-ribofuranose (86)



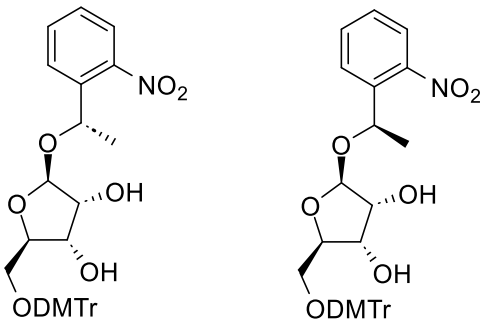
86

Zu einer Lösung aus **85** (10.3 g, 24.2 mmol, 1.00 eq.) in Methanol (12 mL) wurde Na_2CO_3 (2.57 g, 24.2 mmol, 1.00 eq.) hinzugegeben. Das Reaktionsgemisch wurde über Nacht bei

Raumtemperatur gerührt und anschließend mit Ethylacetat (400 mL) verdünnt. Das Gemisch wurde mit einer ges. $NaHCO_3$-Lösung (4 × 150 mL) gewaschen, die vereinten organischen Phasen über $MgSO_4$ getrocknet und *in vacuo* eingeengt. Der Rückstand wurde mittels Säulenchromatographie (Silikagel, EtOAc) aufgereinigt, um **86** (5.07 g, 16.9 mmol, 70 %) als einen farblosen Feststoff zu erhalten. Das Produkt wurde als Diastereomerengemisch charakterisiert.

R_f = 0.27 (EtOAc); **^{1}H NMR** (500 MHz, $CDCl_3$): δ (ppm) = 7.96-7.89 (m, 2H), 7.79-7.70 (m, 4H), 7.57-7.49 (m, 2H), 5.17 (dq, *J* = 15.8 Hz, *J* = 6.4 Hz, 2H), 5.05 (d, *J* = 4.3 Hz, 1H), 4.96 (d, *J* = 4.4 Hz, 1H), 4.90 (d, *J* = 1.0 Hz, 1H), 4.91-4.78 (m, 2H), 4.69 (t, *J* = 5.5 Hz, 1H), 4.42 (d, *J* = 1.0 Hz, 1H), 4.37 (t, *J* = 5.5 Hz, 1H), 3.92 (td, J = 6.9, 4.7 Hz, 1H), 3.82-3.72 (m, 4H), 3.66 (td, J = 6.0, 4.3 Hz, 1H), 3.56 (ddd, *J* = 11.6 Hz, *J* = 5.6 Hz, *J* = 3.7 Hz, 1H), 3.40 (dt, *J* = 11.7 Hz, *J* = 5.9 Hz, 1H), 3.26 (ddd, *J* = 11.4 Hz, *J* = 5.6 Hz, *J* = 4.4 Hz, 1H), 3.01 (dt, *J* = 11.6 Hz, *J* = 5.9 Hz, 1H), 1.43 (d, *J* = 2.8 Hz, 3H), 1.41 (d, *J* = 2.6 Hz, 3H); **^{13}C NMR** (126 MHz, $CDCl_3$): δ (ppm) = 148.4, 147.2, 139.1, 137.9, 133.7,133.4, 128.7, 128.4, 128.3, 128.1, 123.9, 123.8, 105.6, 104.7, 83.7, 83.7, 74.7, 74.3, 71.1, 70.9, 69.5, 68.5, 63.1, 62.9, 23.7, 22.3; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2359 (w), 1523 (m), 1346 (m), 1090 (m), 1023 (s), 997 (s), 855 (m), 823 (w), 433 (m), 422 (m), 415 (m), 409 (m), 402 (m); **HRMS** (ESI): berechnet für $C_{13}H_{17}NNaO_7^+$ $[M+Na]^+$: 322.09027, gefunden: 322.08986.

1'-[-1-(2-Nitrophenyl)ethyl]- 5'-[(4,4'-dimethoxy)triphenyl]-ribofuranose (87)



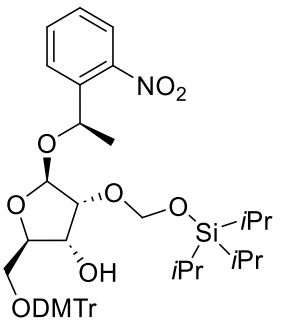
(*S*)-87 **(*R*)-87**

DMT-Cl (8.58 g, 25.3 mmol, 1.50 eq.) wurde 30 min unter Hochvakuum getrocknet und anschließend zu einer Lösung aus **86** (5.05 g, 16.9 mmol, 1.00 eq.) in Pyridin (50 mL) hinzugegeben. Das Gemisch wurde für 2 h bei Raumtemperatur gerührt und anschließend das Lösungsmittel unter reduziertem Druck entfernt. Das Rohprodukt wurde mittels Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 3:1 → 2:1→ 1:1 + 1‰ NEt_3) aufgereinigt, um das *(R)*-Isomer **(*R*)-87** (3.50 g, 5.83 mmol, 35 %) und das *(S)*-Isomer **(*S*)-87** (3.35 g, 5.58 mmol, 33 %) als gelbliche Öle zu erhalten.

***(R)*-Isomer:**

R_f = 0.42 (*i*Hex:EtOAc= 2:1 + 1‰ NEt_3); **^{1}H NMR** (500 MHz, $CDCl_3$): δ (ppm) = 7.87 (dd, *J* = 8.1 Hz, *J* = 1.3 Hz, 1H), 7.71 (dd, *J* = 8.0 Hz, *J* = 1.5 Hz, 1H), 7.63 (td, *J* = 7.6 Hz, *J* = 1.3 Hz, 1H), 7.45-7.27 (m, 6H), 7.18-7.14 (m, 4H), 6.84-6.80 (m, 4H), 5.31 (q, *J* = 6.3 Hz, 1H), 5.08 (s, 1H), 4.26-4.22 (m, 1H), 4.09-4.08 (m, 1H), 3.97-3.90 (m, 1H), 3.79 (s, 6H), 3.57-3.54 (m, 1H), 3.37- 3.33 (m, 1H), 1.53 (d, *J* = 6.5 Hz, 3H); **^{13}C NMR** (126 MHz, $CDCl_3$): δ (ppm) = 171.4, 158.8, 149.0, 147.8, 147.5, 139.6, 139.4, 133.5, 130.1, 130.1, 129.3, 128.4, 128.3, 128.0, 127.9, 127.2, 124.3, 113.3, 106.5, 83.9, 81.6, 75.8, 71.3, 71.0, 62.8, 55.4, 22.8; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2359 (w), 1608 (m), 1523 (m), 1508 (s), 1488 (m), 1445 (w), 1342 (m), 1297 (m), 1247 (s), 1175 (s), 1089 (s), 1030 (s), 953 (s), 855 (m), 828 (s), 789 (s), 750 (s), 701 (s), 679 (s), 606 (s), 584 (s), 467 (m), 434 (m), 425 (m), 403 (m); **HRMS** (ESI): berechnet für $C_{34}H_{34}NO_9^-$ $[M-H]^-$: 600.22391, gefunden: 600.22461.

1'-[-1-(2-Nitrophenyl)ethyl]-5'-[(4,4'-dimethoxy)triphenyl-2'-(triisopropylsilyloxy) methyl]-ribofuranose (88)

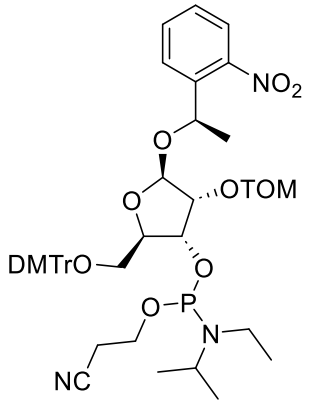


88

Zu einer Lösung aus **(*R*)-87** (4.12 g, 6.85 mmol, 1.00 eq.) in 1,2-Dichlorethan (70 mL) wurden DIPEA (3.58 mL, 20.54 mmol, 3.00 eq.) und Bu_2SnCl_2 (2.08 g, 6.85 mmol, 1.00 eq.) gegeben. Die Lösung wurde für 1 h bei Raumtemperatur gerührt. Anschließend wurde TOM-Cl (1.75 mL, 7.53 mmol, 1.10 eq.) hinzugegeben und die Lösung für weitere 30 min bei 80 °C gerührt. DCM (100 mL) wurde zum abgekühlten Gemisch gegeben und mit einer ges. $NaHCO_3$-Lösung (2 × 150 mL) gewaschen. Die wässrige Phase wurde anschließend mit DCM (2 × 150 mL) extrahiert und die vereinten organischen Phasen über $MgSO_4$ getrocknet. Das Lösungsmittel wurde unter reduziertem Druck entfernt und das Rohprodukt mittels wiederholter Säulenchromatographie (Silikagel, *i*Hex:EtOAc = 1:6→ 1:4 + 1‰ NEt_3) aufgereinigt, um das 2'-geschützte Produkt **88** (3.73 g, 4.73 mmol, 70 %) in Form eines gelblichen Gels zu isolieren.

R_f = 0.34 (*i*Hex:EtOAc= 4:1 + 1‰ NEt_3); **^{1}H NMR** (500 MHz, $CDCl_3$): δ (ppm) = 7.87-7.84 (m, 1H), 7.73-7.71 (m, 1H), 7.39-7.11 (m, 11H), 6.86-6.82 (m, 2H), 6.80-6.73 (m, 2H), 5.39 (q, *J* = 6.3 Hz, 1H), 5.20 (d, *J* = 2.0 Hz, 1H), 5.13 (d, *J* = 4.8 Hz, 1H), 5.01 (d, *J* = 4.8 Hz, 1H), 4.25 (q, *J* = 5.1 Hz, 1H), 4.16-4.09 (m, 1H), 3.97 (q, *J* = 4.7 Hz, 1H), 3.79 (s, 6H), 3.12 (dd, *J* = 10.0 Hz, *J* = 4.7 Hz, 1H), 2.97 (dd, *J* = 10.0 Hz, *J* = 4.5 Hz, 1H), 2.91 (d, *J* = 6.3 Hz, 1H), 1.51 (d, *J* = 6.5 Hz, 3H), 1.10-1.06 (m, 21H); **^{13}C NMR** (126 MHz, $CDCl_3$): δ (ppm) = 158.5, 147.2, 145.0, 139.6, 136.3, 136.2, 133.4, 130.2, 129.3, 128.7, 128.4, 128.0, 127.9, 127.8, 127.7, 127.2, 126.8, 124.2, 113.3, 113.1, 105.3, 90.9, 86.1, 84.2, 83.3, 72.0, 71.8, 64.2, 55.4, 55.3, 23.2, 18.0, 12.0; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 3952 (w), 2970 (m), 2943 (m), 2866 (m), 2359 (s), 1738 (s), 1608 (m), 1576 (w), 1559 (w), 1526 (m), 1508 (s), 1445 (w), 1365 (s), 1300 (m), 1247 (s), 1228 (s), 1217 (s), 1175 (m), 1034 (s), 957 (m), 882 (m), 855 (m), 827 (m), 788 (m), 747 (m), 701 (m), 684 (m), 668 (m), 583 (m), 454 (m), 441 (w), 429 (m), 413 (m); **HRMS** (ESI): berechnet für $C_{44}H_{57}NNaO_{10}Si^+$ $[M+Na]^+$: 810.36494, gefunden: 810.36490.

1'-[-1-(2-Nitrophenyl)ethyl]-5'-[(4,4'-dimethoxy)triphenyl-2'-(triisopropylsilyloxy) methyl]-ribofuranose-3'-(2-cyanoethyldiisopropylphosphoramidit) (89)



89

Das geschützte Nukleosid **88** (200 mg, 254 µmol, 1.00 eq.) wurde durch Lyophilisation aus Benzen (3 ×) vorgetrocknet und bei 0 °C unter Ar-Atmosphäre in DCM (2.5 mL) gelöst. DIPEA (177 µL, 1.02 mmol, 4.00 eq.) und CED-Cl (142 µL, 635 µmol, 4.00 eq.) wurden tropfenweise hinzugegeben. Die resultierende Lösung wurde für 2 h bei Raumtemperatur gerührt und anschließend mit wässriger ges. $NaHCO_3$-Lösung (50 mL) gequencht. Die wässrige Phase wurde mit DCM (3 × 50 mL) extrahiert. Anschließend wurden die vereinten organischen Phasen mit einer ges. NaCl-Lösung (50 mL) gewaschen, über $MgSO_4$ getrocknet und das Lösungsmittel *in vacuo* eingeengt. Der Rückstand wurde mittels Säulenchromatographie (Silikagel, *n*Hex:EtOAc = 1:4 + 1‰ NEt_3) aufgereinigt, um das Phosphoramidit **89** (214 mg, 216 µmol, 85 %) in Form eines gelblichen Gels zu erhalten.

R_f = 0.37 (*i*Hex:EtOAc= 4:1 + 1‰ NEt_3); **^{31}P NMR** (202 MHz, Acetone): δ (ppm) = 149.2, 148.7; **IR** (ATR): $\tilde{\nu}$ (cm^{-1}) = 2867 (m), 1608 (m), 1526 (m), 1508 (s), 1462 (m), 1365 (m), 1300 (w), 1248 (s), 1178 (s), 1034 (s), 882 (m), 827 (m), 790 (m), 748 (m), 702 (m), 584 (w), 445 (w), 431 (m), 413 (w), 402 (w); **HRMS** (ESI): berechnet für $C_{53}H_{75}N_3O_{11}PSi^+$ $[M+H]^+$: 988.49030, gefunden: 988.49081.

5.3. Charakterisierung der synthetisierten Oligonukleotide

Tabelle 1. Charakterisierung der synthetisierten Oligonukleotide **S1^{NPE}- S2^{NPE}** und **S1-S2**.

Strang	Sequenz (5‘→3‘)	berechnet $[M-H]^-$	gefunden (*m/z*)
S1^{NPE}	5‘-**X^{NPE}**AU CGC U-3‘	2194.3	2193.8
S2^{NPE}	5‘-CUU AC**X^{NPE}** CUG A-3‘	3134.4	3134.9
S1	5‘-**X**AU CGC U-3‘	2045.3	2044.8
S2	5‘-CUU AC**X** CUG A-3‘	2985.4	2985.0

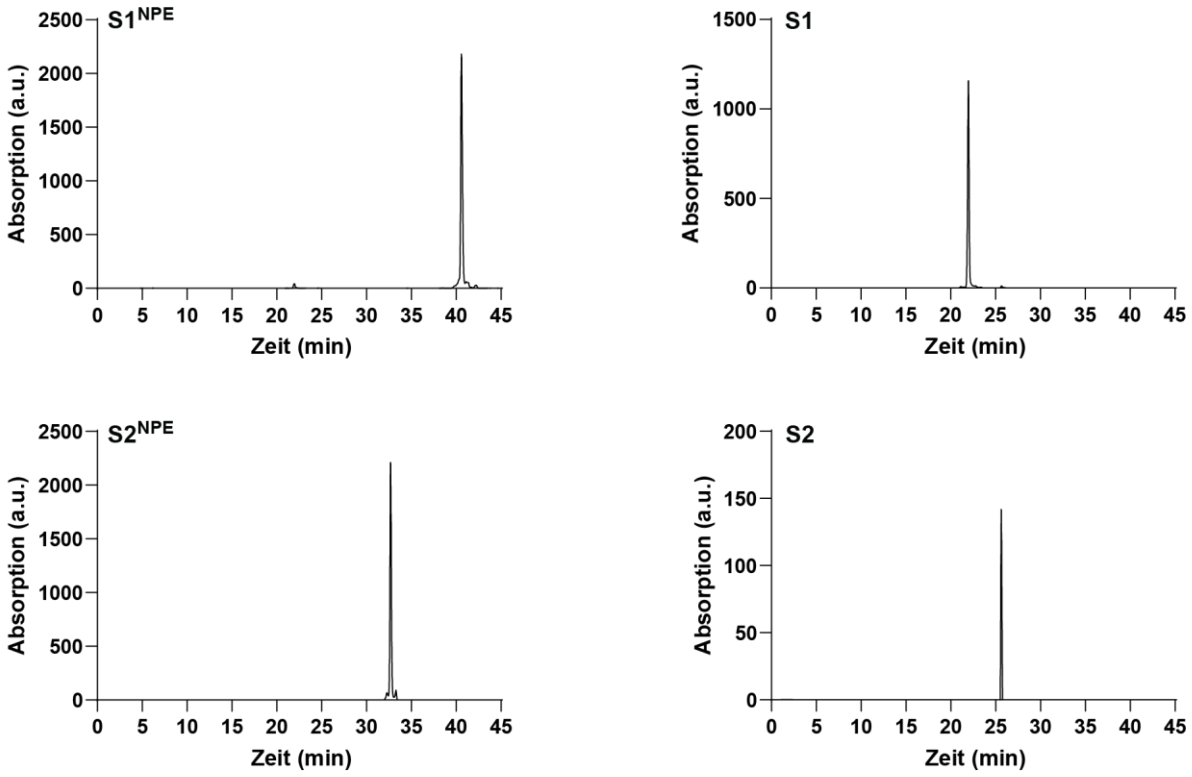


Abbildung 12. HPLC-Chromatogramme (0-30% Puffer B in 45 min) der aufgereinigten Oligonukleotide.

5.4. Präbiotische Reaktionen zur Beladung der abasischen Stellen

Die Oligonukleotide (5 nmol) in einer Lösung aus 50 mM Natriumborat Puffer (pH 8) und 50 mM NaCl gelöst. Die Proto-Nukleobasen wurden in Stocklösungen von 100 mM gelöst/suspendiert und zu der Reaktionsösung gegeben (250 mM). Die Reaktionsansätze wurden über eine *dry-down*-Methode bei 30 °C in einem *Thermoshaker* inkubiert. Danach wurde ein Aliquot (20 µL) entnommen und mittels RP-HPLC (Puffer A: 100 mM NEt_3/AcOH in H_2O (pH 7) und Puffer B: 100 mM $HCOONH_4$ in H_2O/ACN 20/80 (pH 7), Gradient: 0-40% Puffer B in 45 min) und MALDI-TOF-Massenspektrometrie analysiert.

6. Abkürzungsverzeichnis

A	Adenosin (Nukleosid) / Adenin (Nukleobase)
Ac	Acetyl
ACN	Acetonitril
a.u.	*arbitary unit* (Willkürliche Einheit)
br	breit (Signale bei NMR-Spektroskopie)
C	Cytidin (Nukleosid) / Cytosin (Nukleobase)
CE	Cyanoethyl
CED	2-Cyanethyl-*N*,*N*-diisopropylphosphoramidit
CPG	*controlled-pore-glass* (kontrolliertes poröses Glass)
d	Duplettlett (Signale bei NMR-Spektroskopie)
DC	Dünnschichtchromatographie
DCA	Dichloressigsäure
DCM	Dichlormethan
ddH_2O	Bidestilliertes Wasser
DMAP	4-(Dimethylamino)pyridin
DMF	Dimethylformamid
DMTr	4,4′-Dimethoxytrityl
DNA	Desoxyribonukleinsäure
eq.	Äquivalente
EtOAc	Ethylacetat
EI	Elektroionisation (Ionisierungsmethode für die Massenspektrometrie)
ESI	*electron spray ionisation* (Ionisierungsmethode für die Massenspektrometrie)
G	Guanosin (Nukleosid) / Guanin (Nukleobase)
Ga	Giga-Jahr / Jahrmilliarde
ges.	gesättigt
Hex	Hexan
HPLC	Hochleistungsflüssigkeitschromatographie
HR	hochaufgelöst (engl. high resolution)
IR	Infrarotspektroskopie
konz.	konzentriert
LC	*liquid chromatography*
LUCA	*last universal common ancestor*
m	medium (Signale bei IR-Spektroskopie)
m	Multiplett (Signale bei NMR-Spektroskopie)

MeOH	Methanol
MS	Massenspektrometrie
NPE	1-(2-nitropenyl)ethyl
ON	Oligonukleotid
ppm	*parts per million*
q	Quartett (Signale bei NMR-Spektroskopie)
quant.	Quantitativ
R_f	Retentionsfaktor (Dünnschichtchromatographie)
RP	*reverse phase*
RNA	Ribonukleinsäure
rRNA	*ribosomal* Ribonukleinsäure
s	stark (Signale bei IR-Spektroskopie)
s	Singulett (Signale bei NMR-Spektroskopie)
T	Temperatur
t	Triplett (Signale bei NMR-Spektroskopie)
TBS	*tert*-Butyldimethylsilyl
TC	(1,1-Dioxo-λ6-thiomorpholin-4-carbothioat)
THF	Tetrahydrofuran
tf	Triflourmethansulfonsäure
TMS	Trimethylsilan
TOF	*time of flight*
TOM	(Triisopropylsiloxy)methyl
tRNA	*transfer* Ribonukleinsäure
U	Uridin (Nukleosid) / Uracil (Nukleobase)
UV-Vis	*utraviolet-visible*
w	schwach (engl. *weak*) (Signale bei IR-Spektroskopie)

7. Literaturverzeichnis

[1] E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell*, Cambridge University Press, Cambridge, **1944**.

[2] F. Crick, *Nature* **1970**, *227*, 561-563.

[3] C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8392-8396.

[4] M. C. Weiss, F. L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W. F. Martin, *Nat. Microbiol.* **2016**, *1*, 16116.

[5] A. Becerra, L. Delaye, S. Islas, A. Lazcano, *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 361-379.

[6] C. P. McKay, *PLoS Biol.* **2004**, *2*, e302.

[7] S. A. Benner, *Astrobiology* **2010**, *10*, 1021-1030.

[8] D. W. Deamer, G. R. Fleischaker, *Origins of life : the central concepts*, Jones and Bartlett Publishers Boston, Boston, **1994**.

[9] J. Gómez-Márquez, *Mol. Biol. Rep.* **2021**, *48*, 6223-6230.

[10] M. Vitas, A. Dobovišek, *Orig. Life Evol. Biosph.* **2019**, *49*, 77-88.

[11] K. Michaelian, *Foundations* **2022**, *2*, 308-337.

[12] F. Wachowius, J. Attwater, P. Holliger, *Q. Rev. Biophys.* **2017**, *50*, e4.

[13] P. G. Higgs, N. Lehman, *Nat. Rev. Genet.* **2015**, *16*, 7-17.

[14] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99-123.

[15] G. F. Joyce, *Nature* **2002**, *418*, 214-221.

[16] J. W. Szostak, D. P. Bartel, P. L. Luisi, *Nature* **2001**, *409*, 387-390.

[17] A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821-12844.

[18] C. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6854-6859.

[19] G. Dalrymple, *Geol. Soc. Spec. Publ.* **2001**, *190*, 205-221.

[20] C. F. Chyba, *Geochim. Cosmochim. Acta* **1993**, *57*, 3351-3358.

[21] N. H. Sleep, K. J. Zahnle, J. F. Kasting, H. J. Morowitz, *Nature* **1989**, *342*, 139-142.

[22] D. Parkos, A. Pikus, A. Alexeenko, H. J. Melosh, *J. Geophys. Res. Planets* **2018**, *123*, 892-909.

[23] J. Geiss, A. P. Rossi, *Astron. Astrophys. Rev.* **2013**, *21*, 1-54.

[24] N. H. Sleep, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*.

[25] M. Dodd, D. Papineau, T. Greene, J. F. Slack, M. Rittner, F. Pirajno, J. O'Neil, C. Little, *Nature* **2017**, *546*.

[26] J. F. Kasting, *Science* **1993**, *259*, 920-926.

[27] J. Krissansen-Totton, G. N. Arney, D. C. Catling, *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 4105-4110.

[28] H. J. Cleaves, J. H. Chalmers, A. Lazcano, S. L. Miller, J. L. Bada, *Orig. Life Evol. Biosph.* **2008**, *38*, 105-115.

[29] J. F. Kasting, in *Earth's Early Atmosphere and Surface Environment, Vol. 504* (Ed.: G. H. Shaw), Geological Society of America, **2014**.

[30] J. M. Hollis, F. J. Lovas, P. R. Jewell, *Astrophys. J.* **2000**, *540*, L107.

[31] M. Guélin, J. Cernicharo, *Front. Astron. Space Sci.* **2022**, *9*.

[32] 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*, 779-784.

[33] R. Hanel, B. Conrath, F. M. Flasar, V. Kunde, W. Maguire, J. Pearl, J. Pirraglia, R. Samuelson, L. Herath, M. Allison, D. Cruikshank, D. Gautier, P. Gierasch, L. Horn, R. Koppany, C. Ponnamperuma, *Science* **1981**, *212*, 192-200.

[34] A. T. Tokunaga, S. C. Beck, T. R. Geballe, J. H. Lacy, E. Serabyn, *Icarus* **1981**, *48*, 283-289.

[35] R. Wordsworth, A. H. Knoll, J. Hurowitz, M. Baum, B. L. Ehlmann, J. W. Head, K. Steakley, *Nat. Geosci.* **2021**, *14*, 127-132.

[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] Z. R. Todd, K. I. Öberg, *Astrobiology* **2020**, *20*, 1109-1120.

[38] J. R. Cronin, *Adv. Space Res.* **1989**, *9*, 59-64.

[39] C. Ponnamperuma, *Ann. N. Y. Acad. Sci.* **1972**, *194*, 56-70.

[40] K. Kvenvolden, J. Lawless, K. Pering, E. Peterson, J. Flores, C. Ponnamperuma, I. R. Kaplan, C. Moore, *Nature* **1970**, *228*, 923-926.

[41] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *Science* **1966**, *154*, 784-785.

[42] S. L. Miller, G. Schlesinger, *Orig. Life Evol. Biosph.* **1984**, *14*, 83-90.

[43] F. Raulin, P. Bruston, *Adv. Space Res.* **1996**, *18*, 41-49.

[44] A. R. Bossard, F. Raulin, D. Mourey, G. Toupance, *J. Mol. Evol.* **1982**, *18*, 173-178.

[45] S. L. Miller, *Science* **1953**, *117*, 528-529.

[46] A. I. Oparin, *The Origin of Life*, Dover, New York, **1953**.

[47] J. B. S. Haldane, *Ration. Annu.* **1929**, *148*, 3-10.

[48] S. L. Miller, *J. Am. Chem. Soc.* **1955**, *77*, 2351-2361.

[49] S. L. Miller, *Ann. N. Y. Acad. Sci.* **1957**, *69*, 260-275.

[50] C. R. Woese, D. H. Dugre, W. C. Saxinger, S. A. Dugre, *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 966-974.

[51] F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367-379.

[52] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381-393.

[53] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147-157.

[54] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849-857.

[55] T. R. Cech, *Angew. Chem., Int. Ed.* **1990**, *29*, 759-768.

[56] S. Altman, *Angew. Chem., Int. Ed.* **1990**, *29*, 749-758.

[57] C. Guerrier-Takada, S. Altman, *Science* **1984**, *223*, 285-286.

[58] W. Gilbert, *Nature* **1986**, *319*, 618-618.

[59] M. P. Robertson, G. F. Joyce, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*.

[60] H. B. White, 3rd, *J. Mol. Evol.* **1976**, *7*, 101-104.

[61] A. Kirschning, *Angew. Chem., Int. Ed.* **2021**, *60*, 6242-6269.

[62] P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 920-930.

[63] H. S. Bernhardt, *Biol. Direct* **2012**, *7*, 23.

[64] A. Butlerow, *Liebigs Ann.* **1861**, *120*, 295-298.

[65] J. P. Pinto, G. R. Gladstone, Y. L. Yung, *Science* **1980**, *210*, 183-185.

[66] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**.

[67] R. Breslow, *Tetrahedron Lett.* **1959**, *1*, 22-26.

[68] R. Shapiro, *Orig. Life Evol. Biosph.* **1988**, *18*, 71-85.

[69] P. Decker, H. Schweer, R. Pohlamnn, *J. Chromatogr. A* **1982**, *244*, 281-291.

[70] G. Zubay, *Orig. Life Evol. Biosph.* **1998**, *28*, 13-26.

[71] G. Zubay, T. Mui, *Orig. Life Evol. Biosph.* **2001**, *31*, 87-102.

[72] B. E. Prieur, *C. R. Acad. Sci.* **2001**, *4*, 667-670.

[73] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196-196.

[74] 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.

[75] S. A. Benner, H.-J. Kim, M. A. Carrigan, *Acc. Chem. Res.* **2012**, *45*, 2025-2034.

[76] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner, A. Eschenmoser, G. Ohlofjgewidmet, *Helv. Chim. Acta* **1990**, *73*, 1410-1468.

[77] R. Krishnamurthy, S. Pitsch, G. Arrhenius, *Origins of life and evolution of the biosphere* **1999**, *29*, 139-152.

[78] J. Oró, *BBRC* **1960**, *2*, 407-412.

[79] A. W. Schwartz, M. Goverde, *J. Mol. Evol.* **1982**, *18*, 351-353.

[80] H. Wakamatsu, Y. Yamada, T. Saito, I. Kumashiro, T. Takenishi, *J. Org. Chem.* **1966**, *31*, 2035-2036.

[81] J. P. Ferris, L. E. Orgel, *J. Am. Chem. Soc.* **1966**, *88*, 1074-1074.

[82] R. A. Sanchez, J. P. Ferbis, L. E. Orgel, *J. Mol. Biol.* **1967**, *30*, 223-253.

[83] J. P. Ferris, L. E. Orgel, *J. Am. Chem. Soc.* **1966**, *88*, 3829-3831.

[84] J. P. Ferris, L. E. Orgel, *J. Am. Chem. Soc.* **1965**, *87*, 4976-4977.

[85] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 121-128.

[86] J. P. Ferris, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1968**, *33*, 693-704.

[87] M. P. Robertson, S. L. Miller, *Nature* **1995**, *375*, 772-774.

[88] H. Okamura, S. Becker, N. Tiede, S. Wiedemann, J. Feldmann, T. Carell, *Chem. Commun.* **2019**, *55*, 1939-1942.

[89] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25-33.

[90] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249-257.

[91] K. N. Drew, J. Zajicek, G. Bondo, B. Bose, A. S. Serianni, *Carbohydr. Res.* **1998**, *307*, 199-209.

[92] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239-242.

[93] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531-543.

[94] A.-A. Ingar, R. W. A. Luke, B. R. Hayter, J. D. Sutherland, *ChemBioChem* **2003**, *4*, 504-507.

[95] S. Islam, D.-K. Bučar, M. W. Powner, *Nat. Chem.* **2017**, *9*, 584-589.

[96] S. Stairs, A. Nikmal, D.-K. Bučar, S.-L. Zheng, J. W. Szostak, M. W. Powner, *Nat. Commun.* **2017**, *8*, 15270.

[97] J. Xu, V. Chmela, Nicholas J. Green, David A. Russell, Mikołaj J. Janicki, Robert W. Góra, R. Szabla, Andrew D. Bond, John D. Sutherland, *Nature* **2020**, *582*, 60-66.

[98] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833-836.

[99] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76-82.

[100] U. P. Trinks, Zur Chemie der Aminopyrimidine, ETH Zürich **1987**.

[101] K. E. Koch, Über strukturelle Zusammenhänge zwischen Blausäure und natürlichen Purinen und Pteridinen, ETH Zürich **1992**.

[102] E. Maciá, M. V. Hernández, J. Oró, *Orig. Life Evol. Biosph.* **1997**, *27*, 459-480.

[103] A. Gulick, *Am. Sci.* **1955**, *43*, 479-489.

[104] R. Morchio, S. Traverso, *Riv. Biol.* **2005**, *98*, 18-23.

[105] A. Gulick, *Ann. N. Y. Acad. Sci.* **1957**, *69*, 309-313.

[106] R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 64-66.

[107] J. P. Ferris, *Science* **1968**, *161*, 53-54.

[108] M. Halmann, R. A. Sanchez, L. E. Orgel, *J. Org. Chem.* **1969**, *34*, 3702-3703.

[109] C. Degani, M. Halmann, *J. Chem. Soc. C* **1971**, 1459-1465.

[110] M. A. Pasek, *Chem. Rev.* **2020**, *120*, 4690-4706.

[111] A. M. Schoffstall, E. M. Laing, *Orig. Life Evol. Biosph.* **1985**, *15*, 141-150.

[112] A. M. Schoffstall, *Orig. Life Evol. Biosph.* **1976**, *7*, 399-412.

[113] C. Ponnamperuma, R. Mack, *Science* **1965**, *148*, 1221-1223.

[114] M. Takabayashi, Y. Hirakawa, T. Kakegawa, Y. Furukawa, *Geochem. J.* **2023**, *57*, 134-142.

[115] M. A. Pasek, D. S. Lauretta, *Astrobiology* **2005**, *5*, 515-535.

[116] D. E. Bryant, T. P. Kee, *Chem. Commun.* **2006**, 2344-2346.

[117] M. A. Pasek, T. P. Kee, D. E. Bryant, A. A. Pavlov, J. I. Lunine, *Angew. Chem., Int. Ed.* **2008**, *47*, 7918-7920.

[118] M. Gull, M. A. Mojica, F. M. Fernández, D. A. Gaul, T. M. Orlando, C. L. Liotta, M. A. Pasek, *Sci. Rep.* **2015**, *5*, 17198.

[119] R. Osterberg, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 241-248.

[120] J. Rabinowitz, S. Chang, C. Ponnamperuma, *Nature* **1968**, *218*, 442-443.

[121] Y. Yamagata, H. Watanabe, M. Saitoh, T. Namba, *Nature* **1991**, *352*, 516-519.

[122] A. W. Schwartz, *J. Chem. Soc. D* **1969**, 1393a-1393a.

[123] C. Cheng, C. Fan, R. Wan, C. Tong, Z. Miao, J. Chen, Y. Zhao, *Orig. Life Evol. Biosph.* **2002**, *32*, 219-224.

[124] K. Gao, L. E. Orgel, *Orig. Life Evol. Biosph.* **2000**, *30*, 45-51.

[125] Y. Yamagata, *Orig. Life Evol. Biosph.* **1999**, *29*, 511-520.

[126] M. Karki, C. Gibard, S. Bhowmik, R. Krishnamurthy, *Life* **2017**, *7*.

[127] C. Gibard, S. Bhowmik, M. Karki, E. K. Kim, R. Krishnamurthy, *Nat. Chem.* **2018**, *10*, 212-217.

[128] C. Gibard, I. B. Gorrell, E. I. Jiménez, T. P. Kee, M. A. Pasek, R. Krishnamurthy, *Angew. Chem., Int. Ed.* **2019**, *58*, 8151-8155.

[129] R. Krishnamurthy, S. Guntha, A. Eschenmoser, *Angew. Chem., Int. Ed.* **2000**, *39*, 2281-2285.

[130] R. Krishnamurthy, G. Arrhenius, A. Eschenmoser, *Orig. Life Evol. Biosph.* **1999**, *29*, 333-354.

[131] K. Ashe, C. Fernández-García, M. K. Corpinot, A. J. Coggins, D.-K. Bučar, M. W. Powner, *Commun. Chem.* **2019**, *2*, 23.

[132] J. Morávek, *Tetrahedron Lett.* **1967**, *8*, 1707-1710.

[133] J. D. Ibanez, A. P. Kimball, J. Oró, *Science* **1971**, *173*, 444-446.

[134] J. P. Ferris, G. Ertem, V. Agarwal, *Orig. Life Evol. Biosph.* **1989**, *19*, 165-178.

[135] S. A. Rudolph, E. M. Johnson, P. Greengard, *J. Biol. Chem.* **1971**, *246*, 1271-1273.

[136] M. S. Verlander, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1973**, *2*, 303-316.

[137] M. S. Verlander, L. E. Orgel, *J. Mol. Evol.* **1974**, *3*, 115-120.

[138] G. Costanzo, S. Pino, F. Ciciriello, E. Di Mauro, *Journal of Biological Chemistry* **2009**, *284*, 33206-33216.

[139] M. Morasch, C. B. Mast, J. K. Langer, P. Schilcher, D. Braun, *ChemBioChem* **2014**, *15*, 879-883.

[140] R. Lohrmann, *J. Mol. Evol.* **1977**, *10*, 137-154.

[141] A. Mariani, D. A. Russell, T. Javelle, J. D. Sutherland, *J. Am. Chem. Soc.* **2018**, *140*, 8657-8661.

[142] R. Yi, Y. Hongo, A. C. Fahrenbach, *Chem. Commun.* **2018**, *54*, 511-514.

[143] H. Sawai, L. E. Orgel, *J. Am. Chem. Soc.* **1975**, *97*, 3532-3533.

[144] H. Sawai, *J. Am. Chem. Soc.* **1976**, *98*, 7037-7039.

[145] H. Sawai, K. Kuroda, T. Hojo, *Bull. Chem. Soc. Jpn.* **1989**, *62*, 2018-2023.

[146] H. Sawai, K. Higa, K. Kuroda, *J. Chem. Soc., Perkin Trans.* **1992**, 505-508.

[147] H. Sawai, K. Yamamoto, *Bull. Chem. Soc. Jpn.* **1996**, *69*, 1701-1704.

[148] J. P. Ferris, G. Ertem, *J. Am. Chem. Soc.* **1993**, *115*, 12270-12275.

[149] J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59-61.

[150] P. Z. Ding, K. Kawamura, J. P. Ferris, *Orig. Life Evol. Biosph.* **1996**, *26*, 151-171.

[151] K. Kawamura, J. P. Ferris, *Orig. Life Evol. Biosph.* **1999**, *29*, 563-591.

[152] S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti, D. Deamer, *Orig. Life Evol. Biosph.* **2008**, *38*, 57-74.

[153] V. DeGuzman, W. Vercoutere, H. Shenasa, D. Deamer, *J. Mol. Evol.* **2014**, *78*, 251-262.

[154] D. K. O'Flaherty, N. P. Kamat, F. N. Mirza, L. Li, N. Prywes, J. W. Szostak, *J. Am. Chem. Soc.* **2018**, *140*, 5171-5178.

[155] C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8742-8747.

[156] H. F. Noller, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a003681.

[157] M. Kimura, S. Akanuma, *J. Mol. Evol.* **2020**, *88*, 372-381.

[158] D. A. M. Zaia, C. T. B. V. Zaia, H. De Santana, *Orig. Life Evol. Biosph.* **2008**, *38*, 469-488.

[159] A. D. Solis, *BMC Evol. Biol.* **2019**, *19*, 158.

[160] J. E. Jackman, J. D. Alfonzo, *Wiley Interdiscip. Rev. RNA* **2013**, *4*, 35-48.

[161] C. S. Chow, T. N. Lamichhane, S. K. Mahto, *ACS Chem. Biol.* **2007**, *2*, 610-619.

[162] J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, *33*, 7869-7876.

[163] 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, X. Cao, P. A. Limbach, U. L. RajBhandary, V. de Crécy-Lagard, *Mol. Microbiol.* **2015**, *98*, 1199-1221.

[164] J. W. Stuart, Z. Gdaniec, R. Guenther, M. Marszalek, E. Sochacka, A. Malkiewicz, P. F. Agris, *Biochemistry* **2000**, *39*, 13396-13404.

[165] F. V. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186-1191.

[166] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem., Int. Ed.* **2018**, *57*, 5943-5946.

[167] P. Schimmel, *Annu. Rev. Biochem.* **1987**, *56*, 125-158.

[168] F. Schuber, M. Pinck, *Biochimie* **1974**, *56*, 383-390.

[169] S. E. Walker, K. Fredrick, *Methods* **2008**, *44*, 81-86.

[170] M. Di Giulio, *J. Mol. Evol.* **1997**, *45*, 571-578.

[171] P. Canavelli, S. Islam, M. W. Powner, *Nature* **2019**, *571*, 546-549.

[172] B. Jash, P. Tremmel, D. Jovanovic, C. Richert, *Nat. Chem.* **2021**, *13*, 751-757.

[173] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4398-4402.

[174] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Chem. Biol.* **2013**, *20*, 466-474.

[175] A. E. Engelhart, N. V. Hud, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002196.

[176] N. V. Hud, *Nat. Commun.* **2018**, *9*, 5171.

[177] B. J. Cafferty, D. M. Fialho, N. V. Hud, in *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids*, Springer Cham, **2018**, pp. 143-174.

[178] L. E. Orgel, *Biochem. Sci.* **1998**, *23*, 491-495.

[179] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174-4175.

[180] L. Orgel, *Science* **2000**, *290*, 1306-1307.

[181] K. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347-1351.

[182] A. Eschenmoser, *Orig. Life Evol. Biosph.* **2004**, *34*, 277-306.

[183] B. W. F. Colville, M. W. Powner, *Angew. Chem., Int. Ed.* **2021**, *60*, 10526-10530.

[184] D. Whitaker, M. W. Powner, *Nat. Chem.* **2022**, *14*, 766-774.

[185] S. J. Roberts, R. Szabla, Z. R. Todd, S. Stairs, D.-K. Bučar, J. Šponer, D. D. Sasselov, M. W. Powner, *Nat. Commun.* **2018**, *9*, 4073.

[186] N. Martín-Pintado, M. Yahyaee-Anzahaee, R. Campos-Olivas, A. M. Noronha, C. J. Wilds, M. J. Damha, C. González, *Nucleic Acids Res.* **2012**, *40*, 9329-9339.

[187] A. M. Noronha, C. J. Wilds, C.-N. Lok, K. Viazovkina, D. Arion, M. A. Parniak, M. J. Damha, *Biochemistry* **2000**, *39*, 7050-7062.

[188] A. Eschenmoser, *Science* **1999**, *284*, 2118.

[189] M. Bolli, R. Micura, A. Eschenmoser, *Chem. Biol.* **1997**, *4*, 309-320.

[190] N. V. Hud, F. A. L. Anet, *J. Theor. Biol.* **2000**, *205*, 543-562.

[191] J. S. Rice, X. Gao, *Biochemistry* **1997**, *36*, 399-411.

[192] R. J. Jones, K. Y. Lin, J. F. Milligan, S. Wadwani, M. D. Matteucci, *J. Org. Chem.* **1993**, *58*, 2983-2991.

[193] H. D. Bean, F. A. L. Anet, I. R. Gould, N. V. Hud, *Orig. Life Evol. Biosph.* **2006**, *36*, 39-63.

[194] A. L. Weber, *Orig. Life Evol. Biosph.* **2001**, *31*, 71-86.

[195] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25-33.

[196] F. B. Howard, H. T. Miles, *Biochemistry* **1977**, *16*, 4647-4650.

[197] K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1998**, *81*, 375-474.

[198] R. Krishnamurthy, S. Pitsch, M. Minton, C. Miculka, N. Windhab, A. Eschenmoser, *Angew. Chem., Int. Ed.* **1996**, *35*, 1537-1541.

[199] B. D. Heuberger, C. Switzer, *ChemBioChem* **2008**, *9*, 2779-2783.

[200] V. M. Kolb, J. P. Dworkin, S. L. Miller, *J. Mol. Evol.* **1994**, *38*, 549-557.

[201] B. J. Cafferty, I. Gállego, M. C. Chen, K. I. Farley, R. Eritja, N. V. Hud, *J. Am. Chem. Soc.* **2013**, *135*, 2447-2450.

[202] H. Okamura, A. Crisp, S. Hübner, S. Becker, P. Rovó, T. Carell, *Angew. Chem., Int. Ed.* **2019**, *58*, 18691-18696.

[203] L. T. Burgdorf, T. Carell, *Chem. Eur. J.* **2002**, *8*, 293-301.

[204] T. H. Gehrke, U. Lischke, K. L. Gasteiger, S. Schneider, S. Arnold, H. C. Müller, D. S. Stephenson, H. Zipse, T. Carell, *Nat. Chem. Biol.* **2013**, *9*, 455-461.

[205] H. S. Forrest, R. Hull, H. J. Rodda, A. R. Todd, *J. Chem. Soc.* **1951**, 3-7.

[206] C. Temple, Jr., B. H. Smith, J. A. Montgomery, *J. Org. Chem.* **1975**, *40*, 3141-3142.

[207] W. R. Boon, W. G. M. Jones, G. R. Ramage, *J. Chem. Soc.* **1951**, 96-102.

[208] P. G. Baraldi, A. U. Broceta, M. J. P. d. l. Infantas, J. J. D. a. Mochun, A. Espinosa, R. Romagnoli, *Tetrahedron* **2002**, *58*, 7607-7611.

[209] F. Xu, A. Crisp, T. Schinkel, R. C. A. Dubini, S. Hübner, S. Becker, F. Schelter, P. Rovó, T. Carell, *Angew. Chem., Int. Ed.* **2022**, *61*, e202211945.

[210] F. Xu, S. Wiedemann, J. Feldmann, S. Becker, T. Carell, *Chem.Eur.* **2023**, *1*, e202300057.

[211] R. J. Moss, C. R. Petrie, R. B. Meyer, Jr., L. D. Nord, R. C. Willis, R. A. Smith, S. B. Larson, G. D. Kini, R. K. Robins, *J. Med. Chem.* **1988**, *31*, 786-790.

[212] R. Y. Baiazitov, N. Sydorenko, H. Ren, Y.-C. Moon, *J. Org. Chem.* **2017**, *82*, 5881-5889.

[213] A. V. Yakhnin, *Orig. Life Evol. Biosph.* **2013**, *43*, 39-47.

[214] M. Calvin, *Proc. Math. Phys. Eng. Sci.* **1997**, *288*, 441-466.

[215] J. Hu, W. Lei, J. Wang, H.-Y. Chen, J.-J. Xu, *Chem. Commun.* **2019**, *55*, 310-313.

[216] G. Banfalvi, *Int. J. Mol. Sci.* **2021**, *22*, 3857.

[217] H. S. Bernhardt, R. K. Sandwick, *J. Mol. Evol.* **2014**, *79*, 91-104.

[218] G. Banfalvi, *DNA Cell Biol.* **2019**, *38*, 223-228.

[219] J. N. Singer, F. M. Müller, E. Węgrzyn, C. Hölzl, H. Hurmiz, C. Liu, L. Escobar, T. Carell, *Angew. Chem., Int. Ed.* **2023**, *62*, e202302360.

[220] A. Crisp, Non-canonical nucleosides and proto-urea-RNA at the chemical origins of life, LMU München **2021**.

[221] P. A. Küpfer, C. J. Leumann, *Nucleic Acids Res.* **2007**, *35*, 58-68.

Anhang I

nature portfolio

https://doi.org/10.1038/s41586-022-04676-3

Supplementary information

A prebiotically plausible scenario of an RNA–peptide world

In the format provided by the authors and unedited

A prebiotically plausible scenario of an RNA-peptide world

Felix Müller[1]*, Luis Escobar[1]*, Felix Xu[1], Ewa Węgrzyn[1], Milda Nainytė[1], Tynchtyk Amatov[1], Chun-Yin Chan[1], Alexander Pichler[1] and Thomas Carell[1]#

[1] Department of Chemistry, Ludwig-Maximilians-Universität (LMU) München, Butenandtstrasse 5-13, 81377 München, Germany.

* These authors contributed equally.

E-Mail: thomas.carell@lmu.de

Supplementary Information

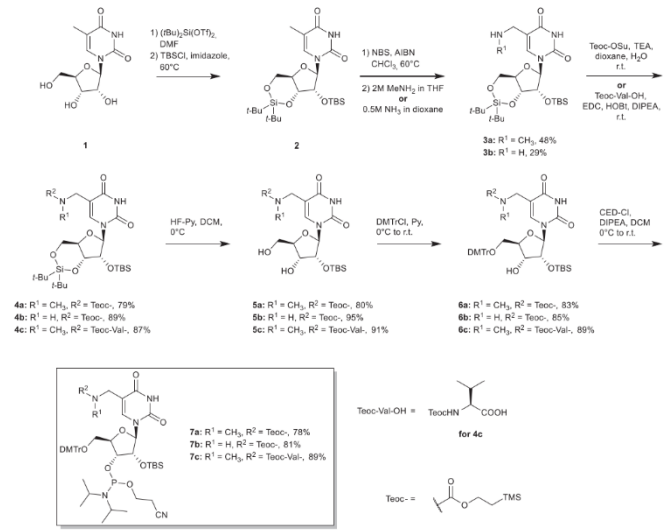
Table of Contents

1. General information and instruments for phosphoramidites, amino acids and peptides

Reagents were purchased from commercial suppliers and used without further purification unless otherwise stated. All anhydrous solvents stored under inert atmosphere were also purchased. All reactions involving air/moisture sensitive reagents/intermediates were performed under inert atmosphere using oven-dried glassware. Routine ^{1}H NMR, $^{13}C\{^1H\}$ NMR and $^{31}P\{^1H\}$ NMR were recorded on a Bruker Ascend 400 spectrometer (400 MHz for ^{1}H NMR, 100 MHz for ^{13}C NMR and 162 MHz for ^{31}P NMR) or a Bruker ARX 600 spectrometer (600 MHz for ^{1}H NMR, 150 MHz for ^{13}C NMR and 243 MHz for ^{31}P NMR). Deuterated solvents used are indicated in the characterization and chemical shifts (δ) are reported in ppm. Residual solvent peaks were used as reference.[1] All NMR *J* values are given in Hz. COSY, HMQC and HMBC experiments were recorded to help with the assignment of ^{1}H and ^{13}C signals. NMR spectra were analyzed using MestReNova software version 10.0. High Resolution Mass Spectra (HRMS) were measured on a Thermo Finnigan LTQ-FT with ESI as ionization mode. IR spectra were recorded on a Perkin-Elmer Spectrum BX II FT-IR instrument equipped with an ATR accessory. Column chromatography was performed with silica gel technical grade (Macherey-Nagel), 40-63 µm particle size. Reaction progress was monitored by Thin Layer Chromatography (TLC) analysis on silica gel 60 F254 and stained with *para*-anisaldehyde, potassium permanganate or cerium ammonium molybdate solution.

2. Synthesis and characterization data

2.1 Nucleobase-modified 5-methyluridine phosphoramidites



Scheme S1. Synthesis of nucleobase-modified 5-methyluridine phosphoramidites.

General procedure for the synthesis of 3a,b:

Silyl-protected 5-methyluridine **2** was synthesized starting from 5-methyluridine **1** following a procedure previously described in literature.[2] A solution of **2** (1.0 equiv.) in dry $CHCl_3$ was heated at 60°C. *N*-bromosuccinimide (NBS) (1.2 equiv., previously purified by recrystallization) and azobisisobutyronitrile (AIBN) (0.12 equiv.) were added and the reaction was stirred under reflux for 1.5 h. After that, the reaction mixture was cooled to r.t. and either $MeNH_2$ (2 M in THF, 5.0 equiv.) for **3a** or NH_3 (0.5 M in 1,4-dioxane, 5.0 equiv.) for **3b** were added. The resulting suspension was stirred for 2 h at r.t. and, subsequently, it was diluted with aq. sat. $NaHCO_3$ solution. The crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated. The crude was purified by silica gel column chromatography to furnish **3a**,**b** as a yellow foam.

3a: Yield: 48%; R_f = 0.11 (9:1 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2931 (w), 2858 (w), 2359 (w), 1682 (s), 1462 (m), 1386 (w), 1254 (m), 1202 (w), 1167 (w), 1115 (m), 1057 (s), 1000 (m), 938 (w), 882 (m), 827 (s), 778 (s), 754 (m), 685 (w); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 7.35 (s, 1H), 5.66 (s, 1H), 4.47 (dd, J = 9.5, 4.7 Hz, 1H), 4.28 (d, J = 4.7 Hz, 1H), 4.18-4.06 (m, 1H), 4.05-3.97 (m, 1H), 3.92 (dd, J = 9.5, 4.7 Hz, 1H), 3.58-3.47 (m, 2H), 2.41 (s, 3H), 1.03 (s, 9H), 1.01 (s, 9H), 0.91 (s, 9H), 0.15 (s, 3H), 0.12 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 164.2, 150.1, 138.1, 111.1, 94.1, 76.1, 75.3, 74.6, 67.7, 47.6, 35.0, 27.6, 27.1, 26.0, 22.9, 20.5, 18.4, -4.2, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{25}H_{48}N_3O_6Si_2$ 542.3076; Found 542.3076.

3b: Yield: 29%; R_f = 0.25 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3052 (w), 2934 (w), 2858 (w), 2363 (w), 1687 (m), 1471 (w), 1422 (w), 1388 (w), 1264 (s), 1204 (w), 1168 (w), 1115 (m), 1059 (m), 999 (m), 938 (w), 896 (w), 882 (m), 828 (s), 780 (m), 731 (s), 702 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 7.29 (s, 1H), 5.69 (s, 1H), 4.49 (dd, J = 9.1, 5.0 Hz, 1H), 4.28 (d, J = 4.8 Hz, 1H), 4.18-4.08 (m, 1H), 4.01 (dd, J = 10.6, 9.1 Hz, 1H), 3.92 (dd, J = 9.5, 4.8 Hz, 1H), 3.60 (s, 2H), 1.05 (s, 9H), 1.02 (s, 9H), 0.92 (s, 9H), 0.16 (s, 3H), 0.13 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 163.2, 149.6, 136.3, 94.1, 76.2, 75.4, 74.6, 67.7, 39.2, 27.6, 27.1, 26.0, 22.9, 20.5, 18.4, -4.2, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{24}H_{48}N_3O_6Si_2$ 528.2920; Found 528.2921.

General procedures for the synthesis of 4a-c:

Procedure A (for compounds 4a,b): To a solution of **3a**,**b** (1.0 equiv.) in 1,4-dioxane and H_2O (1:1 v/v) were added teoc-OSu (1.1 equiv.) and triethylamine (TEA) (1.5 equiv.). The mixture was stirred at r.t. for 16 h. After that, the crude was diluted with water and extracted three times with Et_2O. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated. The obtained residue was purified by silica gel column chromatography to yield the teoc-protected compound **4a,b** as a white solid.

Procedure B (for compound 4c): Teoc-protected valine was synthesized following a previously reported procedure in literature.[3] Teoc-Val-OH (1.2 equiv.) was dissolved in dry DCM and DMF (99:1 v/v). To the solution, 1-hydroxybenzotriazole hydrate (HOBt•H_2O) (1.2 equiv.), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC•HCl) (1.2 equiv.) and *N*,*N*-diisopropylethylamine (DIPEA) (1.2 equiv.) were added. After stirring at r.t. for 30 min, a solution of **3a** (1.0 equiv.) in DCM was added and the reaction was stirred for 24 h. The reaction mixture was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated. Purification by silica gel column chromatography furnished the amino acid conjugate **4c** as a white foam.

4a: Yield: 79%; R_f = 0.34 (4:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3054 (w), 2956 (w), 2359 (w), 1692 (m), 1463 (w), 1422 (w), 1264 (s), 1214 (w), 1167 (w), 1146 (w), 1059 (w), 1000 (w), 938 (w), 895 (m), 838 (m), 730 (s), 702 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 9.14 (s, 1H), 7.54 (s, 1H), 5.65 (s, 1H), 4.48 (dd, J = 9.2, 4.2 Hz, 1H), 4.28 (d, J = 4.2 Hz, 1H), 4.23-3.98 (m, 6H), 3.91 (dd, J = 9.2, 4.2 Hz, 1H), 2.96 (s, 3H), 1.05 (s, 9H), 1.03-0.96 (m, 11H), 0.93 (s, 9H), 0.18 (s, 3H), 0.13 (s, 3H), 0.04 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 163.6, 157.1, 149.7, 139.4, 110.8, 93.8, 76.0, 75.5, 74.9, 67.6, 63.8, 45.0, 35.6, 27.7, 27.1, 26.0, 22.8, 20.5, 18.4, 17.9, -1.3, -4.2, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{31}H_{66}N_3O_8Si_3$ 686.3683; Found 686.3683.

4b: Yield: 89%; R_f = 0.23 (4:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2937 (w), 2359 (w), 2167 (w), 1690 (m), 1470 (w), 1251 (m), 1213 (w), 1127 (w), 1061 (m), 999 (m), 831 (m), 779 (m), 730 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) 8.13 (s, 1H), 7.46 (s, 1H), 5.65 (s, 1H), 5.23 (t, J = 5.9 Hz, 1H), 4.50 (dd, J = 9.0, 4.9 Hz, 1H), 4.28 (d, J = 4.6 Hz, 1H), 4.20-4.05 (m, 4H), 3.98 (d, J = 6.3 Hz, 2H), 3.90 (dd, J = 9.5, 4.6 Hz, 1H), 1.06 (s, 9H), 1.02 (s, 9H), 0.99-0.88 (m, 11H), 0.18 (s, 3H), 0.14 (s, 3H), 0.03 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 162.9, 156.9, 149.4, 138.4, 111.4, 93.9, 76.0, 75.5, 74.9, 67.6, 63.4, 37.7, 27.7, 27.1, 26.0, 22.9, 20.5, 18.4, 17.8, -1.3, -4.1, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{30}H_{58}N_3O_8Si_3$ 672.3526; Found 672.3535.

4c: Yield: 87%; R_f = 0.29 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3053 (w), 2956 (w), 2859 (w), 2359 (w), 1689 (m), 1648 (w), 1586 (w), 1536 (w), 1471 (m), 1382 (w), 1366 (m), 1311 (w), 1264 (s), 1168 (w), 1114 (m), 1059 (m), 1002 (w), 938 (w), 835 (m), 732 (s), 702 (s); For major rotamer: ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 9.22 (s, 1H), 7.66 (s, 1H), 5.69 (s, 1H), 5.40 (d, J = 9.0 Hz, 1H), 4.51-4.42 (m, 2H), 4.25 (d, J = 14.3 Hz, 1H), 4.21-4.01 (m, 6H), 3.97 (dd, J = 9.0, 4.8 Hz, 1H), 3.21 (s, 3H), 1.92-1.87 (m, 1H), 1.09 (s, 9H), 1.05-0.98 (m, 11H), 0.94-0.89 (m, 12H), 0.80 (d, J = 6.7 Hz, 3H), 0.14 (s, 3H), 0.11 (s, 3H), 0.02 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.7, 163.6, 157.0, 149.6, 141.3, 110.1, 93.9, 76.1, 75.6, 74.8, 67.6, 63.4, 55.4, 44.5, 37.3, 31.3, 27.7, 27.1, 26.0, 22.8, 20.5, 19.6, 18.4, 17.8, 17.1, -1.3, -4.2, -5.0; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{36}H_{69}N_4O_9Si_3$ 785.4367; Found 785.4363.

General procedure for the synthesis of 5a-c:

The modified 5-methyluridine **4a-c** (1.0 equiv.) was dissolved in DCM/pyridine (9:1 v/v) and cooled to 0°C in a plastic reaction vessel. Subsequently, a solution of 70% HF-pyridine (5.0 equiv.) was slowly added, and the reaction mixture was stirred at 0°C for 2 h. The reaction was quenched by adding aq. sat. $NaHCO_3$ and the crude was extracted three times with DCM. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated. The crude product was purified by silica gel column chromatography to afford the diol compound **5a-c** as a white foam.

5a: Yield: 80%; R_f = 0.42 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3417 (w), 3060 (w), 2949 (w), 2856 (w), 2359 (w), 1673 (s), 1462 (m), 1401 (w), 1362 (w), 1250 (m), 1214 (w), 1144 (m), 1088 (m), 1060 (m), 1005 (w), 938 (w), 833 (s), 777 (s), 693 (w); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 9.47 (s, 1H), 8.19 (s, 1H), 5.87 (d, J = 5.2 Hz, 1H), 4.48 (t, J = 5.1 Hz, 1H), 4.30-3.85 (m, 7H), 3.83-3.74 (m, 1H), 2.97 (s, 3H), 2.78 (br s, 1H), 1.03-0.92 (m, 2H), 0.88 (s, 9H), 0.06 (s, 6H), 0.02 (s, 9H) (some proton signals appeared too broad for an unequivocal assignment); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 163.8, 157.4, 150.5, 141.9, 111.4, 90.2, 85.8, 75.3, 71.3, 64.1, 62.2, 44.5, 35.6, 25.8, 18.1, -1.4, -4.7 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{23}H_{44}N_3O_8Si_2$ 546.2661; Found 546.2666.

5b: Yield: 95%; R_f = 0.23 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3386 (w), 2950 (w), 2854 (w), 2362 (w), 1674 (s), 1524 (m), 1470 (m), 1390 (w), 1333 (w), 1248 (s), 1179 (w), 1115 (m), 1086 (w), 1060 (s), 1001 (w), 938 (w), 902 (w), 857 (m), 833 (s), 779 (s), 694 (w); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.51 (s, 1H), 8.09 (s, 1H), 5.80 (s, 1H), 5.35 (t, J = 6.2 Hz, 1H), 4.49 (t, J = 4.8 Hz, 1H), 4.33-4.22 (m, 1H), 4.17-4.05 (m, 4H), 4.02-3.92 (m, 3H), 3.81 (dd, J = 12.0, 5.3 Hz, 1H), 3.57 (t, J = 5.3 Hz, 1H), 2.70 (d, J = 4.3 Hz, 1H), 0.99-0.92 (m, 3H), 0.90 (s, 9H), 0.11-0.08 (m, 6H), 0.02 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 163.0, 157.2, 150.2, 141.0, 111.8, 90.9, 85.7, 75.1, 71.0, 63.6, 62.1, 37.2, 25.8, 18.1, 17.8, -1.3, -4.6, -5.0; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{22}H_{42}N_3O_8Si_2$ 532.2505; Found 532.2509.

5c: Yield: 91%; R_f = 0.18 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3440 (w), 3054 (w), 2953 (w), 2857 (w), 2359 (w), 1677 (s), 1463 (m), 1401 (w), 1362 (w), 1264 (s), 1250 (m), 1215 (w), 1137 (m), 1112 (w), 1089 (w), 1060 (w), 1005 (w), 937 (w), 836 (s), 779 (m), 733 (s), 701 (s); For major rotamer: ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.80 (s, 1H), 8.06 (s, 1H), 5.94 (d, J = 5.0 Hz, 1H), 5.45 (d, J = 9.9 Hz, 1H), 4.59 (d, J = 15.0 Hz, 1H), 4.49 (dd, J = 9.9, 5.4 Hz, 1H), 4.43-4.05 (m, 5H), 3.91 (d, J = 15.0 Hz, 2H), 3.80 (d, J = 12.0 Hz, 1H), 3.19 (s, 3H), 2.71 (d, J = 3.1 Hz, 1H), 2.01-1.94 (m, 1H), 1.01-0.93 (m, 5H), 0.93-0.85 (m, 13H), 0.07 (s, 3H), 0.06 (s, 3H), 0.03 (s, 9H) (some proton signals appeared too broad for an unequivocal assignment); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.9, 163.2, 157.5, 150.2, 139.8, 110.0, 89.5, 85.7, 75.9, 71.3, 63.9, 61.8, 55.8, 44.5, 36.7, 31.0, 25.8, 19.7, 18.1, 17.9, 17.1, -1.4, -4.7, -5.1 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{28}H_{53}N_4O_9Si_2$ 645.3346; Found 645.3349.

General procedure for the synthesis of 6a-c:

To a solution of the 3',5'-deprotected 5-methyluridine derivative **5a-c** (1.0 equiv.) in pyridine was added 4,4'-dimethoxytrityl chloride (DMTrCl) (1.5 equiv.). After stirring at r.t. for 16 h, the reaction mixture was concentrated and purified by silica gel column chromatography with an addition of 0.1% of pyridine to the eluent to afford the DMTr-protected compound **6a-c** as a white foam.

6a: Yield: 83%; R_f = 0.57 (1:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3444 (w), 3055 (w), 2953 (w), 2857 (w), 2359 (w), 1678 (s), 1608 (w), 1583 (w), 1508 (m), 1463 (m), 1401 (w), 1342 (w), 1297 (w), 1264 (m), 1248 (s), 1175 (m), 1150 (m), 1113 (w), 1089 (w), 1034 (m), 1006 (w), 938 (w), 910 (w), 830 (s), 780 (m), 733 (s), 701 (s); For major rotamer: ^{1}H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.21 (s, 1H), 7.76 (s, 1H), 7.57-7.46 (m, 2H), 7.45-7.37 (m, 4H), 7.37-7.29 (m, 2H), 7.28-7.19 (m, 1H), 6.90 (d, J = 8.9 Hz, 4H), 5.94 (s, 1H), 4.44 (br s, 1H), 4.22-4.03 (m, 3H), 3.84-3.71 (m, 8H), 3.44 (br s, 2H), 2.90 (br s, 3H), 1.08-0.81 (m, 11H), 0.15 (s, 6H), 0.03 (s, 9H) (some proton signals appeared too broad for an unequivocal assignment); ^{13}C{^{1}H} NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 163.8, 159.6, 156.7, 151.2, 146.1, 136.7, 131.1, 131.0, 129.0, 128.7, 114.0, 114.0, 113.6, 111.4, 89.7, 87.3, 84.3, 76.5, 71.5, 64.6, 63.6, 55.5, 46.5, 35.7, 26.2, 18.7, 18.3, -1.4, -4.6, -4.6 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: [M-H]$^-$ Calcd. for $C_{44}H_{66}N_3O_{10}Si_2$ 846.3823; Found 846.3825.

6b: Yield: 85%; R_f = 0.33 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3342 (w), 2950 (w), 2855 (w), 2358 (w), 1708 (s), 1607 (w), 1582 (w), 1508 (m), 1462 (m), 1390 (w), 1248 (s), 1175 (m), 1116 (m), 1089 (w), 1063 (m), 1035 (m), 969 (w), 937 (w), 859 (m), 835 (s), 780 (s), 754 (m), 726 (w), 699 (m); ^{1}H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.18 (s, 1H), 7.78 (s, 1H), 7.52 (d, J = 7.6 Hz, 2H), 7.41 (d, J = 8.8 Hz, 4H), 7.34 (t, J = 7.6 Hz, 2H), 7.24 (t, J = 7.6 Hz, 1H), 6.92 (d, J = 8.8 Hz, 4H), 6.05 (t, J = 5.0 Hz, 1H), 5.95 (d, J = 4.6 Hz, 1H), 4.46 (t, J = 5.0 Hz, 1H), 4.28-4.22 (m, 1H), 4.17-4.12 (m, 1H), 4.11-4.01 (m, 2H), 3.82 (d, J = 5.8 Hz, 1H), 3.79 (s, 6H), 3.64 (dd, J = 14.5, 5.5 Hz, 1H), 3.56 (dd, J = 14.5, 5.9 Hz, 1H), 3.45 (dd, J = 10.8, 4.2 Hz, 1H), 3.39 (dd, J = 10.8, 2.5 Hz, 1H),

0.97-0.83 (m, 11H), 0.15 (s, 3H), 0.14 (s, 3H), 0.02 (s, 9H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 163.5, 159.6, 157.0, 151.2, 146.0, 138.9, 136.7, 136.6, 131.0, 129.0, 128.8, 127.6, 114.0, 112.2, 89.6, 87.4, 84.3, 76.7, 71.6, 64.4, 62.8, 55.5, 38.5, 26.2, 18.7, 18.4, -1.4, -4.6, -4.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{43}H_{60}N_3O_{10}Si_2$ 834.3812; Found 834.3801.

6c: Yield: 89%; R_f = 0.42 (1:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3054 (w), 2954 (w), 2930 (w), 2857 (w), 2359 (w), 1687 (s), 1644 (w), 1608 (w), 1508 (m), 1463 (m), 1389 (w), 1263 (m), 1249 (s), 1175 (m), 1115 (w), 1083 (w), 1061 (w), 1035 (m), 967 (w), 935 (w), 914 (w), 858 (m), 833 (s), 780 (w), 733 (s), 700 (s); For major rotamer: ^1H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 9.10 (s, 1H), 7.75 (s, 1H), 7.45 (d, *J* = 7.3 Hz, 2H), 7.35 (d, *J* = 8.8 Hz, 4H), 7.31-7.24 (m, 2H), 7.20 (t, *J* = 7.3 Hz, 1H), 6.82 (d, *J* = 8.8 Hz, 4H), 5.88 (d, *J* = 3.8 Hz, 1H), 5.41 (d, *J* = 8.9 Hz, 1H), 4.43-4.37 (m, 1H), 4.34-4.29 (m, 1H), 4.18-4.07 (m, 4H), 4.00 (d, *J* = 14.3 Hz, 1H), 3.77 (s, 6H), 3.54-3.40 (m, 3H), 3.11 (s, 3H), 2.59 (d, *J* = 6.3 Hz, 1H), 1.92-1.78 (m, 1H), 1.04-0.96 (m, 2H), 0.94-0.86 (m, 12H), 0.79 (d, *J* = 6.7 Hz, 3H), 0.12 (s, 3H), 0.11 (s, 3H), 0.03 (s, 9H) (some proton signals appeared too broad for an unequivocal assignment); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.0, 163.2, 158.7, 156.9, 150.0, 149.8, 144.8, 141.2, 136.2, 135.8, 130.4, 130.3, 128.3, 128.0, 127.0, 123.9, 113.3, 110.3, 89.9, 86.8, 83.7, 75.6, 70.8, 63.6, 63.3, 55.3, 45.5, 37.1, 31.4, 25.8, 19.6, 18.1, 17.9, 17.3, -1.4, -4.5, -5.1 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{49}H_{69}N_4O_{11}Si_2$ 945.4507; Found 945.4508.

General procedure for the synthesis of phosphoramidites 7a-c:

A solution of 5'-DMTr-protected compound **6a-c** (1.0 equiv.) and DIPEA (4.0 equiv.) in dry DCM was cooled to 0°C. To this solution was slowly added 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (2.5 equiv.) and the reaction mixture was stirred at r.t. for 5 h. The reaction was quenched by addition of aq. sat. $NaHCO_3$ and the crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. After purification by silica gel column chromatography with an addition of 0.1% pyridine and co-lyophilization from benzene, the desired phosphoramidite **7a-c** was obtained as a mixture of diastereoisomers and rotamers as a white foam.

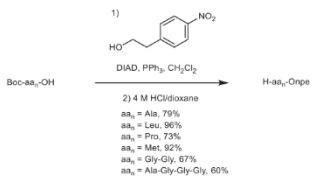
7a: Yield: 78%; R_f = 0.17 (2:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.3, 150.2, 148.8, 148.3; HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{53}H_{77}N_5O_{11}PSi_2$ 1046.4901, Found 1046.4896.

7b: Yield: 81%; R_f = 0.57 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K): δ (ppm): 150.2, 148.3; HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{52}H_{77}N_5O_{11}PSi_2$ 1032.4744; Found 1032.4745.

7c: Yield: 89%; R_f = 0.45 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.0, 149.9, 149.4, 149.3; HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{58}H_{88}N_6O_{12}PSi_2$ 1145.5585; Found 1145.5595.

2.2 Npe-protected amino acids and peptides

The npe-protected amino acids Gly, Val, Thr, Phe and Asp were synthesized following previously reported procedures in the literature.[4]



Scheme S2. Synthesis of npe-protected amino acids and peptides.

Procedure A (for Ala, Leu, Pro and Met): Step 1. **Boc-aa-OH** (1.0 equiv.), 2-(4-nitrophenyl)ethanol (1.3 equiv.) and PPh_3 (1.3 equiv.) were dissolved in dry CH_2Cl_2 and stirred at 0°C under nitrogen atmosphere. Diisopropyl azodicarboxylate (DIAD) (1.3 equiv.) was added dropwise and the reaction was stirred at r.t. overnight. Afterwards, the reaction was stopped and the crude was washed two times with water. The organic layer was dried (Na_2SO_4), filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography affording the **Boc-aa-Onpe** as a white solid. Step 2. **Boc-aa-Onpe** (1.0 equiv.) was dissolved in 4 M HCl in 1,4-dioxane at 0°C. After the reaction was stirred at r.t. for 1 h, the mixture was concentrated obtaining a white solid. The white solid was triturated with Et_2O, filtered and washed with additional Et_2O. The npe-protected amino acid **H-aa-Onpe** chloride salt was isolated as a white solid. For **Pro**, an oil was obtained which was washed with aq. sat. $NaHCO_3$ and the

crude was extracted with EtOAc. The organic layer was dried (Na_2SO_4), filtered and concentrated under reduced pressure. The crude was purified by silica gel column chromatography affording the **H-Pro-Onpe** as a pale-yellow oil.

H-Ala-Onpe•HCl: Yield: 79% over two steps; IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2843 (m), 1730 (s), 1598 (m), 1345 (s), 1269 (w), 1233 (s), 1195 (m), 1115 (m), 820 (m), 746 (m); ^1H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm): 8.59 (br s, 3H), 8.19-8.17 (m, 2H), 7.61-7.59 (m, 2H), 4.50-4.36 (m, 2H), 4.00 (q, *J* = 7.2 Hz, 1H), 3.10 (t, *J* = 6.3 Hz, 2H), 1.33 (d, *J* = 7.2 Hz, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 169.9, 146.4, 146.3, 130.4, 123.5, 65.3, 47.8, 33.9, 15.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{11}H_{15}N_2O_4$: 239.1026; Found 239.1027.

H-Leu-Onpe•HCl: Yield: 96% over two steps; IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3663 (w), 2871 (m), 1737 (s), 1589 (m), 1516 (s), 1503 (s), 1380 (s), 1260 (w), 1207 (m), 1109 (w), 959 (w), 856 (m), 812 (m), 735 (s); ^1H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 8.63 (s, 3H), 8.17 (d, *J* = 8.7 Hz, 2H), 7.60 (d, *J* = 8.7 Hz, 2H), 4.51-4.38 (m, 2H), 3.82 (t, *J* = 6.5 Hz, 1H), 3.11 (t, *J* = 6.5 Hz, 2H), 1.54-1.43 (m, 3H), 0.75 (t, *J* = 5.3 Hz, 6H); $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 169.8, 146.4, 130.4, 123.4, 65.3, 50.4, 33.8, 23.7, 22.2, 21.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{14}H_{21}N_2O_4$ 281.1496; Found 281.1495.

H-Pro-Onpe: Yield: 73% over two steps; R_f = 0.30 (9:1 CH_2Cl_2/IPA); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3400 (w), 2879 (w), 1649 (s), 1513 (s), 1432 (m), 1318 (s), 1159 (w), 1048 (m), 856 (m), 747 (m); ^1H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.18-8.16 (m, 2H); 7.40-7.38 (m, 2H); 4.40-4.37 (m, 2H); 3.74-3.71 (m, 1H); 3.07 (t, *J* = 6.7 Hz, 2H); 3.05-2.87 (m, 2H); 2.29 (br s, 1H); 2.13-2.02 (m, 1H); 1.77-1.67 (m, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 175.4, 147.0, 145.6, 129.9, 123.9, 64.3, 59.8, 47.1, 35.0, 30.4, 25.6; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{13}H_{17}N_2O_4$ 265.1183; Found 265.1179.

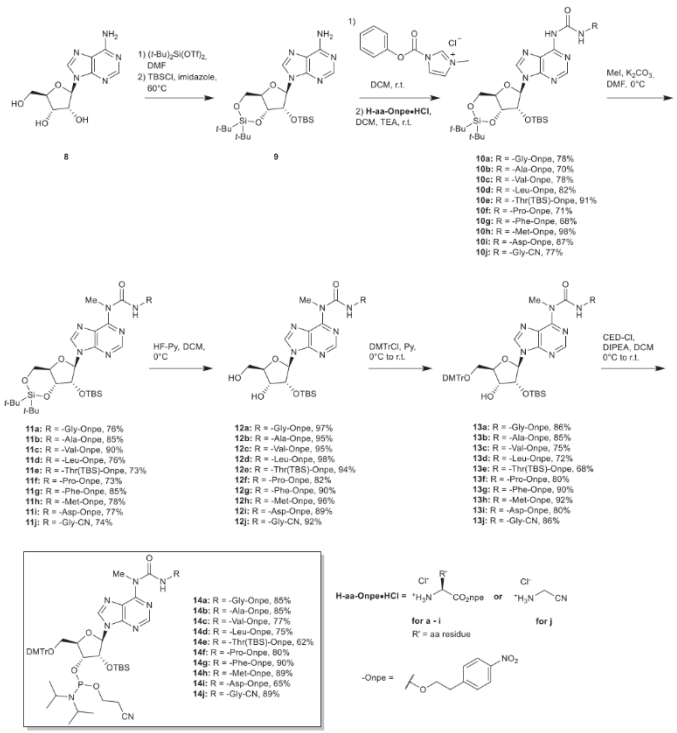
H-Met-Onpe•HCl: Yield: 92% over two steps; IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2852 (w), 1756 (m), 1743 (m), 1598 (w), 1567 (w), 1509 (s), 1347 (s), 1279 (m), 1256 (w), 1230 (w), 1206 (m), 1194 (m), 1148 (w), 1109 (w), 1066 (m), 1000 (w), 856 (m), 827 (m), 793 (w), 769 (m), 744 (s), 694 (m); ^1H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 8.77 (s, 3H), 8.18-8.15 (m, 2H), 7.62-7.59 (m, 2H), 4.50-4.41 (m, 2H), 4.01 (br s, 1H), 3.11 (t, *J* = 6.3 Hz, 2H), 2.55-2.43 (m, 1H), 2.38-2.27 (m, 1H), 2.03-1.86 (m, 5H); $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 169.1, 146.3, 146.3, 130.4, 123.5, 65.4, 50.8, 33.8, 29.3, 28.2, 14.1; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{13}H_{19}N_2O_4S$ 299.1060; Found 299.1058.

Procedure B (for Gly-Gly and Ala-Gly-Gly-Gly): Step 1. **Boc-aa$_n$-OH** (1.0 equiv.) and 2-(4-nitrophenyl)ethanol (1.3 equiv.) were suspended in dry ACN. Dry pyridine was added giving a solution. The solution was stirred at 0°C under nitrogen atmosphere. *N,N'*-dicyclohexylcarbodiimide (DCC) (1.3 equiv.) and HOBt (1.3 equiv.) were added and the reaction was stirred at r.t. overnight. After that, the reaction was quenched with 1 M aq. citric acid solution at r.t. for 30 min. The crude was diluted with EtOAc and filtered. The precipitate was washed with EtOAc two times. The organic layer was washed with aq. sat. $NaHCO_3$ solution and water. The organic layer was dried (Na_2SO_4), filtered and concentrated. The crude was purified by silica gel column chromatography affording the boc- and npe-protected peptide. Step 2. **Boc-aa$_n$-Onpe** (1.0 equiv.) was dissolved in 4 M HCl in 1,4-dioxane at 0°C. After the reaction was stirred at r.t. for 2 h, the mixture was concentrated obtaining a white solid. The white solid was triturated with Et_2O, filtered and washed with additional Et_2O. The npe-protected peptide **H-aa$_n$-Onpe** chloride salt was isolated as a white solid.

H-Gly-Gly-Onpe•HCl: Yield: 67% over two steps; IR (ATR) $\tilde{\nu}$ (cm^{-1}): 1741 (s), 1677 (m), 1660 (s), 1570 (w), 1512 (s), 1479 (w), 1229 (w), 1207 (s), 1199 (s), 1109 (m); ^1H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 8.91 (t, *J* = 5.7 Hz, 1H), 8.21 (br s, 3H), 8.19-8.17 (m, 2H), 7.58-7.56 (m, 2H), 4.34 (t, *J* = 6.4 Hz, 2H), 3.91 (d, *J* = 5.7 Hz, 2H), 3.59 (s, 2H), 3.06 (t, *J* = 6.4 Hz, 2H); $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 169.4, 166.6, 146.5, 146.3, 130.3, 123.5, 64.3, 40.6, 39.7 (the signal overlaps with that of the solvent), 34.0; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{12}H_{16}N_3O_5$ 282.1084; Found 282.1086.

H-Ala-Gly-Gly-Gly-Onpe•HCl: Yield: 60% over two steps; IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3222 (w), 2931 (w), 1743 (w), 1654 (s), 1514 (s), 1188 (s), 1117 (m), 871 (m), 856 (m), 697 (m); ^1H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 8.80 (t, *J* = 5.5 Hz, 1H), 8.38-8.33 (m, 2H), 8.25 (br s, 3H), 8.18-8.16 (m, 2H), 7.58-7.55 (m, 2H), 4.31 (t, *J* = 6.4 Hz, 2H), 3.91-3.87 (m, 1H), 3.84-3.79 (m, 4H), 3.76-3.72 (m, 2H), 3.04 (t, *J* = 6.4 Hz, 2H), 1.36 (d, *J* = 6.9 Hz, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 169.9, 169.7, 169.3, 168.6, 146.5, 146.3, 130.3, 123.5, 64.2, 48.2, 42.0, 41.7, 40.6, 34.0, 17.1; HRMS (ESI) *m/z* $[M+H]^+$: Calcd. for $C_{17}H_{24}N_5O_7$ 410.1670; Found: 410.1671.

2.3 Nucleobase-modified N^6-carbamoyl adenosine phosphoramidites



Scheme S3. Synthesis of nucleobase-modified N^6-carbamoyl adenosine phosphoramidites.

General procedure for the synthesis of 10a-j:

The compounds **10a**, **10c**, **10e**, **10g** and **10i** were previously reported in the literature.[4]

To a solution of silyl-protected adenosine **9**[5] (1.0 equiv.) in DCM was added 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (2.0 equiv.). The resulting suspension was stirred at r.t. for 16 h and then **H-aa-Onpe•HCl** (2.0 equiv.) together with NEt_3 (2.0 equiv.) was added. After stirring for 16 h, the reaction mixture was quenched by the addition of aq. sat. $NaHCO_3$ and the crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by silica gel column chromatography furnished the amino acid-modified adenosine derivative **10a-j** as a white foam.

10b: Yield: 70%; R_f = 0.70 (4:3 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3239 (w), 2929 (w), 2855 (w), 1744 (m), 1698 (s), 1610 (m), 1586 (m), 1519 (s), 1463 (m), 1344 (s), 1250 (s), 1138 (s), 1054 (s), 998 (m), 894 (s), 825 (s), 782 (s), 749 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.02 (d, J = 7.3 Hz, 1H), 8.82 (s, 1H), 8.51 (s, 1H), 8.27 (s, 1H), 8.07 (d, J = 8.7 Hz, 2H), 7.38 (d, J = 8.7 Hz, 2H), 6.00 (s, 1H), 4.68-4.59 (m, 1H), 4.58 (d, J = 4.6 Hz, 1H), 4.55-4.37 (m, 4H), 4.27-4.22 (m, 1H), 4.14-4.05 (m, 1H), 3.09 (t, J = 6.5 Hz, 2H), 1.49 (d, J = 7.2 Hz, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.94 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 173.0, 153.6, 151.2, 150.3, 149.8, 146.9, 145.5, 141.8, 130.0, 129.9, 123.8, 121.1, 92.5, 75.9, 75.8, 74.9, 67.9, 64.7, 49.2, 35.0, 27.6, 27.1, 26.0, 22.8, 20.5, 18.5, 18.4, -4.2, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{36}H_{56}N_7O_9Si_2$ 786.3673; Found 786.3682.

10d: Yield: 82%; R_f = 0.43 (4:3 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3237 (w), 2168 (w), 1666 (s), 1572 (w), 1511 (s), 1429 (w), 1335 (m), 1271 (s), 1227 (m), 1178 (w), 1151 (w), 1119 (m), 1090 (s), 1019 (w), 908 (m), 843 (s), 781 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 9.87 (d, J = 7.7 Hz, 1H), 8.50-8.49 (m, 2H), 8.18 (s, 1H), 8.06 (d, J = 8.6 Hz, 2H), 7.38 (d, J = 8.6 Hz, 2H), 5.98 (s, 1H), 4.64-4.56 (m, 2H), 4.54-4.38 (m, 4H), 4.24 (td, J = 10.0, 5.1 Hz, 1H), 4.07 (dd, J = 10.0, 9.1 Hz, 1H), 3.09 (t, J = 6.5 Hz, 2H), 1.71-1.65 (m, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.97-0.91 (m, 15H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 173.0, 153.8, 151.2, 150.3, 149.8, 146.9, 145.7, 141.6, 129.9, 123.7, 121.1, 92.6, 75.9, 75.7, 74.9, 67.9, 64.5, 52.1, 41.2, 34.9, 27.6, 27.1, 26.0, 25.2, 23.0, 22.9, 22.0, 20.5, 18.4, -4.2, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{39}H_{62}N_7O_9Si_2$ 828.4142; Found 828.4149.

10f: Yield: 71%; R_f = 0.25 (6:4 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2933 (w), 1741 (w), 1649 (w), 1519 (m), 1401 (m), 1344 (s), 1166 (m), 1140 (m), 1057 (s), 750 (m); For major rotamer: ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.65 (s, 1H), 8.10-8.08 (m, 2H), 7.97 (s, 1H), 7.40-7.38 (m, 2H), 5.96 (s, 1H), 4.64 (d, J = 4.6 Hz, 1H), 4.61-4.59 (m, 1H), 4.50 (dd, J = 9.4, 4.6 Hz, 2H), 4.45-4.42 (m, 2H), 4.24 (ddd, J = 9.4, 9.4, 4.6 Hz, 1H), 4.04 (dd, J = 9.4, 9.4 Hz, 1H), 3.67-3.64 (m, 2H), 3.10-3.06 (m, 2H), 2.26-2.19 (m, 1H), 2.06-2.00 (m, 3H), 1.07 (s, 9H), 1.04 (s, 9H), 0.93 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.2, 152.8, 150.8, 147.0, 140.7, 129.9, 123.8, 123.7, 123.2, 92.6, 75.9, 75.6, 74.9, 67.9, 64.3, 59.6, 47.0, 35.0, 29.7, 27.6, 27.1, 26.0, 24.7, 22.9, 20.5, 18.4, -4.1, -4.9 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{38}H_{58}N_7O_9Si_2$ 812.3829; Found 812.3835.

10h: Yield: 98%; R_f = 0.30 (2:1 *i*-Hexane/EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.01 (d, J = 7.8 Hz, 1H), 8.49 (s, 1H), 8.19 (s, 1H), 8.12 (s, 1H), 8.06 (d, J = 8.7 Hz, 2H), 7.38 (d, J = 8.7 Hz, 2H), 5.97 (s, 1H), 4.74 (td, J = 7.8, 5.1 Hz, 1H), 4.62 (d, J = 4.6 Hz, 1H), 4.55-4.41 (m, 4H), 4.24 (td, J = 10.0, 5.1 Hz, 1H), 4.05 (dd, J = 10.5, 9.2 Hz, 1H), 3.09 (t, J = 6.5 Hz, 2H), 2.54 (dd, J = 8.1, 6.0 Hz, 1H), 2.28-1.98 (m, 1H), 2.10-2.05 (m, 4H), 1.08 (s, 9H), 1.05 (s, 9H), 0.94 (s, 9H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 171.9, 153.6, 151.2, 150.1, 149.8, 146.9, 145.5, 141.3, 129.9, 123.8, 121.1, 92.6, 75.9, 75.7, 74.9, 67.9, 64.8, 52.6, 34.9, 31.7, 30.2, 27.6, 27.2, 26.0, 22.9, 20.5, 18.5, 15.6, -4.1, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{38}H_{60}N_7O_9SSi_2$ 846.3706; Found 846.3704.

10j: Yield: 77%; R_f = 0.29 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3119 (w), 2930 (m), 2857 (m), 2168 (w), 1706 (s), 1658 (m), 1612 (m), 1525 (m), 1466 (m), 1394 (w), 1353 (m), 1249 (s), 1141 (s), 1058 (s), 997 (m), 892 (m), 825 (s), 783 (s); ^{1}H NMR (600 MHz, $CDCl_3$, 298 K) δ (ppm): 10.10 (t, J = 5.7 Hz, 1H), 8.54 (s, 1H), 8.45 (s, 1H), 8.10 (s, 1H), 5.98 (s, 1H), 4.59 (d, J = 4.6 Hz, 1H), 4.51 (dd, J = 9.3, 4.6 Hz, 1H), 4.45 (dd, J = 9.3, 4.6 Hz, 1H), 4.37 (d, J = 5.7 Hz, 2H), 4.29-4.21 (m, 1H), 4.06 (dd, J = 10.5, 9.3 Hz, 1H), 1.09 (s, 9H), 1.05 (s, 9H), 0.94 (s, 9H), 0.17 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (150 MHz, $CDCl_3$, 298 K) δ (ppm): 153.7, 151.1, 150.1, 149.8, 141.5, 121.2, 116.4, 92.6, 76.0, 75.8, 75.0, 67.9, 28.4, 27.6, 27.2, 26.0, 22.9, 20.5, 18.5, -4.1, -4.8; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{27}H_{46}N_7O_9Si_2$ 604.3093; Found 604.3094.

General procedure for the synthesis of 11a-j:

The amino acid-modified adenosine derivative **10a-j** (1.0 equiv.) was dissolved in DMF and cooled to 0°C. To the solution were added K_2CO_3 (3.0 equiv.) together with MeI (2.0 equiv.) and the reaction was stirred at r.t. for 2 h. The reaction mixture was diluted with H_2O and extracted three times with EtOAc. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated. The obtained residue was purified by silica gel column chromatography to give **11a-j** as a white foam.

11a: Yield: 76%; R_f = 0.22 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3235 (w), 2932 (w), 2858 (w), 1749 (w), 1686 (m), 1568 (w), 1521 (s), 1470 (m), 1347 (s), 1264 (s), 1167 (w), 1135 (m), 1055 (m), 1000 (w), 894 (w), 827 (m), 732 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.97 (t, J = 5.4 Hz, 1H), 8.50 (s, 1H), 8.11 (d, J = 8.7 Hz, 2H), 7.97 (s, 1H), 7.38 (d, J = 8.7 Hz, 2H), 6.01 (s, 1H), 4.56 (d, J = 4.6 Hz, 1H), 4.51 (dd, J = 9.2, 5.2 Hz, 1H), 4.45-4.38 (m, 3H), 4.27-4.22 (m, 1H), 4.16 (dd, J = 5.2, 1.7 Hz, 2H), 4.07-3.95 (m, 4H), 3.08 (t, J = 6.6 Hz, 2H), 1.07 (s, 9H), 1.04 (s, 9H), 0.94 (s, 9H), 0.18 (s, 3H), 0.15 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.3, 156.2, 153.2, 151.7, 150.2, 147.0, 145.5, 139.3, 129.9, 123.8, 122.7, 92.4, 76.1, 75.7, 74.8, 68.0, 64.6, 43.0, 35.0, 34.8, 27.6, 27.1, 26.0, 22.9, 20.5, 18.5, -4.1, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{36}H_{56}N_7O_9Si_2$ 786.3673; Found 786.3674.

11b: Yield: 85%; R_f = 0.60 (5:3 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3190 (w), 2934 (w), 2858 (w), 1744 (m), 1686 (m), 1567 (m), 1519 (s), 1469 (s), 1344 (s), 1250 (m), 1166 (m), 1133 (m), 1057 (s), 1000 (m), 895 (m), 825 (s), 777 (m), 749 (m); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.51 (s, 1H), 8.10 (d, J = 8.7 Hz, 2H), 8.01 (s, 1H), 7.38 (d, J = 8.7 Hz, 2H), 6.01 (s, 1H), 4.63-4.54 (m, 2H), 4.54-4.48 (m, 1H), 4.46-4.39 (m, 3H), 4.30-4.20 (m, 1H), 4.07-4.00 (m, 1H), 3.96 (s, 3H), 3.09 (t, J = 6.5 Hz, 2H), 1.46 (d, J = 7.2 Hz, 3H), 1.07 (s, 9H), 1.04 (s, 9H), 0.94 (s, 9H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 173.4, 155.4, 153.3, 151.6, 150.2, 147.0,

145.6, 139.3, 129.9, 123.8, 122.8, 92.5, 77.4, 76.0, 75.7, 74.8, 68.0, 64.6, 50.0, 35.0, 34.7, 27.6, 27.1, 26.0, 22.9, 20.5, 18.5, 18.4, -4.1, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{37}H_{58}N_7O_9Si_2$ 800.3829; Found 800.3836.

11c: Yield: 90%; R_f = 0.34 (1:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3246 (m), 2977 (w), 1732 (m), 1686 (s), 1524 (s), 1469 (w), 1372 (m), 1254 (s), 1177 (m), 1147 (w), 1107 (s), 1050 (s), 1020 (m), 926 (m), 853 (w), 790 (m), 744 (w); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 11.07 (d, *J* = 7.6 Hz, 1H), 8.48 (s, 1H), 8.09 (d, *J* = 8.7 Hz, 2H), 7.97 (s, 1H), 7.38 (d, *J* = 8.7 Hz, 2H), 5.99 (s, 1H), 4.59 (d, *J* = 4.6 Hz, 1H), 4.54-4.47 (m, 2H), 4.46-4.39 (m, 3H), 4.27-4.22 (m, 1H), 4.03 (dd, *J* = 10.5, 9.2 Hz, 1H), 3.97 (s, 3H), 3.09 (t, *J* = 6.6 Hz, 2H), 2.25-2.18 (m, 1H), 1.07 (s, 9H), 1.04 (s, 9H), 0.99 (d, *J* = 6.8 Hz, 3H), 0.95-0.93 (m, 12H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.4, 156.0, 153.3, 151.6, 150.0, 146.9, 145.7, 139.3, 129.9, 123.8, 122.8, 92.5, 76.0, 75.6, 74.9, 68.0, 64.4, 59.8, 35.0, 34.7, 30.8, 27.6, 27.1, 26.0, 22.9, 20.5, 19.6, 18.5, 18.2, -4.1, -4.8; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{39}H_{62}N_7O_9Si_2$ 828.4142; Found 828.4143.

11d: Yield: 76%; R_f = 0.47 (5:3 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3230 (w), 2933 (w), 1740 (s), 1690 (s), 1580 (s), 1520 (s), 1469 (s), 1345 (s), 1259 (s), 1134 (s), 1057 (s), 1013 (s), 900 (w), 826 (s), 780 (s), 750 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.92 (d, *J* = 6.8 Hz, 1H), 8.47 (s, 1H), 8.09 (d, *J* = 8.7 Hz, 2H), 7.97 (s, 1H), 7.38 (d, *J* = 8.7 Hz, 2H), 6.00 (s, 1H), 4.59 (d, *J* = 4.6 Hz, 1H), 4.57-4.48 (m, 2H), 4.47-4.39 (m, 3H), 4.25 (td, *J* = 10.1, 5.1 Hz, 1H), 4.03 (dd, *J* = 10.5, 9.2 Hz, 1H), 3.96 (s, 3H), 3.08 (t, *J* = 6.6 Hz, 2H), 1.73-1.61 (m, 3H), 1.07 (s, 9H), 1.05 (s, 9H), 0.95-0.94 (m, 12H), 0.93 (s, 3H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 173.4, 155.7, 153.3, 151.6, 150.1, 146.9, 145.7, 139.3, 129.9, 123.8, 122.8, 92.5, 76.0, 75.6, 74.9, 68.0, 64.5, 53.0, 41.2, 35.0, 34.7, 27.6, 27.1, 26.0, 25.3, 23.0, 22.9, 22.1, 20.5, 18.5, -4.1, -4.8; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{40}H_{64}N_7O_9Si_2$ 842.4299; Found 842.4296.

11e: Yield: 73%; R_f = 0.34 (4:3 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3237 (w), 2931 (s), 2857 (s), 1737 (s), 1701 (s), 1610 (s), 1520 (s), 1465 (s), 1345 (s), 1250 (s), 1136 (w), 1057 (s), 998 (w), 894 (w), 840 (s), 777 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 11.00 (d, *J* = 8.7 Hz, 1H), 8.43 (s, 1H), 8.01-7.96 (m, 3H), 7.32 (d, *J* = 8.7 Hz, 2H), 6.02 (s, 1H), 4.60 (d, *J* = 4.6 Hz, 1H), 4.58 (dd, *J* = 8.7, 1.7 Hz, 1H), 4.53-4.46 (m, 3H), 4.43-4.30 (m, 2H), 4.29-4.22 (m, 1H), 4.04 (t, *J* = 9.5 Hz, 1H), 3.98 (s, 3H), 3.03 (t, *J* = 6.5 Hz, 2H), 1.23 (d, *J* = 6.2 Hz, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.88 (s, 9H), 0.19 (s, 3H), 0.16 (s, 3H), 0.05 (s, 3H), -0.05 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 171.3, 156.4, 153.4, 151.6, 150.2, 146.8, 145.7, 139.4, 129.9, 123.7, 122.8, 92.5, 76.0, 75.7, 74.9, 68.9, 68.0, 64.6, 60.6, 34.9, 27.6, 27.2, 26.0, 25.7, 22.9, 21.3, 20.5, 18.5, 18.0, -4.1, -4.2, -4.9, -5.3; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{44}H_{74}N_7O_{10}Si_3$ 944.4799; Found: 944.4793.

11f: Yield: 73%; R_f = 0.30 (8:2 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2933 (w), 1744 (w), 1683 (m), 1583 (m), 1392 (m), 1345 (s), 1166 (m), 1056 (s), 1002 (m), 783 (m); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.46 (s, 1H), 8.17-8.15 (m, 2H), 7.84 (s, 1H), 7.39-7.38 (m, 2H), 5.93 (s, 1H), 4.58-4.57 (m, 1H), 4.54 (br s, 2H), 4.49 (dd, *J* = 9.8, 5.1 Hz, 2H), 4.39-4.35 (m, 1H), 4.22 (ddd, *J* = 9.8, 9.8, 5.1 Hz, 1H), 4.03 (dd, *J* = 9.8, 9.8 Hz, 1H), 3.55 (br s, 3H), 3.06 (br s, 2H), 2.18-2.16 (m, 1H), 1.88-1.86 (m, 3H), 1.08 (s, 9H), 1.04 (s, 9H), 0.92 (s, 9H), 0.15 (s, 3H), 0.14 (s, 3H) (some proton signals of proline appeared too broad for an unequivocal assignment); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.1, 156.9, 153.3, 152.5, 150.8, 147.0, 145.6, 139.4, 129.9, 123.9, 92.6, 76.0, 75.6, 74.8, 67.9, 64.6, 60.0, 48.0, 35.0, 34.9, 27.6, 27.1, 26.0, 24.2, 22.9, 20.5, 18.5, -4.2, -4.8 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{39}H_{60}N_7O_9Si_2$ 826.3985; Found 826.3991.

11g: Yield: 85%; R_f = 0.50 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2931 (w), 2857 (w), 1738 (w), 1682 (s), 1568 (s), 1518 (s), 1469 (s), 1344 (s), 1261 (s), 1166 (s), 1134 (s), 1056 (s), 1011 (s), 895 (w), 826 (s), 778 (s); ^{1}H NMR (400 MHz, CD_2Cl_2, 298 K) δ (ppm): 10.87 (d, *J* = 6.9 Hz, 1H), 8.30 (s, 1H), 8.04 (d, *J* = 8.7 Hz, 2H), 7.99 (s, 1H), 7.34 (d, *J* = 8.7 Hz, 2H), 7.29-7.24 (m, 3H), 7.20-7.12 (m, 2H), 6.02 (s, 1H), 4.77 (q, *J* = 6.5 Hz, 1H), 4.59 (d, *J* = 4.6 Hz, 1H), 4.53-4.42 (m, 2H), 4.38 (td, *J* = 6.5, 3.9 Hz, 2H), 4.25 (td, *J* = 10.0, 5.1 Hz, 1H), 4.06 (dd, *J* = 10.5, 9.2 Hz, 1H), 3.88 (s, 3H), 3.13 (dd, *J* = 6.4, 2.1 Hz, 2H), 3.03 (t, *J* = 6.4 Hz, 2H), 1.09 (s, 9H), 1.06 (s, 9H), 0.96 (s, 9H), 0.19 (s, 3H), 0.17 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, CD_2Cl_2, 298 K) δ (ppm): 172.4, 155.9, 153.5, 152.1, 150.3, 147.3, 146.5, 139.9, 137.2, 130.4, 129.9, 129.0, 127.6, 124.0, 123.1, 92.8, 76.5, 76.1, 75.3, 68.3, 65.0, 56.3, 38.3, 35.3, 34.9, 27.8, 27.4, 26.2, 23.1, 20.8, 18.8, -4.0, -4.7; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{43}H_{62}N_7O_9Si_2$ 876.4142; Found 876.4148.

11h: Yield: 78%; R_f = 0.30 (2.5:1 *i*-Hexane/EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 11.08 (d, *J* = 7.2 Hz, 1H), 8.49 (s, 1H), 8.09 (d, *J* = 8.6 Hz, 2H), 7.97 (s, 1H), 7.38 (d, *J* = 8.6 Hz, 2H), 6.00 (s, 1H), 4.70 (td, *J* = 7.2, 5.1 Hz, 1H), 4.59 (d, *J* = 4.6 Hz, 1H), 4.52 (dd, *J* = 9.2, 5.1 Hz, 1H), 4.46-4.41 (m, 3H), 4.25 (td, *J* = 10.1, 5.1 Hz, 1H), 4.03 (dd, *J* = 10.5, 9.2 Hz, 1H), 3.96 (s, 3H), 3.09 (t, *J* = 6.5 Hz, 2H), 2.52 (td, *J* = 7.3, 1.9 Hz, 2H), 2.24-1.99 (m, 2H), 2.07 (s, 3H), 1.07 (s, 8H), 1.04 (s, 9H), 0.95 (s, 9H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.3, 155.7, 153.2, 151.7, 150.1, 147.0, 145.6, 139.4, 129.9, 123.8, 122.8, 92.5, 76.0, 75.7, 74.9, 68.0, 64.8, 53.5, 35.0, 34.8, 31.7, 30.3, 27.6, 27.1, 26.0, 22.9, 20.5, 18.5, 15.6, -4.1, -4.8; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{39}H_{62}N_7O_9SSi_2$ 860.3863; Found 860.3858.

11i: Yield: 77%; R_f = 0.40 (8:2 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2933 (w), 1737 (m), 1683 (m), 1569 (m), 1518 (s), 1344 (s), 1166 (m), 1057 (m), 1000 (m), 780 (m); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 11.32 (d, *J* = 7.5 Hz, 1H), 8.40 (s, 1H), 8.09-8.06 (m, 2H), 8.02-7.99 (m, 2H), 7.98 (s, 1H), 7.34-7.32 (m, 4H), 6.01 (s, 1H), 4.87 (dt, *J* = 7.5, 4.5 Hz, 1H), 4.61 (d, *J* = 4.5 Hz, 1H), 4.52 (dd, *J* = 9.2, 4.5 Hz, 1H), 4.46-4.40 (m, 3H), 4.40-4.23 (m, 3H), 4.04 (dd, *J* = 9.2, 9.2 Hz, 1H), 3.94 (s, 3H), 3.06-2.99 (m, 4H), 2.97-2.95 (m, 2H), 1.07 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.19 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 171.0, 170.9, 155.6, 152.9, 151.7, 149.9, 146.9, 146.8, 145.6, 145.5, 139.5, 129.8 (×2), 123.8, 123.7, 122.6, 92.5, 76.0, 75.6, 74.9, 68.0, 65.0, 64.4, 50.5, 36.6, 34.9, 34.8, 34.7, 27.6, 27.1, 26.0, 22.9, 20.5, 18.5, -4.1, -4.9; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{46}H_{65}O_{13}N_8Si_2$ 993.4203; Found 993.4215.

11j: Yield: 74%; R_f = 0.34 (4:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3121 (w), 2933 (m), 2896 (w), 2857 (m), 2168 (w), 1692 (s), 1570 (s), 1525 (s), 1469 (s), 1422 (w), 1360 (m), 1328 (m), 1308 (w), 1299 (w), 1278 (m), 1249 (m), 1218 (w), 1198 (w), 1165 (s), 1141 (s), 1111 (m), 1062 (s), 1024 (s), 1001 (s), 968 (w), 889 (m), 825 (s), 784 (s); ^{1}H NMR (600 MHz, $CDCl_3$, 298 K) δ (ppm): 11.14 (t, *J* = 5.6 Hz, 1H), 8.53 (s, 1H), 7.99 (s, 1H), 6.01 (s, 1H), 4.56 (d, *J* = 4.6 Hz, 1H), 4.52 (dd, *J* = 9.4, 4.6 Hz, 1H), 4.39 (dd, *J* = 9.4, 4.6 Hz, 1H), 4.34 (dd, *J* = 5.2, 2.0 Hz, 2H), 4.26 (td, *J* = 9.4, 5.2 Hz, 1H), 4.04-4.02 (m, 1H), 4.02 (s, 3H), 1.07 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.18 (s, 3H), 0.16 (s, 3H); ^{13}C{^{1}H} NMR (150 MHz, $CDCl_3$, 298 K) δ (ppm): 156.0, 152.9, 151.9, 150.1, 139.6, 122.8, 116.7, 92.5, 76.1, 75.7, 74.9, 67.9, 35.0, 29.2, 27.6, 27.2, 26.0, 22.9, 20.5, 18.5, -4.1, -4.8; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{28}H_{48}N_7O_9Si_2$ 618.3250; Found 618.3256.

General procedure for the synthesis of 12a-j:

A solution of the modified adenosine derivative **11a-j** (1.0 equiv.) in DCM/pyridine (9:1 v/v) inside a plastic reaction vessel was cooled to 0°C. Subsequently, a solution of 70% HF-pyridine (5.0 equiv.) was slowly added and the reaction mixture was stirred at 0°C for 2 h. The reaction mixture was diluted with aq. sat. $NaHCO_3$ solution and extracted three times with DCM. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude product was purified by silica gel column chromatography to isolate the 3',5'-deprotected adenosine derivative **12a-j** as a white foam.

12a: Yield: 97%; R_f = 0.37 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2932 (w), 2857 (w), 1738 (w), 1688 (m), 1606 (w), 1581 (m), 1571 (m), 1518 (s), 1471 (m), 1445 (w), 1345 (s), 1253 (m), 1219 (m), 1135 (m), 1083 (m), 1031 (m), 858 (w), 838 (s), 780 (s), 750 (m); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.89 (t, *J* = 5.3 Hz, 1H), 8.52 (s, 1H), 8.14 (d, *J* = 8.6 Hz, 2H), 7.96 (s, 1H), 7.39 (d, *J* = 8.6 Hz, 2H), 5.90 (d, *J* = 11.6 Hz, 1H), 5.81 (d, *J* = 7.3 Hz, 1H), 5.12 (dd, *J* = 7.3, 4.9 Hz, 1H), 4.43 (td, *J* = 6.7, 1.7 Hz, 2H), 4.39-4.33 (m, 2H), 4.24-4.10 (m, 2H), 4.03-3.91 (m, 4H), 3.82-3.70 (m, 1H), 3.09 (t, *J* = 6.7 Hz, 2H), 2.81 (s, 1H), 0.80 (s, 9H), -0.18 (s, 3H), -0.39 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.2, 155.9, 153.8, 151.2, 149.7, 147.0, 145.4, 141.6, 129.9, 123.9, 123.7, 91.4, 87.7, 74.2, 72.8, 64.8, 63.5, 43.0, 35.0, 25.6, 17.9, -5.2, -5.3; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{28}H_{46}N_7O_9Si$ 646.2651; Found 646.2645.

12b: Yield: 95%; R_f = 0.40 (100:5 DCM/IPA); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3191 (w), 2927 (w), 2856 (w), 1739 (m), 1681 (s), 1610 (m), 1568 (s), 1519 (s), 1469 (m), 1344 (s), 1261 (m), 1211 (w), 1143 (w), 1018 (m), 998 (m), 836 (s), 779 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.73 (br s, 1H), 8.58 (s, 1H), 8.15 (d, *J* = 8.7 Hz, 2H), 8.11 (s, 1H), 7.41 (d, *J* = 8.7 Hz, 2H), 5.84 (d, *J* = 6.9 Hz, 1H), 5.13-5.06 (m, 1H), 4.66-4.53 (m, 1H), 4.49-4.40 (m, 2H), 4.39-4.35 (m, 2H), 4.00-3.93 (m, 4H), 3.78 (d, *J* = 12.9 Hz, 1H), 3.10 (t, *J* = 6.6 Hz, 2H), 1.47 (d, *J* = 7.2 Hz, 3H), 0.82 (s, 9H), -0.14 (s, 3H), -0.34 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 173.4, 155.1, 153.9, 151.2, 149.7, 147.1, 145.6, 141.5, 130.0, 123.9, 123.8, 91.5, 87.7, 74.2, 72.9, 64.8, 63.5, 50.1, 35.0, 34.9, 25.6, 18.4, 18.0, -5.2, -5.3; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{29}H_{42}N_7O_9Si$ 660.2808; Found: 660.2807.

12c: Yield: 95%; R_f = 0.16 (100:3 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3244 (w), 2952 (w), 2929 (w), 2359 (w), 1736 (w), 1681 (m), 1571 (m), 1518 (s), 1469 (m), 1422 (w), 1345 (s), 1255 (m), 1187 (m), 1145 (m), 1089 (m), 1046 (w), 1016 (m), 907 (m), 857 (m), 837 (s), 780 (s), 746 (m); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 11.02 (d, *J* = 7.4 Hz, 1H), 8.54 (s, 1H), 8.15 (d, *J* = 8.6 Hz, 2H), 7.95 (s, 1H), 7.41 (d, *J* = 8.6 Hz, 2H), 5.93 (dd, *J* = 11.9, 1.5 Hz, 1H), 5.81 (d, *J* = 7.4 Hz, 1H), 5.15 (dd, *J* = 7.4, 4.8 Hz, 1H), 4.51-4.34 (m, 5H), 4.02-3.91 (m, 4H), 3.81-3.71 (m, 1H), 3.10 (t, *J* = 6.6 Hz, 2H), 2.81 (s, 1H), 2.29-2.15 (m, 1H), 0.98 (d, *J* = 6.8 Hz, 3H), 0.94 (d, *J* = 6.9 Hz, 3H), 0.81 (s, 9H), -0.16 (s, 3H), -0.37 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 172.4, 155.7, 154.0, 151.2, 149.6, 147.0, 145.6, 141.5, 129.9, 123.9, 123.7, 91.5, 87.7, 74.1, 72.8, 64.6, 63.5, 59.8, 35.0, 34.9, 30.8, 25.6, 19.6, 18.1, 17.9, -5.2, -5.3; HRMS (ESI) *m/z*: [M+H]$^+$ Calcd. for $C_{31}H_{48}N_7O_9Si$ 688.3121; Found 688.3120.

12d: Yield: 98%; R_f = 0.52 (9:1 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (s), 1469 (s), 1345 (s), 1313 (w), 1250 (s), 1129 (w), 1093 (s), 835 (s), 760 (s); ^{1}H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.86 (d, *J* = 7.1 Hz, 1H), 8.63 (s, 1H), 8.58 (s, 1H), 8.13 (d, *J* = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 6.13 (d, *J* = 5.8 Hz, 1H), 5.04 (dd, *J* = 8.3, 3.7 Hz, 1H), 4.98 (t, *J* = 4.7 Hz, 1H), 4.49-4.41 (m, 3H), 4.40-4.37 (m, 1H), 4.21 (dd, *J* = 2.6 Hz, 1H), 3.98 (d, *J* = 4.0 Hz, 1H), 3.92 (s, 3H), 3.90-

3.87 (m, 1H), 3.80- 3.75 (m, 1H), 3.15 (t, *J* = 6.4 Hz, 2H), 1.76-1.53 (m, 3H), 0.91 (dd, *J* = 6.4, 3.3 Hz, 6H), 0.81 (s, 9H), -0.05 (s, 3H), -0.18 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 173.4, 155.9, 154.0, 152.7, 150.5, 147.6, 147.5, 142.5, 131.1, 124.1, 123.6, 90.5, 87.5, 76.6, 72.4, 65.1, 62.8, 53.7, 41.7, 35.3, 34.8, 26.0, 25.8, 23.1, 22.1, 18.6, -4.9, -5.1; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{32}H_{48}N_7O_9Si$ 702.3277; Found: 702.3279.

12e: Yield: 94%; R_f = 0.37 (9:1 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (s), 1469 (s), 1345 (s), 1313 (w), 1250 (s), 1129 (w), 1093 (s), 835 (s), 760 (s); ^1H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.90 (d, *J* = 8.6 Hz, 1H), 8.48 (s, 1H), 8.10 (d, *J* = 8.6 Hz, 2H), 7.96 (s, 1H), 7.37 (d, *J* = 8.6 Hz, 2H), 5.82 (d, *J* = 7.3 Hz, 1H), 5.14 (dd, *J* = 7.3, 4.8 Hz, 1H), 4.56 (dd, *J* = 8.6, 1.8 Hz, 1H), 4.50-4.41 (m, 2H), 4.39-4.35 (m, 2H), 4.29-4.23 (m, 1H), 3.99 (s, 3H), 3.96 (dd, *J* = 13.0, 1.8 Hz, 1H), 3.76 (dd, *J* = 13.0, 1.8 Hz, 1H), 3.06 (t, *J* = 6.7 Hz, 2H), 1.22 (d, *J* = 6.2 Hz, 3H), 0.87 (s, 9H), 0.81 (s, 9H), 0.03 (s, 3H), -0.06 (s, 3H), -0.16 (s, 3H), -0.37 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 171.2, 156.1, 154.1, 151.2, 149.8, 147.0, 145.6, 141.5, 130.0, 123.9, 91.5, 87.7, 74.2, 72.9, 68.8, 64.9, 63.5, 60.7, 35.1, 35.0, 25.7, 25.6, 21.3, 18.0, 17.9, -4.1, -5.2, -5.3, -5.3; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{36}H_{58}N_7O_{10}Si_2$ 804.3778; Found: 804.3768.

12f: Yield: 82%; R_f = 0.15 (98:2 DCM/IPA); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2929 (w), 1743 (w), 1679 (m), 1585 (s), 1519 (m), 1391 (m), 1344 (s), 1090 (m), 1046 (m), 780 (m); ^1H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 8.47 (s, 1H), 8.17-8.14 (m, 2H), 7.81 (s, 1H), 7.39 (br s, 2H), 6.35 (d, *J* = 12.1 Hz, 1H), 5.76 (d, *J* = 7.4 Hz, 1H), 5.15 (dd, *J* = 7.4, 4.8 Hz, 1H), 4.61-4.46 (m, 2H), 4.38 (s, 1H), 4.35 (d, *J* = 4.8 Hz, 1H), 3.96 (d, *J* = 12.1 Hz, 1H), 3.75 (dd, *J* = 12.1, 12.1 Hz, 1H), 3.58 (br s, 3H), 3.06 (br s, 2H), 2.80 (s, 1H), 2.18-2.13 (m, 1H), 1.92-1.88 (m, 3H), 0.79 (s, 9H), -0.19 (s, 3H), -0.41 (s, 3H) (some proton signals of proline appeared too broad for an unequivocal assignment); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 171.9, 156.3, 153.7, 152.0, 150.1, 147.0, 141.3, 129.9, 123.9, 91.3, 87.8, 74.1, 73.0, 64.7, 63.5, 59.9, 47.9, 35.0, 34.9, 25.6, 24.2, 17.9, -5.2, -5.4 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{31}H_{44}N_7O_9Si$ 686.2964; Found 686.2963.

12g: Yield: 90%; R_f = 0.50 (98:2 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3391 (w), 3194 (w), 2951 (w), 2855 (w), 1738 (s), 1681 (s), 1568 (s), 1516 (s), 1469 (s), 1344 (s), 1261 (s), 1171 (s), 1128 (s), 1091 (s), 1016 (s), 836 (s), 779 (s); ^1H NMR (400 MHz, CD_2Cl_2, 298 K) δ (ppm): 10.81 (d, *J* = 6.7 Hz, 1H), 8.31 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.99 (s, 1H), 7.39 (d, *J* = 8.7 Hz, 2H), 7.33-7.20 (m, 3H), 7.17-7.13 (m, 2H), 5.83 (d, *J* = 7.3 Hz, 1H), 5.67 (dd, *J* = 11.9, 2.0 Hz, 1H), 5.10 (dd, *J* = 7.3, 4.7 Hz, 1H), 4.78 (td, *J* = 6.8, 5.7 Hz, 1H), 4.46-4.36 (m, 2H), 4.36-4.32 (m, 2H), 3.93-3.89 (m, 4H), 3.79-3.68 (m, 1H), 3.13 (dd, *J* = 6.3, 3.6 Hz, 2H), 3.09-2.99 (m, 2H), 2.81 (s, 1H), 0.80 (s, 9H), -0.18 (s, 3H), -0.38 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, CD_2Cl_2, 298 K) δ (ppm): 172.3, 155.6, 154.1, 151.7, 150.3, 149.9, 147.4, 146.4, 142.1, 137.1, 130.4, 129.9, 129.1, 127.7, 124.1, 91.7, 88.2, 74.7, 73.3, 65.2, 63.7, 56.3, 38.3, 35.3, 35.0, 25.8, 18.2, -5.1, -5.2; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{35}H_{46}N_7O_9Si$ 736.3121; Found 736.3118.

12h: Yield: 96%; R_f = 0.40 (100:5 DCM/MeOH); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.93 (d, *J* = 7.1 Hz, 1H), 8.63 (s, 1H), 8.58 (s, 1H), 8.12 (d, *J* = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 6.12 (d, *J* = 5.8 Hz, 1H), 5.02 (dd, *J* = 8.3, 3.7 Hz, 1H), 4.97 (dd, *J* = 5.9, 4.7 Hz, 1H), 4.60 (td, *J* = 7.5, 5.3 Hz, 1H), 4.46 (td, *J* = 6.5, 1.9 Hz, 2H), 4.38 (td, *J* = 4.4, 2.9 Hz, 1H), 4.22-4.18 (m, 1H), 3.96 (d, *J* = 4.0 Hz, 1H), 3.92-3.87 (m, 4H), 3.82-3.74 (m, 1H), 3.16 (t, *J* = 6.3 Hz, 2H), 2.52 (t, *J* = 7.9 Hz, 2H), 2.15-1.97 (m, 5H), 0.80 (s, 9H), -0.05 (s, 3H), -0.18 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 172.6, 155.9, 154.0, 152.8, 150.5, 147.7, 147.4, 142.5, 131.1, 124.2, 123.6, 90.4, 87.5, 76.6, 72.4, 65.4, 62.8, 54.2, 35.3, 34.9, 32.2, 30.6, 26.0, 18.6, 15.1, -4.9, -5.1; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{31}H_{46}N_7O_9SSi$ 720.2841; Found 720.2833.

12i: Yield: 89%; R_f = 0.15 (97:3 DCM/IPA); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2930 (w), 1735 (m), 1682 (m), 1570 (m), 1516 (s), 1468 (m), 1261 (m), 1018 (m), 837 (m), 781 (m); ^1H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 11.31 (d, *J* = 7.5 Hz, 1H), 8.48 (s, 1H), 8.13-8.10 (m, 4H), 7.96 (s, 1H), 7.38-7.33 (m, 4H), 5.87 (d, *J* = 12.5 Hz, 1H), 5.81 (d, *J* = 7.3 Hz, 1H), 5.13 (dd, *J* = 7.3, 4.8 Hz, 1H), 4.87 (dt, *J* = 7.3, 4.8 Hz, 1H), 4.49-4.25 (m, 6H), 3.97 (s, 3H), 3.97-3.94 (m, 1H), 3.77 (dd, *J* = 12.5, 12.5 Hz, 1H), 3.06 (t, *J* = 6.7 Hz, 2H), 3.01 (t, *J* = 6.7 Hz, 2H), 2.99-2.94 (m, 2H), 2.80 (s, 1H), 0.81 (s, 9H), -0.16 (s, 3H), -0.38 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.9, 170.8, 155.4, 153.7, 151.3, 150.0, 149.6, 147.0, 145.5, 145.3, 141.7, 129.9, 129.8, 123.9, 123.8, 123.7, 91.5, 87.7, 74.2, 72.8, 65.2, 64.5, 63.5, 50.6, 36.6, 34.9, 34.8 (×2), 25.6, 17.9, -5.2, -5.3; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{38}H_{49}O_{13}N_8Si$ 853.3182; Found 853.3187.

12j: Yield: 92%; R_f = 0.44 (10:1 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3347 (w), 2929 (m), 2857 (m), 1731 (w), 1681 (s), 1570 (s), 1515 (s), 1462 (s), 1422 (m), 1360 (w), 1329 (w), 1262 (s), 1217 (m), 1126 (s), 1035 (s), 994 (m), 901 (m), 866 (m), 835 (s), 779 (s); ^1H NMR (600 MHz, $CDCl_3$, 298 K) δ (ppm): 11.02 (t, *J* = 5.7 Hz, 1H), 8.55 (s, 1H), 7.98 (s, 1H), 5.82 (d, *J* = 7.4 Hz, 1H), 5.77 (d, *J* = 11.4 Hz, 1H), 5.13 (dd, *J* = 7.4, 4.8 Hz, 1H), 4.42-4.29 (m, 4H), 4.04 (s, 3H), 4.00-3.92 (m, 1H), 3.82-3.73 (m, 1H), 2.79 (s, 1H), 0.81 (s, 9H), -0.16 (s, 3H), -0.38 (s, 3H); $^{13}C\{^1H\}$

NMR (150 MHz, $CDCl_3$, 298 K) δ (ppm): 155.7, 153.5, 151.5, 149.6, 141.9, 123.8, 116.6, 110.2, 91.5, 87.7, 74.2, 72.9, 63.5, 35.2, 29.2, 25.6, 18.0, -5.1, -5.3; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{20}H_{32}N_7O_5Si$ 478.2229; Found 478.2231.

General procedure for the synthesis of 13a-j:

The 3',5'-deprotected adenosine derivative **12a-j** (1.0 equiv.) was dissolved in pyridine and DMTrCl (1.5 equiv.) was added. The reaction mixture was stirred at r.t. for 16 h and afterwards the solvents were removed *in vacuo*. Purification by silica gel column chromatography with an addition of 0.1% pyridine afforded the DMTr-protected adenosine derivative **13a-j** as a white or pale-yellow foam.

13a: Yield: 86%; R_f = 0.16 (1:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3320 (w), 2929 (w), 2853 (w), 1749 (w), 1681 (m), 1606 (w), 1568 (m), 1510 (s), 1466 (m), 1345 (s), 1300 (w), 1250 (s), 1213 (m), 1176 (s), 1066 (w), 1034 (s), 1005 (w), 916 (w), 856 (m), 834 (s), 782 (m), 699 (m); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.85 (t, *J* = 5.6 Hz, 1H), 8.49 (s, 1H), 8.46 (s, 1H), 8.12 (d, *J* = 8.7 Hz, 2H), 7.58 (d, *J* = 8.7 Hz, 2H), 7.53-7.47 (m, 2H), 7.37 (dd, *J* = 9.0, 2.3 Hz, 4H), 7.32-7.19 (m, 3H), 6.86 (dd, *J* = 9.0, 2.3 Hz, 4H), 6.18 (d, *J* = 4.4 Hz, 1H), 5.07 (t, *J* = 4.4 Hz, 1H), 4.54-4.50 (m, 1H), 4.43 (t, *J* = 6.4 Hz, 2H), 4.31-4.26 (m, 1H), 4.11 (d, *J* = 5.8 Hz, 2H), 3.98 (d, *J* = 5.8 Hz, 1H), 3.93 (s, 3H), 3.77 (s, 6H), 3.49-3.43 (m, 2H), 3.13 (t, *J* = 6.4 Hz, 2H), 0.86 (s, 9H), 0.07 (s, 3H), -0.03 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 170.8, 159.6, 156.5, 153.7, 153.2, 150.7, 147.6, 146.1, 141.8, 136.7, 131.1, 131.0, 129.0, 128.6, 127.6, 124.2, 123.2, 113.9, 89.9, 87.1, 84.7, 76.5, 71.9, 65.1, 64.3, 55.5, 43.4, 35.3, 34.8, 26.1, 18.7, -4.6, -4.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{49}H_{58}N_7O_{11}Si$ 948.3958; Found 948.3949.

13b: Yield: 85%; R_f = 0.70 (4:1 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2928 (w), 1741 (w), 1681 (w), 1610 (m), 1568 (m), 1508 (s), 1463 (m), 1344 (m), 1251 (s), 1174 (m), 1018 (m), 835 (s), 781 (m); ^1H NMR (400 MHz, CD_2Cl_2, 298 K) δ (ppm) 10.91 (d, *J* = 6.5 Hz, 1H), 8.44 (s, 1H), 8.17 (s, 1H), 8.10 (d, *J* = 8.7 Hz, 2H), 7.49-7.44 (m, 2H), 7.42 (d, *J* = 8.7 Hz, 2H), 7.35 (d, *J* = 8.9 Hz, 4H), 7.32-7.20 (m, 3H), 6.82 (d, *J* = 8.9 Hz, 4H), 6.08 (d, *J* = 4.9 Hz, 1H), 4.97 (t, *J* = 4.9 Hz, 1H), 4.58-4.45 (m, 1H), 4.45-4.33 (m, 3H), 4.25-4.20 (m, 1H), 3.92 (s, 3H), 3.77 (s, 6H), 3.49 (dd, *J* = 10.7, 3.1 Hz, 1H), 3.39 (dd, *J* = 10.7, 4.2 Hz, 1H), 3.09 (t, *J* = 6.5 Hz, 2H), 2.64 (br s, 1H), 1.44 (d, *J* = 7.2 Hz, 3H), 0.86 (s, 9H), 0.02 (s, 3H), -0.09 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, CD_2Cl_2, 298 K) δ (ppm): 173.7, 155.7, 153.7, 152.8, 150.5, 147.4, 146.5, 145.4, 140.6, 136.2, 130.6, 130.4, 128.6, 128.4, 127.4, 124.0, 123.1, 113.7, 89.2, 87.1, 84.6, 76.1, 71.9, 65.0, 63.9, 55.8, 50.5, 35.4, 34.9, 25.9, 18.5, 18.4, -4.6, -4.9; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{50}H_{60}N_7O_{11}Si$ 962.4115; Found 962.4128.

13c: Yield: 75%; R_f = 0.15 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2950 (w), 2850 (w), 1730 (w), 1670 (w), 1607 (m), 1577 (s), 1508 (s), 1464 (w), 1347 (s), 1250 (s), 1177 (s), 1150 (w), 1090 (s), 1035 (m), 981 (w), 913 (s), 866 (s), 839 (s), 701 (s); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 11.03 (d, *J* = 7.7 Hz, 1H), 8.52 (s, 1H), 8.49 (s, 1H), 8.09 (d, *J* = 8.7 Hz, 2H), 7.58 (d, *J* = 8.7 Hz, 2H), 7.52-7.48 (m, 2H), 7.41-7.34 (m, 4H), 7.29-7.16 (m, 3H), 6.86 (dd, *J* = 9.0, 2.7 Hz, 4H), 6.19 (d, *J* = 4.3 Hz, 1H), 5.06 (t, *J* = 4.3 Hz, 1H), 4.54-4.37 (m, 4H), 4.32-4.28 (m, 1H), 3.97 (d, *J* = 5.9 Hz, 1H), 3.93 (s, 3H), 3.77 (s, 6H), 3.48-3.44 (m, 2H), 3.14 (t, *J* = 6.3 Hz, 2H), 2.81-2.80 (m, 2H), 0.98 (d, *J* = 6.8 Hz, 3H), 0.94 (d, *J* = 6.8 Hz, 3H), 0.87 (s, 9H), 0.08 (s, 3H), -0.01 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 172.4, 159.5, 156.2, 153.8, 153.1, 150.5, 147.5, 147.4, 146.1, 141.8, 136.7, 136.6, 131.0, 129.1, 128.9, 128.6, 127.5, 124.1, 123.2, 113.9, 90.0, 87.1, 84.6, 76.5, 71.8, 65.0, 64.3, 60.5, 55.5, 35.3, 34.8, 31.4, 26.1, 19.7, 18.7, 18.4, -4.6, -4.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{52}H_{64}N_7O_{11}Si$ 990.4428; Found 990.4430.

13d: Yield: 72%; R_f = 0.20 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2950 (w), 2852 (w), 1729 (w), 1670 (w), 1607 (s), 1577 (s), 1508 (s), 1464 (w), 1347 (s), 1250 (s), 1177 (s), 1152 (w), 1091 (s), 1035 (s), 981 (w), 913 (s), 866 (s), 839 (s), 699 (s); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.89 (d, *J* = 7.1 Hz, 1H), 8.50 (s, 1H), 8.49 (s, 1H), 8.10 (d, *J* = 8.8 Hz, 2H), 7.58 (d, *J* = 8.8 Hz, 2H), 7.50 (d, *J* = 7.2 Hz, 2H), 7.38 (dd, *J* = 9.0, 2.5 Hz, 4H), 7.31-7.25 (m, 2H), 7.25-7.19 (m, 1H), 6.86 (dd, *J* = 9.0, 2.5 Hz, 4H), 6.18 (d, *J* = 4.3 Hz, 1H), 5.05 (t, *J* = 4.3 Hz, 1H), 4.53-4.37 (m, 4H), 4.29 (dd, *J* = 4.3, 4.3 Hz, 1H), 3.97 (d, *J* = 5.9 Hz, 1H), 3.91 (s, 3H), 3.77 (s, 6H), 3.51-3.42 (m, 2H), 3.14 (t, *J* = 6.3 Hz, 2H), 1.74-1.56 (m, 3H), 0.92 (d, *J* = 2.0 Hz, 3H), 0.92 (d, *J* = 2.0 Hz, 3H), 0.86 (s, 9H), 0.07 (s, 3H), -0.02 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 173.4, 159.6, 156.0, 153.8, 153.1, 150.6, 147.6, 147.5, 146.1, 141.9, 136.7, 131.1, 131.0, 131.0, 129.1, 129.0, 128.6, 127.6, 124.1, 123.3, 113.9, 90.0, 87.1, 84.7, 76.5, 71.9, 65.1, 64.3, 55.5, 53.7, 41.7, 35.3, 34.8, 26.1, 25.8, 23.1, 22.2, 18.7, -4.6, -4.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{53}H_{66}N_7O_{11}Si$ 1004.4584; Found 1004.4579.

13e: Yield: 68%; R_f = 0.22 (2:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2908 (w), 1757 (w), 1718 (w), 1670 (w), 1608 (w), 1507 (s), 1441 (w), 1294 (w), 1248 (s), 1177 (s), 1090 (s), 1034 (s), 975 (s), 913 (s), 869 (s), 776 (s), 703 (s); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.89 (d, *J* = 8.7 Hz, 1H), 8.50 (s, 1H), 8.42 (s, 1H), 8.03 (d, *J* = 8.7 Hz, 2H), 7.52 (d, *J* = 8.8 Hz, 4H), 7.41-7.32 (m, 4H), 7.28 (t, *J* = 7.4 Hz, 2H), 7.25-7.17 (m, 1H), 6.85 (dd, *J* = 8.8, 1.8 Hz, 4H), 6.20 (d, *J* = 4.5 Hz, 1H), 5.11 (t, *J* = 4.5 Hz, 1H), 4.54-4.48 (m, 3H), 4.45-4.31 (m, 2H), 4.31-4.26

(m, 1H), 3.98 (d, *J* = 5.7 Hz, 1H), 3.95 (s, 3H), 3.76 (s, 6H), 3.48 (qd, *J* = 10.5, 4.1 Hz, 2H), 3.12 (t, *J* = 6.3 Hz, 2H), 1.25 (d, *J* = 6.3 Hz, 3H), 0.88 (s, 9H), 0.86 (s, 9H), 0.07 (s, 6H), 0.03 (s, 6H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 171.7, 159.6, 153.1, 150.4, 142.0, 136.7, 131.1, 131.0, 129.0, 128.6, 127.6, 124.1, 123.4, 113.9, 90.0, 87.1, 84.8, 76.4, 71.9, 69.7, 65.5, 64.4, 61.2, 55.5, 35.3, 35.1, 26.1, 26.0, 21.6, 18.7, 18.4, -4.2, -4.6, -4.8, -5.2; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{57}H_{76}N_7O_{12}Si_2$ 1106.5085; Found: 1106.5103.

13f: Yield: 80%; R_f = 0.30 (6:4 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2930 (w), 1743 (w), 1680 (m), 1582 (s), 1509 (m), 1391 (m), 1345 (s), 1249 (s), 1174 (s), 782 (m); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 8.37 (s, 1H), 8.33 (s, 1H), 8.18-8.16 (m, 2H), 7.56 (br s, 2H), 7.49 (d, *J* = 7.3 Hz, 2H), 7.37-7.35 (m, 4H), 7.28 (dd, *J* = 7.3, 7.3 Hz, 2H), 7.21 (t, *J* = 7.3 Hz, 1H), 6.88-6.84 (m, 4H), 6.11 (d, *J* = 4.8 Hz, 1H), 5.10 (dd, *J* = 4.8, 4.8 Hz, 1H), 4.53-4.49 (m, 1H), 4.45-4.38 (m, 2H), 4.25 (dd, *J* = 8.2, 4.8 Hz, 1H), 3.94 (d, *J* = 4.8 Hz, 1H), 3.78 (s, 6H), 3.50-3.39 (m, 6H), 3.12-3.10 (m, 2H), 1.84-1.79 (m, 2H), 1.71 (br s, 1H), 0.83 (s, 9H), 0.04 (s, 3H), -0.08 (s, 3H) (some proton signals of proline appeared too broad for an unequivocal assignment); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 172.6, 159.6, 153.8, 152.9, 152.2, 147.7, 147.3, 146.0, 141.7, 136.7, 136.6, 131.0, 130.9, 129.1, 128.6, 127.6, 124.3, 113.9, 89.5, 87.1, 84.8, 84.7, 76.3, 72.0, 71.9, 65.2, 64.3, 60.7, 55.5, 48.4, 35.3, 34.6, 30.4, 26.1, 18.7, -4.7, -4.9; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{52}H_{62}N_7O_{11}Si$ 988.4270; Found 988.4280.

13g: Yield: 90%; R_f = 0.50 (5:1 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3538 (w), 2953 (w), 2855 (w), 1738 (w), 1681 (w), 1568 (s), 1508 (s), 1463 (s), 1344 (s), 1249 (s), 1174 (s), 1031 (s), 1016 (s), 834 (s), 781 (s); ^1H NMR (400 MHz, CD_2Cl_2, 298 K) δ (ppm): 10.89 (d, *J* = 6.8 Hz, 1H), 8.22 (s, 1H), 8.15 (s, 1H), 8.07 (d, *J* = 8.7 Hz, 2H), 7.50-7.45 (m, 2H), 7.40-7.16 (m, 12H), 7.14 (dd, *J* = 7.3, 2.1 Hz, 2H), 6.82 (d, *J* = 8.9 Hz, 4H), 6.06 (d, *J* = 5.0 Hz, 1H), 4.97 (t, *J* = 5.0 Hz, 1H), 4.79-4.75 (m, 1H), 4.41-4.35 (m, 3H), 4.26-4.20 (m, 1H), 3.89 (s, 3H), 3.76 (s, 6H), 3.48 (dd, *J* = 10.7, 3.1 Hz, 1H), 3.38 (dd, *J* = 10.7, 4.2 Hz, 1H), 3.12 (d, *J* = 6.3 Hz, 2H), 3.03 (t, *J* = 6.5 Hz, 2H), 2.64 (d, *J* = 4.8 Hz, 1H), 0.86 (s, 9H), 0.02 (s, 3H), -0.10 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, CD_2Cl_2, 298 K) δ (ppm): 172.4, 159.2, 155.9, 153.5, 152.7, 150.3, 150.2, 147.3, 146.5, 145.4, 140.6, 140.1, 137.2, 136.2, 136.2, 130.6, 130.6, 130.4, 129.9, 129.6, 129.0, 128.6, 128.4, 128.3, 128.2, 127.6, 127.4, 124.0, 123.0, 113.7, 113.6, 89.1, 87.1, 84.6, 76.1, 71.9, 65.1, 64.0, 56.3, 55.7, 38.3, 35.3, 34.9, 25.9, 18.4, -4.6, -4.9; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{56}H_{64}N_7O_{11}Si$ 1038.4428; Found 1038.4447.

13h: Yield: 92%; R_f = 0.25 (100:5 DCM/EtOAc); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.98 (d, *J* = 7.1 Hz, 1H), 8.50 (s, 1H), 8.49 (s, 1H), 8.14-8.07 (d, *J* = 8.9 Hz, 2H), 7.59 (d, *J* = 8.6 Hz, 2H), 7.50 (d, *J* = 7.2 Hz, 2H), 7.37 (dd, *J* = 9.0, 2.7 Hz, 4H), 7.29 (t, *J* = 7.4 Hz, 2H), 7.22 (t, *J* = 7.2 Hz, 1H), 6.86 (dd, *J* = 8.9, 3.2 Hz, 4H), 6.18 (d, *J* = 4.3 Hz, 1H), 5.07 (t, *J* = 4.6 Hz, 1H), 4.61 (td, *J* = 7.5, 5.3 Hz, 1H), 4.52 (q, *J* = 5.4 Hz, 1H), 4.47 (td, *J* = 6.1, 4.3 Hz, 2H), 4.29 (q, *J* = 4.4 Hz, 1H), 3.98 (d, *J* = 5.9 Hz, 1H), 3.91 (s, 3H), 3.77 (s, 6H), 3.47 (dd, *J* = 4.1, 2.1 Hz, 2H), 3.16 (t, *J* = 6.3 Hz, 2H), 2.53 (td, *J* = 7.2, 1.5 Hz, 2H), 2.17-1.95 (m, 5H), 0.86 (s, 9H), 0.07 (s, 3H), -0.03 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 172.6, 159.6, 159.6, 156.0, 153.7, 153.1, 150.7, 150.6, 147.6, 147.5, 146.1, 141.9, 136.7, 136.7, 131.0, 131.0, 131.0, 129.0, 128.6, 127.6, 124.6, 124.1, 123.3, 113.9, 90.0, 87.1, 84.7, 76.4, 71.9, 65.3, 64.3, 55.5, 54.1, 35.3, 34.8, 32.2, 30.7, 26.1, 18.7, 15.2, -4.6, -4.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{52}H_{64}N_7O_{11}SSi$ 1022.4148; Found 1022.4137.

13i: Yield: 80%; R_f = 0.25 (95:5 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2932 (w), 1737 (m), 1682 (m), 1571 (m), 1518 (s), 1464 (s), 1251 (s), 1176 (s), 1018 (m), 836 (s); ^1H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 11.21 (d, *J* = 7.5 Hz, 1H), 8.49 (s, 1H), 8.39 (s, 1H), 8.09-8.04 (m, 4H), 7.55-7.49 (m, 6H), 7.39-7.35 (m, 4H), 7.29 (dd, *J* = 7.5, 7.5 Hz, 2H), 7.21 (t, *J* = 7.5 Hz, 1H), 6.88-6.83 (m, 4H), 6.18 (d, *J* = 4.2 Hz, 1H), 5.04 (dd, *J* = 4.2, 4.2 Hz, 1H), 4.82 (dt, *J* = 7.5, 5.3 Hz, 1H), 4.53 (dd, *J* = 4.2, 4.2 Hz, 1H), 4.44-4.27 (m, 5H), 3.96 (d, *J* = 6.1 Hz, 1H), 3.90 (s, 3H), 3.77 (s, 6H), 3.51-3.44 (m, 2H), 3.11 (t, *J* = 6.2 Hz, 2H), 3.06 (t, *J* = 6.2 Hz, 2H), 2.96-2.94 (m, 2H), 0.86 (s, 9H), 0.08 (s, 3H), -0.01 (s, 3H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 171.5, 171.2, 159.6, 155.9, 153.5, 153.1, 150.4, 147.6, 147.5, 147.4, 147.2, 146.1, 141.9, 136.7, 131.0 (×2), 130.9, 129.0, 128.6, 127.6, 124.1, 124.0, 123.2, 113.9, 90.1, 87.1, 84.5, 76.5, 71.8, 65.6, 65.0, 64.2, 55.5, 51.4, 37.1, 35.3, 35.2, 34.8, 26.1, 18.7, -4.6, -4.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{59}H_{67}O_{15}N_8Si$ 1155.4489; Found 1155.4504.

13j: Yield: 86%; R_f = 0.21 (5:2 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3397 (w), 2954 (m), 2926 (s), 2854 (m), 2168 (w), 1682 (m), 1607 (m), 1569 (s), 1508 (s), 1462 (s), 1445 (m), 1362 (w), 1297 (w), 1249 (s), 1174 (s), 1134 (m), 1032 (s), 994 (m), 904 (w), 833 (s), 781 (m), 700 (m); ^1H NMR (600 MHz, $CDCl_3$, 298 K) δ (ppm): 11.02 (t, *J* = 5.7 Hz, 1H), 8.50 (s, 1H), 8.48 (s, 1H), 7.50 (d, *J* = 7.2 Hz, 2H), 7.38-7.36 (m, 4H), 7.29 (t, *J* = 7.6 Hz, 2H), 7.24-7.21 (m, 1H), 6.87-6.83 (m, 4H), 6.18 (d, *J* = 4.6 Hz, 1H), 5.08 (dd, *J* = 4.6, 4.6 Hz, 1H), 4.54-4.49 (m, 1H), 4.39 (d, *J* = 5.7 Hz, 2H), 4.28 (td, *J* = 4.7, 3.4 Hz, 1H), 3.98 (s, 3H), 3.96 (d, *J* = 5.8 Hz, 1H), 3.78 (s, 6H), 3.48 (dd, *J* = 10.2, 3.8 Hz, 1H), 3.45 (dd, *J* = 10.2, 4.7 Hz, 1H), 0.85 (s, 9H), 0.06 (s, 3H), -0.05 (s, 3H); $^{13}C\{^1H\}$ NMR (150 MHz, $CDCl_3$, 298 K) δ (ppm): 159.6, 159.6, 156.5, 153.5, 153.4, 150.7, 146.1, 142.2, 142.2, 136.7, 131.0, 131.0, 129.0, 129.0, 128.6, 127.6, 123.3, 118.2, 113.9, 89.9, 84.8, 76.4, 71.9, 64.4, 55.5, 55.5, 34.9, 29.7, 26.1, 18.7, -4.6, -4.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{41}H_{50}N_7O_7Si$ 780.3535; Found 780.3538.

General procedure for the synthesis of 14a-j:

To a solution of 5′-DMTr-protected adenosine derivative **13a-j** (1.0 equiv.) in anhydrous DCM, *N,N*-diisopropylethylamine (DIPEA) (4.0 equiv.) was added. After cooling down to 0°C, 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (2.5 equiv.) was added dropwise and the reaction mixture was stirred at r.t. for 5 h. Afterwards aq. sat. $NaHCO_3$ solution was added to the reaction mixture and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography with addition of 0.1% pyridine and co-lyophilized from benzene to afford the desired phosphoramidite **14a-j** as a mixture of diastereoisomers, as a white or pale-yellow foam.

14a: Yield: 85%; R_f = 0.15 (2:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.1, 148.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{68}H_{76}N_9O_{12}PSi$ 1148.5037; Found 1148.5052.

14b: Yield: 85%; R_f = 0.50 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, CD_2Cl_2, 298 K) δ (ppm): 150.6, 149.2; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{59}H_{77}N_9O_{12}PSi$ 1162.5193; Found 1162.5221.

14c: Yield: 77%; R_f = 0.35 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.1, 148.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{61}H_{81}N_9O_{12}PSi$ 1190.5506; Found 1190.5492.

14d: Yield: 75%; R_f = 0.38 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.1, 148.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{62}H_{83}N_9O_{12}PSi$ 1204.5663; Found 1204.5682.

14e: Yield: 62%; R_f = 0.43 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.2, 148.5; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{66}H_{93}N_9O_{13}PSi_2$ 1306.6164; Found 1306.6189.

14f: Yield: 80%; R_f = 0.30 (6:4 DCM/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm) 150.2, 148.6; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{61}H_{79}N_9O_{12}PSi$ 1188.5350; Found 1188.5388.

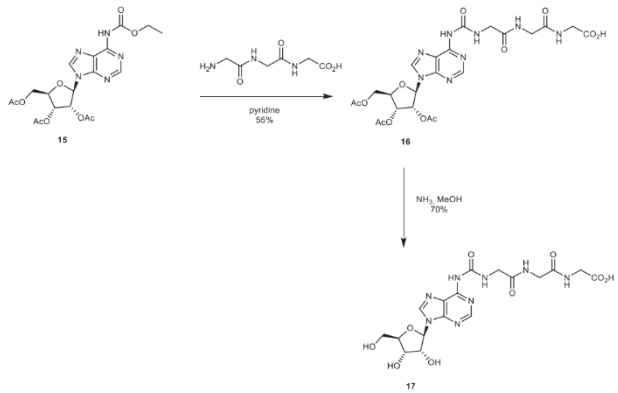
14g: Yield: 90%; R_f = 0.30 (5:3 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm) 150.7, 149.1; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{65}H_{81}N_9O_{12}PSi$ 1238.5506; Found 1238.5530.

14h: Yield: 89%; R_f = 0.30 (6:4 DCM/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm) 150.1, 148.6; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{61}H_{81}N_9O_{12}PSSi$ 1222.5227; Found 1222.5215.

14i: Yield: 65%; R_f = 0.15 (92:8 DCM/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.1, 148.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{68}H_{84}O_{16}N_{10}PSi$ 1355.5567; Found 1355.5590.

14j: Yield: 89%; R_f = 0.39 (2:1 EtOAc/*i*-Hexane); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.3, 148.6; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{50}H_{67}N_9O_8PSi$ 980.4614; Found 980.4611.

2.4 Nucleobase-modified N^6-triglycinylcarbamoyl adenosine nucleoside



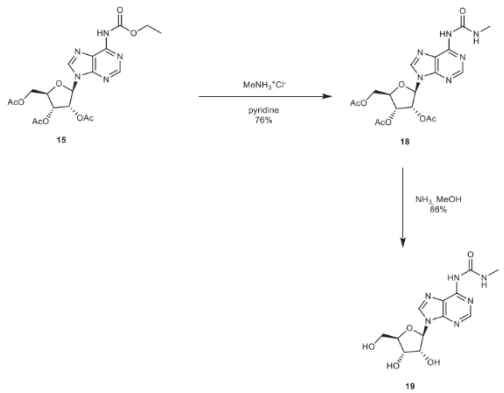
Scheme S4. Synthesis of N^6-triglycinylcarbamoyl adenosine **17**.

The compound **15** was synthesized according to a procedure previously described in the literature.[6,7]

Acetyl protected N^6-triglycinylcarbamoyl adenosine 16: Carbamate derivative **15** (0.25 g, 0.54 mmol, 1.0 equiv.) was dissolved in dry pyridine and **H-Gly-Gly-Gly-OH** (0.20 g, 1.1 mmol, 2.0 equiv.) was added. The mixture was stirred under reflux for 7 h and at r.t. overnight. After that, the crude was filtered and concentrated. The crude was resuspended in toluene and concentrated. Finally, the crude was crystallized from EtOH affording the product **16** as a white solid (0.18 g, 0.30 mmol, 55% yield). IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3353 (w), 1745 (m), 1732 (m), 1697 (m), 1608 (w), 1590 (w), 1515 (s), 1216 (s), 1038 (s), 902 (w); ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 12.58 (br s, 1H), 9.94 (s, 1H), 9.66 (t, J = 5.3 Hz, 1H), 8.65 (s, 1H), 8.59 (s, 1H), 8.38 (t, J = 5.8 Hz, 1H), 8.21 (t, J = 5.8 Hz, 1H), 6.30 (d, J = 5.4 Hz, 1H), 6.03 (dd, J = 5.4, 5.4 Hz, 1H), 5.63 (dd, J = 5.4, 5.4 Hz, 1H), 4.44-4.38 (m, 2H), 4.29-4.24 (m, 1H), 3.98 (d, J = 5.3 Hz, 2H), 3.77-3.75 (m, 4H), 2.12 (s, 3H), 2.04 (s, 3H), 2.01 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 171.2, 170.1, 169.5, 169.3, 169.2 (x2), 153.5, 151.1, 150.5, 150.1, 142.7, 120.5, 85.8, 79.6, 72.0, 70.0, 62.7, 43.1, 41.8, 40.6, 20.5, 20.4, 20.2; HRMS (ESI) m/z: $[M+Na]^+$ Calcd. for $C_{23}H_{28}O_{12}N_8Na$ 631.1718; Found 631.1721.

***N^6*-triglycinylcarbamoyl adenosine 17:** Protected adenosine derivative **16** (0.12 g, 0.20 mmol, 1.0 equiv.) was dissolved in 7 N NH_3 in MeOH. The reaction was heated at 40°C for 1.5 h and at r.t. overnight. After that, the crude was concentrated. Finally, the crude product was recrystallized from EtOH (5 mL) affording the product as a white solid (67 mg, 0.14 mmol, 70% yield). IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3281 (m), 2936 (w), 1691 (s), 1658 (s), 1551 (s), 1470 (s), 1240 (s), 1058 (s), 794 (m), 690 (s); ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 9.70 (t, J = 5.0 Hz, 1H), 8.68 (s, 1H), 8.56 (s, 1H), 8.41 (t, J = 5.7 Hz, 1H), 7.83 (t, J = 5.0 Hz, 1H), 5.98 (d, J = 5.6 Hz, 1H), 4.59 (dd, J = 5.6, 5.6 Hz, 1H), 4.18 (dd, J = 5.6, 5.6 Hz, 1H), 3.99-3.98 (m, 3H), 3.74 (d, J = 5.7 Hz, 2H), 3.71-3.56 (m, 4H); ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 171.2, 169.2, 168.5, 153.6, 150.9, 150.4, 150.3, 142.2, 120.4, 87.7, 85.7, 73.8, 70.3, 61.3, 43.0, 42.1, 42.0; HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{17}H_{23}O_9N_8$ 483.1582; Found 483.1583.

2.5 Nucleobase-modified N^6-methylurea adenosine nucleoside

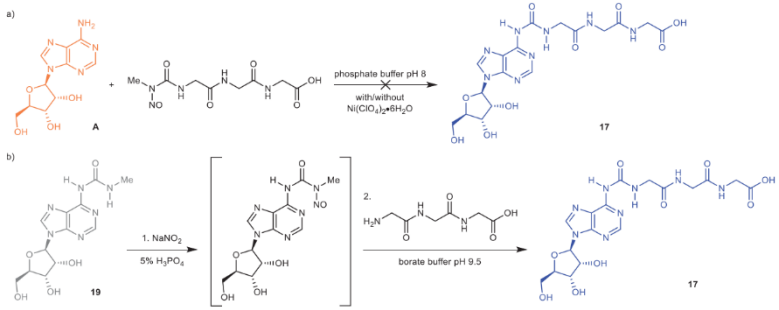


Scheme S5. Synthesis of N^6-methylurea adenosine **19**.

Acetyl protected N^6-methylurea adenosine 18: Carbamate derivative **15** (0.40 g, 0.86 mmol, 1.0 equiv.) was dissolved in dry pyridine and methylammonium chloride (0.17 g, 2.6 mmol, 3.0 equiv.) was added. The mixture was stirred under reflux overnight. After that, the crude was filtered, washed with EtOAc and concentrated. The crude was suspended in toluene and concentrated. Finally, the crude was purified by silica gel column chromatography (20 g, 95:5 DCM/IPA) affording the product **18** as a white foam (0.30 g, 0.65 mmol, 76% yield). IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3246 (w), 1744 (m), 1699 (m), 1590 (m), 1544 (m), 1469 (w), 1365 (w), 1212 (s), 1046 (m), 797 (w); ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 9.74 (s, 1H), 9.20 (c, J = 4.6 Hz, 1H), 8.63 (s, 1H), 8.57 (s, 1H), 6.29 (d, J = 5.3 Hz, 1H), 6.03 (dd, J = 5.3, 5.3 Hz, 1H), 5.64 (dd, J = 5.3, 5.3 Hz, 1H), 4.43-4.23 (m, 3H), 2.83 (d, J = 4.6 Hz, 3H), 2.12 (s, 3H), 2.04 (s, 3H), 2.01 (s, 3H); ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 170.1, 169.5, 169.3, 153.9, 151.1, 150.5, 150.0, 142.6, 120.3, 85.8, 79.6, 72.0, 70.0, 62.7, 26.3, 20.5, 20.4, 20.2; HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{18}H_{23}O_8N_6$ 451.1571; Found 451.1573.

***N^6*-methylurea adenosine 19:** Protected N^6-methylurea adenosine **18** (0.26 g, 0.58 mmol, 1.0 equiv.) was dissolved in 7 N NH_3 in MeOH. The reaction was heated at 40°C for 1.5 h and at r.t. overnight. After that, the crude was concentrated. Finally, the crude was triturated in EtOH, filtered and washed with EtOH affording the product as a white solid (0.16 g, 0.50 mmol, 86% yield). IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3360 (w), 1703 (m), 1584 (m), 1537 (m), 1462 (m), 1297 (m), 1245 (s), 1103 (m), 1057 (m), 795 (m); ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 9.57 (br s, 1H), 9.24 (c, J = 4.5 Hz, 1H), 8.65 (s, 1H), 8.54 (s, 1H), 5.97 (d, J = 5.7 Hz, 1H), 5.53 (d, J = 5.7 Hz, 1H), 5.24 (d, J = 4.8 Hz, 1H), 5.15 (dd, J = 5.7, 5.7 Hz, 1H); 4.62-4.58 (m, 1H), 4.19-4.15 (m, 1H), 3.98-3.95 (m, 1H), 3.71-3.54 (m, 2H), 2.83 (d, J = 4.5 Hz, 3H); ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 154.0, 150.8, 150.3, 150.2, 142.2, 120.2, 87.7, 85.7, 73.8, 70.3, 61.3, 26.3; HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{12}H_{17}O_5N_6$ 325.1254; Found 325.1257.

2.6 Nucleobase-modified N^6-triglycinylcarbamoyl adenosine nucleoside under prebiotic conditions



Scheme S6. Synthesis of N^6-triglycinylcarbamoyl adenosine **17** under prebiotic conditions using: a) nitroso derivative of the *N*-methylurea peptide and b) N^6-methylurea adenosine **19**.

Method A:[8] Adenosine **A** (2.67 mg, 10 µmol, 1.0 equiv.) was dissolved in 30 mM phosphate buffer pH 8 (370 µL). The nitroso derivative of the *N*-methylurea peptide (5.50 mg, 20 µmol, 2.0 equiv.) was dissolved in water (40 µL) and added to the adenosine's solution. Either water (40 µL) or $Ni(ClO_4)_2 \cdot 6H_2O$ (91.34 mg, 250 µmol, 25 equiv.) in water (40 µL) was added and the reaction was heated at 70°C for 24 h in a ThermoMixer. Finally, an aliquot (50 µL) of the reaction crude was diluted with water (up to 1 mL), filtered and analyzed by LC-MS (Buffer A: 2 mM $HCOONH_4$ pH 5.5 in H_2O and buffer B: 2 mM $HCOONH_4$ pH 5.5 in 20:80 H_2O/MeCN; Gradient: 0-20% of B in 30 min; Flow rate = 0.15 mL·min^{-1} and Injection: 5 µL).

This prebiotic synthetic method did not afford the N^6-triglycinylcarbamoyl adenosine **17**. We only detected the formation of traces of inosine when using the Ni(II) salt.

Method B: Step 1. N^6-methylurea adenosine **19** (1 mg, 3.08 µmol, 1.0 equiv.) was dissolved in 5% H_3PO_4 in water (140 µL) and cooled to 0°C in an ice bath. $NaNO_2$ (2.66 mg, 38.54 µmol, 12.5 equiv.) was dissolved in water (10 µL) and added to the previous solution. The reaction was incubated at 0°C for 2 h and -20°C for 22 h. After that, the adenosine's solution was allowed to reach 0°C. Step 2. The peptide (5.83 mg, 30.84 µmol, 10 equiv.) was dissolved in 30 mM borate buffer pH 9.5 (3 mL) and cooled down to 0°C. The adenosine's solution was added to the peptide's solution and the pH was adjusted to 9.5 with 4 N NaOH (60 µL). The reaction was stirred at r.t. for 1 h. Finally, an aliquot (25 µL) of the reaction crude was diluted with water (up to 1 mL), filtered and analyzed by LC-MS (Buffer A: 2 mM $HCOONH_4$ pH 5.5 in H_2O and buffer B: 2 mM $HCOONH_4$ pH 5.5 in 20:80 H_2O/MeCN; Gradient: 0-20% of B in 30 min; Flow rate = 0.15 mL·min^{-1} and Injection: 5 µL).

This prebiotic synthetic method afforded the N^6-triglycinylcarbamoyl adenosine **17** in 65% yield. The assignment and amount of the compounds observed in the HPL-chromatogram (Figure S1) was performed by analyzing separate solutions of those synthesized using non-prebiotic methods. Mass spectrometry analyses confirmed the assignments.

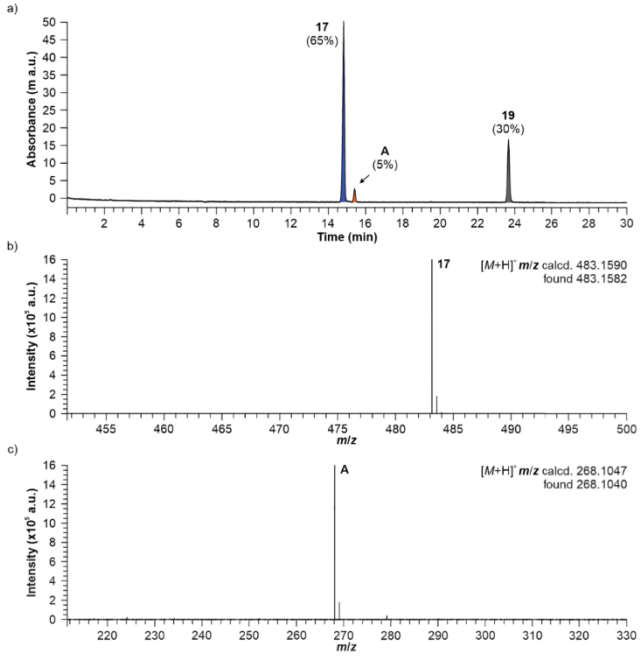
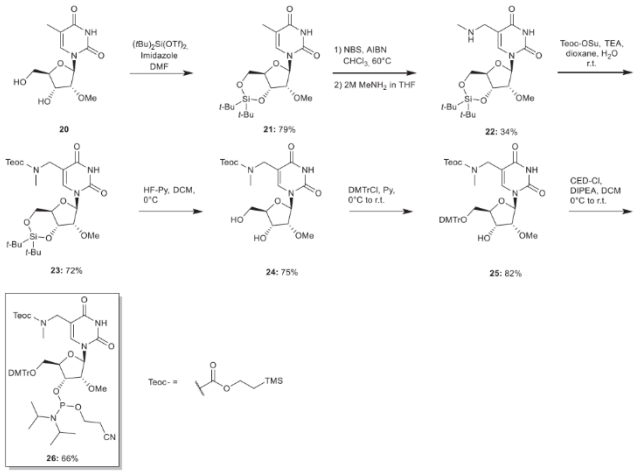


Figure S1. a) HPL-chromatogram of the reaction crude using the **Method B** shown in Scheme S6; mass spectra of the chromatographic peaks observed at: b) 14.8 and c) 15.4 min. The mass spectra confirmed the formation of the compounds **17** and **A**.

2.7 Nucleobase-modified 5-methyluridine 2'-methoxy phosphoramidite



Scheme S7. Synthesis of nucleobase-modified 5-methyluridine 2'-methoxy phosphoramidite.

General procedure for the synthesis of 21:

A suspension of 2'-OMe m^5U **20** (1.0 equiv.) in DMF was cooled to 0°C. Di-*tert*-butylsilyl *bis*(trifluoromethanesulfonate) (1.1 equiv.) was added dropwise and the mixture was stirred at r.t. for 30 min. To the reaction was added imidazole (2.5 equiv.) and the resulting solution was stirred at r.t. for 16 h. The crude was concentrated under reduced pressure and the residue was redissolved in EtOAc and washed with water, aq. sat. $NaHCO_3$ solution and brine. The organic layer was dried ($MgSO_4$), filtered and concentrated. The crude was purified by silica gel column chromatography to yield **21** as a white foam.

21: Yield: 79%; R_f = 0.29 (5:1 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2933 (w), 2859 (w), 1681 (m), 1471 (m), 1365 (w), 1323 (w), 1266 (w), 1148 (m), 1131 (m), 1064 (s), 1038 (m), 1919 (w), 952 (w), 907 (s), 826 (s), 727 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 7.02 (s, 1H), 5.63 (d, J = 0.8 Hz, 1H), 4.48-4.45 (m, 1H), 4.07-3.95 (m, 3H), 3.94-3.90 (m, 1H), 3.61 (s, 3H), 1.93 (s, 3H), 1.07 (s, 9H), 1.03 (s, 9H); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 163.6, 149.7, 136.1, 111.2, 91.9, 82.2, 77.4, 74.5, 67.4, 59.3, 27.5, 27.2, 22.9, 20.5, 12.8; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{19}H_{33}N_2O_8Si$ 413.2102; Found 413.2106.

General procedure for the synthesis of 22:

A solution of **21** (1.0 equiv.) in dry $CHCl_3$ was heated at 60°C. *N*-bromosuccinimide (NBS) (1.2 equiv., previously purified by recrystallization) and azobisisobutyronitrile (AIBN) (0.12 equiv.) were added and the reaction was stirred under reflux for 1.5 h. After that, the reaction mixture was cooled to r.t. and $MeNH_2$ (2 M in THF, 5.0 equiv.) was added. The resulting suspension was stirred for 2 h at r.t. and, subsequently, it was diluted with aq. sat. $NaHCO_3$ solution. The crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated. The crude was purified by silica gel column chromatography to furnish **22** as a yellow foam.

22: Yield: 34%; R_f = 0.30 (9:1 DCM/IPA); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2934 (w), 2859 (w), 1680 (s), 1468 (m), 1245 (s), 1201 (w), 1132 (m), 1064 (m), 1034 (m), 961 (w), 852 (w), 826 (s), 735 (w); ^{1}H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 8.02 (s, 1H), 5.81 (s, 1H), 4.43-4.32 (m, 2H), 4.25-4.18 (m, 1H), 4.11 (d, J = 5.0 Hz, 1H), 4.08-3.98 (m, 2H), 3.84 (d, J = 7.0 Hz, 2H), 3.58 (s, 3H), 2.62 (s, 3H), 1.07 (s, 9H), 1.03 (s, 9H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 163.9, 150.5, 142.6, 107.5, 91.7, 82.9, 77.8, 75.4, 67.8, 59.2, 45.8, 33.3, 27.8, 27.5, 23.1, 20.9; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{26}H_{36}N_3O_8Si$ 442.2368; Found 442.2370.

General procedure for the synthesis of 23:

To a solution of **22** (1.0 equiv.) in 1,4-dioxane and H_2O (1:1 v/v) were added teoc-OSu (1.1 equiv.) and triethylamine (TEA) (1.5 equiv.). The mixture was stirred at r.t. for 16 h. After that, the crude was diluted with water and extracted three times with Et_2O. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated. The obtained residue was purified by silica gel column chromatography to yield the teoc-protected compound **23** as a white solid.

23: Yield: 72%; R_f = 0.53 (95:5 DCM/IPA); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2948 (w), 2894 (w), 2859 (w), 1725 (m), 1464 (w), 1384 (w), 1280 (w), 1245 (s), 1198 (m), 1139 (m), 1057 (m), 1029 (m), 955 (w), 920 (w), 826 (s), 744 (m), 691 (w); For major rotamer: ^{1}H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.27 (br s, 1H), 7.54 (s, 1H), 5.76 (s, 1H), 4.47 (d, J = 4.1 Hz, 1H), 4.27-3.95 (m, 8H), 3.59 (s, 3H), 2.94 (s, 3H), 1.08 (s, 9H), 1.04-1.00 (m, 11H), 0.06 (s, 9H); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 150.6, 139.7, 111.1, 91.4, 83.0, 77.7, 75.4, 68.1, 63.8, 59.2, 45.5, 35.3, 27.8, 27.5, 23.2, 20.9, 18.4, -1.3; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{26}H_{48}N_3O_8Si_2$ 586.2975; Found 586.2981.

General procedure for the synthesis of 24:

The modified 2'-OMe 5-methyluridine **23** (1.0 equiv.) was dissolved in DCM/pyridine (9:1 v/v) and cooled to 0°C in a plastic reaction vessel. Subsequently, a solution of 70% HF-pyridine (5.0 equiv.) was slowly added, and the reaction mixture was stirred at 0°C for 2 h. The reaction was quenched by adding aq. sat. $NaHCO_3$ and the crude was extracted three times with DCM. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated. The crude product was purified by silica gel column chromatography to afford the diol compound **24** as a white foam.

24: Yield: 75%; R_f = 0.22 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3060 (w), 2951 (w), 1710 (m), 1463 (m), 1401 (m), 1249 (s), 1214 (m), 1114 (m), 1086 (m), 1062 (m), 988 (w), 938 (w), 838 (s), 769 (m), 694 (w); For major rotamer: ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.17 (br s, 1H), 8.09 (s, 1H), 5.99 (d, J = 4.3 Hz, 1H), 4.34 (s, 1H), 4.28-4.12 (m, 3H), 4.11-3.92 (m, 5H), 3.92-3.74 (m, 2H), 3.47 (s, 3H), 2.95 (s, 3H), 1.02 (s, 2H), 0.04 (s, 9H) (some proton signals appeared too broad for an unequivocal assignment); $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 163.9, 157.2, 151.2, 140.3, 138.6, 111.0, 87.9, 84.4, 69.9, 63.9, 62.1, 58.5, 45.8, 35.3, 18.3, -1.4 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{18}H_{32}N_3O_8Si$ 446.1953; Found 446.1954.

General procedure for the synthesis of 25:

To a solution of the 2'-OMe 3',5'-deprotected 5-methyluridine derivative **24** (1.0 equiv.) in pyridine was added 4,4'-dimethoxytrityl chloride (DMTrCl) (1.5 equiv.). After stirring at r.t. for 16 h, the reaction mixture was concentrated and purified by silica gel column chromatography with an addition of 0.1% of pyridine to the eluent to afford the DMTr-protected compound **25** as a white foam.

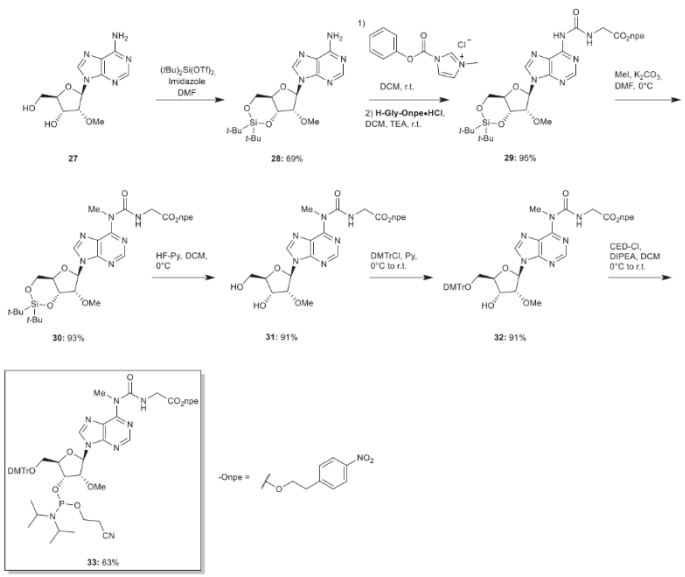
25: Yield: 82%; R_f = 0.34 (1:1 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2953 (w), 1694 (m), 1607 (w), 1508 (m), 1461 (m), 1397 (w), 1344 (w), 1298 (w), 1245 (s), 1175 (m), 1166 (m), 1063 (m), 1032 (s), 962 (w), 832 (s), 756 (w), 726 (w); For major rotamer: ^{1}H NMR (400 MHz, acetone-d_6, 298 K) δ (ppm): 10.20 (br s, 1H), 7.74 (s, 1H), 7.60-7.50 (m, 2H), 7.46-7.38 (m, 4H), 7.32 (t, J = 7.8 Hz, 2H), 7.25-7.20 (m, 1H), 6.89 (d, J = 8.9 Hz, 4H), 5.95 (s, 1H), 4.45-4.21 (m, 1H), 4.17-3.90 (m, 4H), 3.87-3.66 (m, 8H), 3.59-3.37 (m, 5H), 2.88 (s, 3H), 1.00-0.91 (m, 2H), 0.02 (s, 9H) (some proton signals appeared too broad for an unequivocal assignment); $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K) δ (ppm): 163.9, 159.6, 156.8, 151.0, 146.1, 140.4, 136.9, 131.1, 129.1, 128.7, 127.5, 114.0, 88.5, 87.2, 84.1, 70.3, 64.5, 63.6, 58.7, 55.5, 46.3, 35.6, 18.4, -1.4 (some carbon signals appeared too broad for an unequivocal assignment); HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{39}H_{50}N_3O_{10}Si$ 746.3114; Found 746.3113.

General procedure for the synthesis of phosphoramidite 26:

A solution of 5'-DMTr-protected compound **25** (1.0 equiv.) and DIPEA (4.0 equiv.) in dry DCM was cooled to 0°C. To this solution was slowly added 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (2.5 equiv.) and the reaction mixture was stirred at r.t. for 5 h. The reaction was quenched by addition of aq. sat. $NaHCO_3$ and the crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. After purification by silica gel column chromatography with an addition of 0.1% pyridine and co-lyophilization from benzene the desired phosphoramidite **26** was obtained as a mixture of diastereoisomers and rotamers as a white foam.

26: Yield: 66%; R_f = 0.19 (1:1 DCM/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.0, 149.9, 149.8, 149.7; HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{48}H_{67}N_5O_{11}PSi$ 948.4338; Found 948.4333.

2.8 Nucleobase-modified 2'-methoxy N^6-carbamoyl adenosine phosphoramidite



Scheme S8. Synthesis of nucleobase-modified 2'-methoxy N^6-carbamoyl adenosine phosphoramidite.

General procedure for the synthesis of 28:

A suspension of 2′-OMe adenosine **27** (1.0 equiv.) in DMF was cooled to 0°C. Di-*tert*-butylsilyl *bis*(trifluoromethanesulfonate) (1.1 equiv.) was added dropwise and the mixture was stirred at r.t. for 30 min. To the reaction was added imidazole (2.5 equiv.) and the resulting solution was stirred at r.t. for 16 h. The reaction was concentrated under reduced pressure and the residue was redissolved in EtOAc and washed with water, aq. sat. $NaHCO_3$ solution and brine. The organic layer was dried ($MgSO_4$), filtered and concentrated. The crude was purified by silica gel column chromatography to yield **28** as a white foam.

28: Yield: 69%; R_f = 0.21 (100:1 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3319 (w), 3161 (m), 2933 (w), 1669 (s), 1600 (s), 1472 (s), 1367 (m), 1328 (m), 1260 (m), 1207 (m), 1133 (s), 1069 (s), 1027 (s), 966 (s), 907 (w), 829 (s), 739 (s), 653 (s); ^{1}H NMR (400 MHz, DMSO-d_6, 298 K) δ (ppm): 8.32 (s, 1H), 8.13 (s, 1H), 7.36 (s, 2H), 6.01 (s, 1H), 4.89 (dd, J = 9.0, 4.8 Hz, 1H), 4.34 (d, J = 4.8 Hz, 1H), 4.31 (d, J = 4.8 Hz, 1H), 4.03-3.94 (m, 2H), 3.54 (s, 3H), 1.08 (s, 9H), 1.01 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K) δ (ppm): 156.6, 153.2, 149.2, 140.4, 119.6, 88.6, 82.1, 76.9, 74.5, 67.3, 58.8, 27.7, 27.4, 22.7, 20.4; HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{19}H_{32}N_5O_4Si$ 422.2218; Found 422.2220.

General procedure for the synthesis of 29:

To a solution of silyl-protected 2′-OMe adenosine **28** (1.0 equiv.) in DCM was added 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (2.0 equiv.). The resulting suspension was stirred at r.t. for 16 h and then **H-aa-Onpe•HCl** (2.0 equiv.) together with NEt_3 (2.0 equiv.) was added. After stirring for 16 h, the reaction mixture was quenched by the addition of aq. sat. $NaHCO_3$ and the crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by silica gel column chromatography furnished the amino acid-modified adenosine derivative **29** as a white foam.

29: Yield: 95%; R_f = 0.23 (100:1 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3235 (w), 2934 (w), 2856 (m), 1747 (m), 1702 (s), 1587 (m), 1518 (s), 1467 (s), 1343 (s), 1257 (m), 1187 (s), 1138 (s), 1062 (s), 1014 (m), 825 (s), 736 (m), 651 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 9.99 (t, J = 5.6 Hz, 1H), 8.77 (s, 1H), 8.51 (s, 1H), 8.26 (s, 1H),

8.08 (d, J = 8.7 Hz, 2H), 7.38 (d, J = 8.7 Hz, 2H), 6.01 (s, 1H), 4.65 (dd, J = 9.6, 4.6 Hz, 1H), 4.50-4.38 (m, 3H), 4.27 (d, J = 4.6 Hz, 1H), 4.22-4.14 (m, 3H), 4.05 (dd, J = 9.6, 9.6 Hz, 1H), 3.69 (s, 3H), 3.09 (t, J = 6.6 Hz, 2H), 1.09 (s, 9H), 1.06 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.0, 154.3, 151.3, 150.3, 149.9, 146.9, 145.5, 142.1, 129.9, 123.8, 121.1, 89.7, 82.4, 77.3, 74.9, 67.6, 64.7, 59.5, 42.2, 35.0, 27.5, 27.2, 22.9, 20.5; HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{30}H_{42}N_7O_9Si$ 672.2808; Found 672.2808.

General procedure for the synthesis of 30:

The amino acid-modified 2′-OMe adenosine derivative **29** (1.0 equiv.) was dissolved in DMF and cooled to 0°C. To the solution were added K_2CO_3 (3.0 equiv.) together with MeI (2.0 equiv.) and the reaction was stirred at r.t. for 2 h. The reaction mixture was diluted with H_2O and extracted three times with EtOAc. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated. The obtained residue was purified by silica gel column chromatography to give **30** as a white foam.

30: Yield: 93%; R_f = 0.32 (1:1 *i*-Hexane/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2932 (w), 1857 (w), 1746 (m), 1682 (s), 1567 (s), 1517 (s), 1467 (s), 1343 (s), 1266 (m), 1192 (m), 1135 (s), 1062 (s), 1027 (s), 826 (s), 735 (m), 651 (s); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.95 (t, J = 5.4 Hz, 1H), 8.51 (s, 1H), 8.10 (d, J = 8.7 Hz, 2H), 7.98 (s, 1H), 7.37 (d, J = 8.7 Hz, 2H), 6.02 (s, 1H), 4.62-4.54 (m, 1H), 4.48 (dd, J = 9.2, 5.0 Hz, 1H), 4.43 (t, J = 6.6 Hz, 2H), 4.27-4.12 (m, 4H), 4.03 (d, J = 10.5 Hz, 1H), 3.98 (s, 3H), 3.69 (s, 3H), 3.08 (t, J = 6.6 Hz, 2H), 1.09 (s, 9H), 1.05 (s, 9H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.3, 156.2, 153.2, 151.7, 150.3, 147.0, 145.6, 139.6, 129.9, 123.8, 122.8, 89.7, 82.3, 77.3, 74.8, 67.6, 64.6, 59.5, 43.0, 35.0, 34.8, 27.5, 27.2, 22.9, 20.5; HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{31}H_{44}N_7O_9Si$ 686.2964; Found 686.2967.

General procedure for the synthesis of 31:

A solution of the modified 2′-OMe adenosine derivative **30** (1.0 equiv.) in DCM/pyridine (9:1 v/v) inside a plastic reaction vessel was cooled to 0°C. Subsequently, a solution of 70% HF-pyridine (5.0 equiv.) was slowly added and the reaction mixture was stirred at 0°C for 2 h. The reaction mixture was diluted with aq. sat. $NaHCO_3$ solution and extracted three times with DCM. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude product was purified by silica gel column chromatography to isolate the 3',5'-deprotected adenosine derivative **31** as a white foam.

31: Yield: 91%; R_f = 0.25 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 3201 (w), 2935 (w), 1743 (m), 1677 (m), 1568 (s), 1514 (s), 1464 (m), 1343 (s), 1268 (m), 1209 (m), 1110 (m), 1036 (m), 856 (m), 795 (s), 697 (m), 645 (m); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.85 (t, J = 5.4 Hz, 1H), 8.51 (s, 1H), 8.13 (d, J = 8.8 Hz, 2H), 8.01 (s, 1H), 7.39 (d, J = 8.8 Hz, 2H), 5.94-5.91 (m, 2H), 4.72 (dd, J = 7.4, 4.7 Hz, 1H), 4.60 (d, J = 4.7 Hz, 1H), 4.43 (t, J = 6.6 Hz, 2H), 4.37 (d, J = 1.0 Hz, 1H), 4.25-4.09 (m, 2H), 4.01 (s, 3H), 4.00-3.92 (m, 1H), 3.84-3.74 (m, 1H), 3.37 (s, 3H), 3.09 (t, J = 6.6 Hz, 2H), 2.69 (d, J = 1.7 Hz, 1H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.2, 156.0, 153.8, 151.2, 149.6, 147.0, 145.5, 141.6, 129.9, 123.9, 123.9, 89.7, 88.2, 82.3, 70.6, 64.7, 63.4, 59.0, 43.1, 35.0; HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{23}H_{28}N_7O_9$ 546.1943; Found 546.1943.

General procedure for the synthesis of 32:

The 3',5'-deprotected 2′-OMe adenosine derivative **31** (1.0 equiv.) was dissolved in pyridine and DMTrCl (1.5 equiv.) was added. The reaction mixture was stirred at r.t. for 16 h and afterwards the solvents were removed *in vacuo*. Purification by silica gel column chromatography with an addition of 0.1% pyridine afforded the DMTr-protected adenosine derivative **32** as a pale-yellow foam.

32: Yield: 91%; R_f = 0.45 (100:5 DCM/MeOH); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2358 (w), 1682 (m), 1568 (m) 1509 (s), 1463 (m), 1344 (s), 1249 (m), 1174 (m), 1033 (s), 701 (w), 667 (w); ^{1}H NMR (400 MHz, $CDCl_3$, 298 K) δ (ppm): 10.84 (t, J = 5.4 Hz, 1H), 8.48-8.42 (m, 2H), 8.10 (d, J = 8.6 Hz, 2H), 7.57 (d, J = 8.6 Hz, 2H), 7.48 (d, J = 7.4 Hz, 2H), 7.39-7.32 (m, 4H), 7.28 (t, J = 7.4 Hz, 2H), 7.24-7.20 (m, 1H), 6.91-6.78 (m, 4H), 6.27 (d, J = 4.0 Hz, 1H), 4.74-4.64 (m, 1H), 4.59 (t, J = 4.5 Hz, 1H), 4.43 (t, J = 6.4 Hz, 2H), 4.28-4.21 (m, 2H), 4.11 (d, J = 5.6 Hz, 2H), 3.92 (s, 3H), 3.77 (s, 6H), 3.53 (s, 3H), 3.45 (d, J = 4.6 Hz, 2H), 3.13 (t, J = 6.4 Hz, 2H); ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K) δ (ppm): 170.7, 159.6, 156.5, 153.7, 153.0, 150.7, 147.6, 147.4, 146.0, 141.7, 136.7, 131.0, 130.9, 129.0, 128.6, 127.6, 124.1, 123.2, 113.8, 87.6, 87.1, 84.9, 83.8, 70.6, 65.0, 64.3, 58.8, 55.5, 43.4, 35.3, 34.8; HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{44}H_{46}N_7O_{11}$ 848.3249; Found 848.3234.

General procedure for the synthesis of 33:

To a solution of 5'-DMTr-protected 2′-OMe adenosine derivative **32** (1.0 equiv.) in anhydrous DCM, *N,N*-diisopropylethylamine (DIPEA) (4.0 equiv.) was added. After cooling down to 0°C, 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (2.5 equiv.) was added dropwise and the reaction mixture was stirred

at r.t. for 5 h. After that, aq. sat. $NaHCO_3$ solution was added to the reaction mixture and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography with addition of 0.1% pyridine and co-lyophilized from benzene to afford the desired phosphoramidite **33** as a mixture of diastereoisomers and as a white foam.

33: Yield: 63%; R_f = 0.25 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.2, 149.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{53}H_{63}N_9O_{12}P$ 1048.4328; Found 1048.4309.

3. General information and instruments for oligonucleotides

3.1 Synthesis and purification of oligonucleotides

Phosphoramidites of canonical ribonucleosides (Bz-A-CE, Dmf-G-CE, Ac-C-CE and U-CE) were purchased from LinkTech and Sigma-Aldrich. Oligonucleotides (ONs) were synthesized on a 1 µmol scale using RNA SynBase™ CPG 1000/110 and High Load Glen UnySupport™ as solid supports for strands containing amino acid-modified carbamoyl adenosine and 5-(methyl)aminomethyl uridine derivatives, respectively, using an RNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with a standard phosphoramidite chemistry. ONs were synthesized in DMT-OFF mode using DCA as a deblocking agent in CH_2Cl_2, BTT or Activator 42® as activator in MeCN, Ac_2O as capping reagent in pyridine/THF and I_2 as oxidizer in pyridine/H_2O.

Deprotection of npe and teoc groups

For the deprotection of the *para*-nitrophenylethyl (npe) group in ONs containing amino acid-modified carbamoyl adenosine derivatives, the solid support beads were suspended in a 9:1 THF/DBU solution mixture (1 mL) and incubated at r.t. for 2 h.[9] After that, the supernatant was removed and the beads were washed with THF (3×1 mL).

For the deprotection of the 2-(trimethylsilyl)ethoxycarbonyl (teoc) group in ONs containing 5-(methyl)aminomethyl uridine derivatives, the solid support beads were suspended in a saturated solution of $ZnBr_2$ in 1:1 $MeNO_2$/IPA (1 mL) and incubated at r.t. overnight.[10] After that, the supernatant was removed and the beads were washed with 0.1 M EDTA in water (1 mL) and water (1 mL).

Cleavage from beads, deprotection of TBS groups and precipitation of the synthesized ON

The solid support beads were suspended in a 1:1 aqueous solution mixture (0.6 mL) of 30% NH_4OH and 40% $MeNH_2$. The suspension was heated at 65ºC (8 min for SynBase™ CPG 1000/110 and 60 min for High Load Glen UnySupport™). Subsequently, the supernatant was collected and the beads were washed with water (2×0.3 mL). The combined aqueous solutions were concentrated under reduced pressure using a SpeedVac concentrator. After that, the crude was dissolved in DMSO (100 µL) and triethylamine trihydrofluoride (125 µL) was added. The solution was heated at 65ºC for 1.5 h. Finally, the ON was precipitated by adding 3 M NaOAc in water (25 µL) and *n*-butanol (1 mL). The mixture was kept at -80ºC for 2 h and centrifuged at 4ºC for 1 h. The supernatant was removed and the white precipitate was lyophilized.

Purification of the synthesized ON by HPLC and desalting

The crude was purified by semi-preparative HPLC (1260 Infinity II Manual Preparative LC System from Agilent equipped with a G7114A detector) using a reverse-phase (RP) VP 250/10 Nucleodur 100-5 C18ec column from Macherey-Nagel. Buffers: A) 0.1 M AcOH/Et_3N in H_2O at pH 7 and B) 0.1 M AcOH/Et_3N in 80% (v/v) MeCN in H_2O. Gradient: 0-25% of B in 45 min. Flow rate = 5 mL·min^{-1}. The purified ON was analyzed by RP-HPLC (1260 Infinity II LC System from Agilent equipped with a G7165A detector) using an EC 250/4 Nucleodur 100-3 C18ec from Macherey-Nagel. Gradient: 0-30% or 0-40% of B in 45 min. Flow rate = 1 mL·min^{-1}. Finally, the purified ON was desalted using a C18 RP-cartridge from Waters.

Determination of the concentration and the mass of the synthesized ON

The absorbance of the synthesized ON in H_2O solution was measured using an IMPLEN NanoPhotometer® N60/N50 at 260 nm. The extinction coefficient of the single stranded ONs was calculated using the OligoAnalyzer Version 3.0 from Integrated DNA Technologies. For ONs incorporating non-canonical bases, the extinction coefficients were assumed to be identical to those containing only canonical counterparts.

The synthesized ON (2-3 µL) was desalted on a 0.025 µm VSWP filter (Millipore), co-crystallized in a 3-hydroxypicolinic acid matrix (HPA, 1 µL) and analyzed by MALDI-TOF mass spectrometry (negative mode).

3.2 Analysis of coupling and cleavage reactions by HPLC and MALDI-TOF mass spectrometry

The crudes of the coupling and cleavage reactions were analyzed by RP-HPLC using an EC 250/4 Nucleodur 100-3 C18ec column from Macherey-Nagel. Buffers: A) 0.1 M AcOH/Et_3N in H_2O at pH 7 and B) 0.1 M AcOH/Et_3N in 80% (v/v) MeCN in H_2O. Gradient: 0-40% of B in 45 min. Flow rate = 1 mL·min^{-1}. Injection: 20 µL (1 nmol). The same HPLC method was used for the purification of the products obtained in the coupling and cleavage reactions. The yields of the reactions were calculated by integration of the chromatographic peaks of the products and the use of the calibration curves of the corresponding canonical ONs (see Section 5). In order to simplify the calculations, we assumed that the formed products and the canonical oligonucleotides used for calibration featured identical extinction coefficients, which were calculated for single stranded RNAs. It is expected that double strands and/or secondary structures are disrupted under the HPLC conditions used.

The crudes of the reactions and the isolated products (2-3 µL) were desalted on a 0.025 µm VSWP filter (Millipore), co-crystallized in a 3-hydroxypicolinic acid matrix (HPA, 1 µL) and analyzed by MALDI-TOF mass spectrometry (negative mode).

3.3 Coupling of amino acids and peptides to ONs anchored to the solid support beads

Oligonucleotides (ONs) were synthesized on a 4 µmol scale using the High Load Glen UnySupport™ for strands containing glycine-modified carbamoyl adenosine and 5-valine-methylaminomethyl uridine derivatives using an RNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with a standard phosphoramidite chemistry. The npe and teoc protecting groups were removed as described in Section 3.1 and the solid support beads were dried using a SpeedVac concentrator.

The solid support beads (1 µmol) in an Eppendorf tube were washed with dry DMF (0.3 mL). In a separate Eppendorf tube, Boc-protected amino acid (for altering of the mnm^5U derivatives), npe-protected amino acid (for altering of the m^6g^6A derivatives) or protected peptide (100 µmol), DMTMM•BF_4 (100 µmol) as activator and dry DIPEA (200 µmol) were dissolved in dry DMF (0.6 mL). Subsequently, the amino acid or peptide solution was added to the solid support beads and the reaction was incubated in an orbital shaker at r.t. for 1 h. The suspension was centrifuged and the supernatant was removed. The solid support beads were washed with dry DMF (2×0.3 mL) and dry MeCN (2×0.3 mL). Finally, the beads were dried using a SpeedVac concentrator.

For the deprotection of the *tert*-butyloxycarbonyl (Boc) group in ONs after the coupling of a Boc-protected amino acid or peptide, the solid support beads were suspended in a 1:1 TFA/CH_2Cl_2 solution mixture (0.5 mL) and incubated for 5 min at r.t.[11] After that, the supernatant was removed and the solid support beads were washed with CH_2Cl_2 (2×0.5 mL). The deprotection of the npe-protected adenosine derivatives was performed as described in Section 3.1.

The ONs containing 5-peptide-methylaminomethyl uridine derivatives were cleaved from the solid support beads using a 1:1 aqueous solution mixture (0.6 mL) of 30% NH_4OH and 40% $MeNH_2$ at 65ºC for 60 min. The ONs containing peptide-modified carbamoyl adenosine derivatives were cleaved from the solid support beads using a 30% NH_4OH aqueous solution (0.6 mL) at r.t. overnight. The following work-up and purification steps were identical to those described in Section 3.1. Based on HPLC analyses, we calculated that the coupling reaction using the solid support beads and DMTMM•BF_4 as activator proceeded in an extent larger than 70%.

4. Synthesized oligonucleotides using a DNA/RNA automated synthesizer

4.1 Canonical oligonucleotides (CON)

RNA sequences:

CON1; 5'-AAU CGC U-3'

CON2; 5'-GUA CAG CGA UU-3'

CON3; 5'-GUA CAG CGA UUA AUC GCU-3'

CON4; 5'-AmAmUm CmGmCm Um-3'

CON5; 5'-GmUmCm AmGmUm AmCmAm GmCmGm AmUmUm-3'

CON6; 5'-GmUmCm AmGmUm AmCmAm GmCmGm AmUmUm AmAmUm CmGmCm Um-3'

Table S1. HPLC retention times (0-30% of B in 45 min) and MALDI-TOF mass spectrometric analysis (negative mode) of canonical oligonucleotides.

Strand	t_R (min)	*m/z* calcd. for [*M*-H]$^-$	found
CON1	23.6	2162.3	2162.0
CON2	23.1	3487.5	3486.9
CON3	23.9	5712.8	5711.7

Table S2. HPLC retention times (0-40% of B in 45 min) and MALDI-TOF mass spectrometric analysis (negative mode) of canonical oligonucleotides.

Strand	t_R (min)	*m/z* calcd. for [*M*-H]$^-$	found
CON4	23.3	2261.6	2260.1
CON5	18.8	4772.7	4772.8
CON6	18.6	6998.0	6995.1

The sequences of **CON1**-**6** are similar to those of the modified ONs used in the coupling reactions. These canonical ONs were used for the development of HPLC calibration curves in Section 5.

4.2 Donor oligonucleotides (ON1) with a complementary sequence

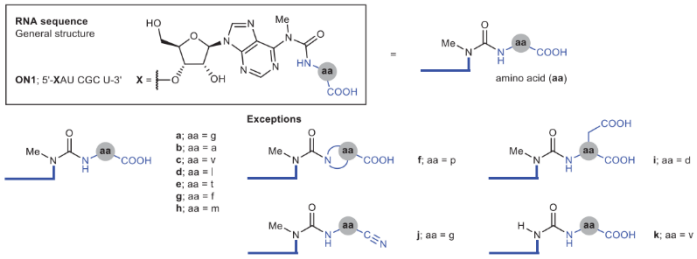


Figure S2. RNA sequence and general structure of amino acid-modified carbamoyl adenosine derivatives.

Other RNA donor strands with longer sequences:

ON1l: 5'-**X**CU AUU GAG U-3'; **X** = $m^6\underline{v}^6A$

ON1m: 5'-**X**^{1}AU CGC UGU ACC CUA UUG AGU **X**^{2}-3'; **X**^{1} = $m^6\underline{v}^6A$; **X**^{2} = $m^6\underline{g}^6A$

ON1n: 5'-**X**AU CGC UGU AC-3'; **X** = $m^6\underline{v}^6A$

ON1o: 5'-**X**AmUm CmGmCm Um-3'; **X** = $m^6\underline{g}^6Am$

ON1p: 5'-**X**AmUm CmGm-3'; **X** = $m^6\underline{g}^6Am$

ON1q: 5'-**X**AmUm-3'; **X** = $m^6\underline{g}^6Am$

Table S3. HPLC retention times (0-40% of B in 45 min) and MALDI-TOF mass spectrometric analysis (negative mode) of **ON1**.

Strand	t_R (min)	*m/z* calcd. for [*M*-H]$^-$	found
ON1a; **X** = $m^6\underline{g}^6A$	18.8	2277.4	2278.4
ON1b; **X** = $m^6\underline{a}^6A$	20.2	2291.4	2290.0
ON1c; **X** = $m^6\underline{v}^6A$	22.2	2319.4	2317.8
ON1d; **X** = $m^6\underline{l}^6A$	24.3	2333.4	2331.6
ON1e; **X** = $m^6\underline{t}^6A$	18.9	2321.4	2320.0
ON1f; **X** = $m^6\underline{p}^6A$	18.0	2317.4	2316.8
ON1g; **X** = $m^6\underline{f}^6A$	24.5	2368.6	2365.4
ON1h; **X** = $m^6\underline{m}^6A$	23.2	2351.4	2350.4
ON1i; **X** = $m^6\underline{d}^6A$	17.2	2335.4	2334.3
ON1j; **X** = $m^6\underline{g}^6A$ (amino nitrile)	21.2	2258.4	2258.5
ON1k; **X** = $\underline{v}^6A$ (non-methylated)	20.6	2305.4	2302.2
ON1l; **X** = $m^6\underline{v}^6A$	22.3	3300.5	3301.1
ON1m; **X**^{1} = $m^6\underline{v}^6A$ and **X**^{2} = $m^6\underline{g}^6A$	23.1	7231.0	7233.7
ON1n; **X** = $m^6\underline{v}^6A$	23.2	3604.6	3603.4
ON1o; **X** = $m^6\underline{g}^6Am$	23.8	2375.5	2374.4
ON1p; **X** = $m^6\underline{g}^6Am$	23.6	1736.4	1735.1
ON1q; **X** = $m^6\underline{g}^6Am$	23.1	1058.2	1058.2

4.3 Acceptor oligonucleotides (ON2) with a complementary sequence

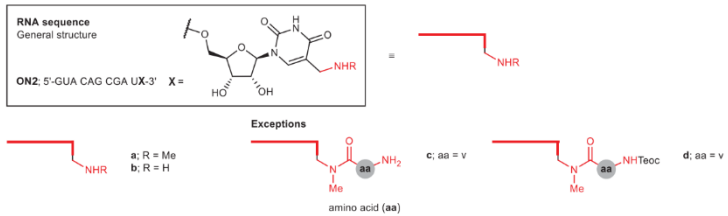


Figure S3. RNA sequence and general structure of (methyl)aminomethyl uridine derivatives.

Other RNA acceptor strands with longer sequences:

ON2e 5'-GUA CAG CGA U**X**^{1}A CUC AAU AG**X**^{2}-3'; **X**^{1} = $\underline{g}mnm^5U$; **X**^{2} = nm^5U

ON2f: 5'-GUA CAG CGA U**X**A CUC AAU AGG-3'; **X** = $\underline{v}mnm^5U$

ON2g: 5'-GmUmAm CmAmGm CmGmAm UmX-3'; **X** = mnm^5U

ON2h: 5'-GmUmCm AmGmUm AmCmAm GmCmGm AmUm**X**-3'; **X** = mnm^5Um

Table S4. HPLC retention times (0-40% of B in 45 min) and MALDI-TOF mass spectrometric analysis (negative mode) of **ON2**.

Strand	t_R (min)	*m/z* calcd. for [*M*-H]$^-$	found
ON2a; **X** = mnm^5U	17.4	3530.5	3529.7
ON2b; **X** = nm^5U	17.8	3516.5	3515.9
ON2c; **X** = $\underline{v}mnm^5U$	18.6	3629.6	3627.2
ON2d; **X** = Teoc-$\underline{v}mnm^5U$	37.7	3773.7	3776.9
ON2e; **X**^{1} = $\underline{g}mnm^5U$ and **X**^{2} = nm^5U	18.7	6806.0	6806.4
ON2f; **X** = $\underline{v}mnm^5U$	19.9	6858.0	6857.7
ON2g; **X** = mnm^5U	23.0	3670.5	3670.4
ON2h; **X** = mnm^5Um	24.2	5025.9	5026.0

4.4 Donor oligonucleotides with non-complementary sequences

RNA sequences that are not fully complementary to the acceptor **ON2**:

ON1r: 5'-**X**AU AGC U-3'; **X** = m^6g^6A (one mismatch marked in red)

ON1s: 5'-**X**AG CCC U-3'; **X** = m^6g^6A (two mismatches marked in red)

Table S5. HPLC retention times (0-40% of B in 45 min) and MALDI-TOF mass spectrometric analysis (negative mode) of **ON1r** and **ON1s**.

Strand	t_R (min)	*m/z* calcd. for [*M*-H]$^-$	found
ON1r; **X** = m^6g^6A	20.0	2301.4	2301.2
ON1s; **X** = m^6g^6A	19.5	2276.4	2275.7

5. HPLC calibration curves using canonical oligonucleotides (CON1-6) and hairpin-type intermediate (ON3a)

Canonical oligonucleotides, **CON1-6**, and hairpin-type intermediate, **ON3a**, were used for the development of HPLC calibration curves. Separate stock solutions of **CON1-6** and **ON3a** were prepared in water (100 μM). Separate standard solutions containing 1.2; 1.0; 0.8; 0.6; 0.4; 0.2 and 0.1 nmol of **CON1-6** and **ON3a** were prepared in a final volume of 20 μL. The standard solutions were injected in an analytical HPLC equipped with a C18 column and using buffers A and B (gradient: 0-30% or 0-40% of B in 45 min; flow rate = 1 mL·min^{-1}). The absorbance was monitored at 260 nm and the areas of the chromatographic peaks were determined by integration of the HPL-chromatograms. The plot of the chromatographic area (a.u.) versus the amount (nmol) of each oligonucleotide followed a linear relationship.

Calibration curve of **CON1**

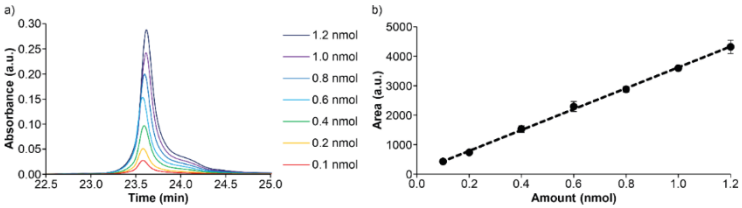


Figure S4. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON1**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Calibration curve of **CON2**

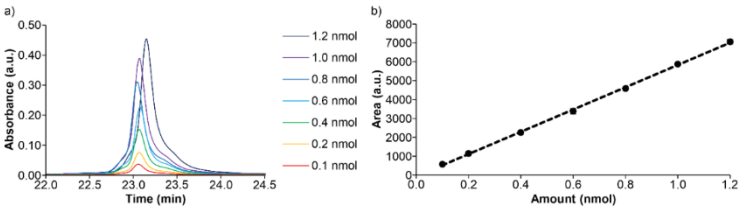


Figure S5. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON2**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Calibration curve of **CON3**

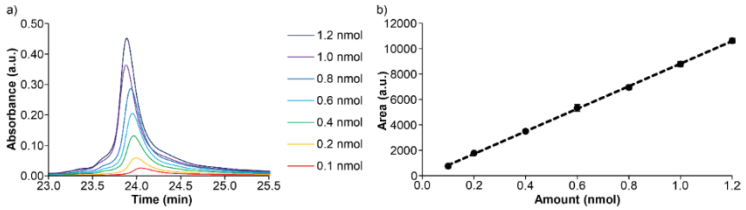


Figure S6. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON3**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Calibration curve of **ON3a**

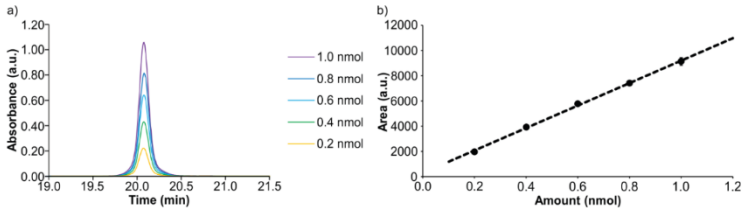


Figure S7. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **ON3a**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

The results of the calibration curves of **CON3** (canonical oligonucleotide) and **ON3a** (hairpin-type intermediate) were very similar (Table S6).

Calibration curve of **CON4**

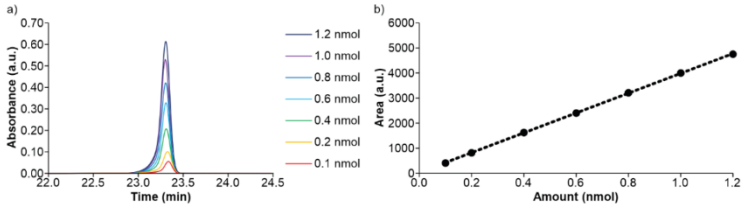


Figure S8. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON4**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Calibration curve of **CON5**

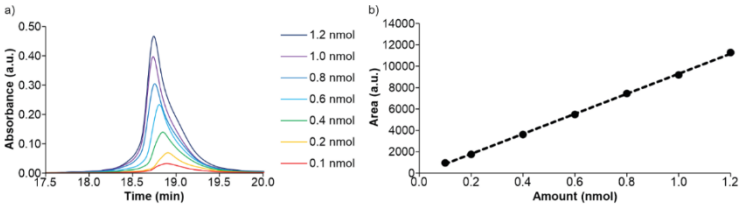


Figure S9. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON5**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Calibration curve of **CON6**

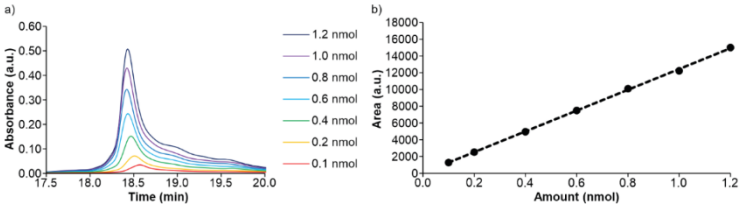


Figure S10. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON6**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Table S6. Calibration curves (y = mx + n) obtained by HPLC analyses of **CON1**-**6** and **ON3a** and calculated extinction coefficients of **CON1**-**6** using the OligoAnalyzer Version 3.0 from Integrated DNA Technologies.

Strand	Slope, m (nmol^{-1})	Intercept, n	r^2	ε (M^{-1}·cm^{-1})
CON1	3534.2	82.3	0.9989	65500
CON2	5903.7	-73.3	0.9994	107200
CON3	8885.4	-64.6	0.9997	170700
ON3a	8890.5	299.2	0.9986	170700[a]
CON4	3952.4	32.36	0.9999	68800
CON5	9376.2	-83.62	0.9995	153800
CON6	12405.0	41.49	0.9996	221900

[a] In order to simplify the calculations, the extinction coefficient of **ON3a** was assumed to be identical to that of **CON3**.

6. Coupling reactions between donor and acceptor oligonucleotides, ON1 and ON2

Stock solutions of pH buffer (400 mM), NaCl (1 M) and activator (500 mM, Figure S11) were prepared in water. Subsequently, equimolar amounts of **ON1** and **ON2** (3-5 nmol) were annealed at 95ºC for 4 min in water containing NaCl (half of the volume required for the reaction). Finally, buffer, NaCl, activator solutions and water were added to the ONs' solution and the reaction was incubated in a ThermoMixer at 25°C for 24 h.

Concentration of the components in the reaction mixture: 50 µM of **ON1**, 50 µM of **ON2**, 100 mM of buffer, 100 mM of NaCl and 50 mM of activator (see figure footnotes for details).

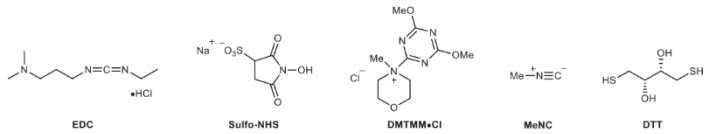


Figure S11. Activators of carboxylic acid and nitrile groups.

The crudes of the reactions (20 µL, 1 nmol) were analyzed as indicated in Section 3.2.

6.1 Control experiments

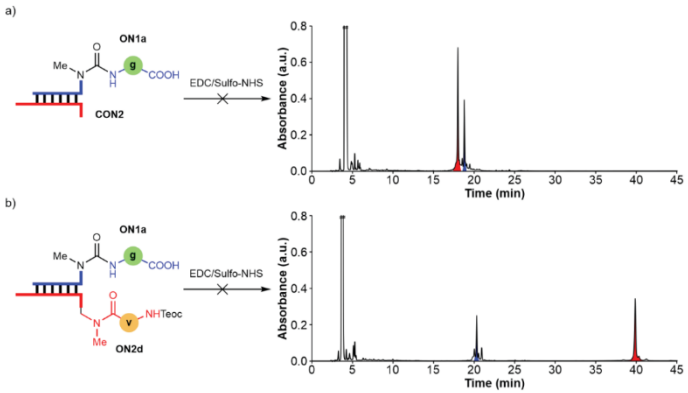


Figure S12. HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g^{6}A with: a) **CON2** (complementary canonical ON) and b) **ON2d**; **X** = Teoc-vmnm^{5}U in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Control reactions using the donor strand **1a** and the RNA strand lacking the mnm group on the 3'-terminal uridine base **CON2** or the protected 3'-vmnm^{5}U-RNA-5' acceptor strand **ON2d** did not provide noticeable evidence for the formation of the corresponding hairpin-type intermediate products.

6.2 Screening of activators using ON1a (m^{6}g^{6}A) and ON2a (mnm^{5}U)

Scheme S9. Coupling of **ON1a**; **X** = m^{6}g^{6}A with **ON2a**. The formed peptide bond is marked in purple.

MES buffer at pH 6 (adjusted with NaOH)

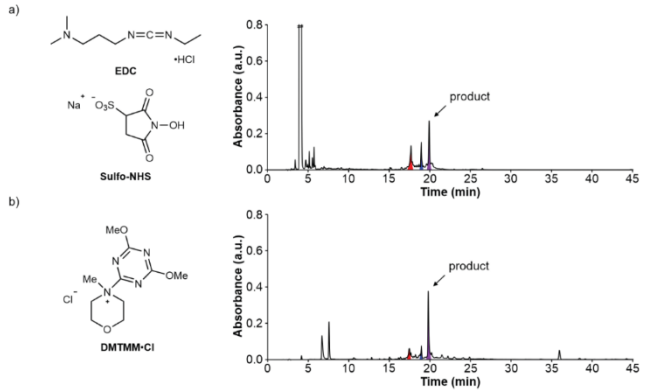


Figure S13. HPL-chromatograms of the reactions of **ON1a**; **X** = m^6g^6A with **ON2a** using: a) EDC/Sulfo-NHS and b) DMTMM•Cl as activators.

Table S7. Results obtained in the coupling reactions of **ON1a**; **X** = m^6g^6A with **ON2a** (average of, at least, two experiments).

Activators	pH	Time (h)	Average Yield ± Error (%)[a]
EDC/Sulfo-NHS	6	24	16±4
DMTMM•Cl	6	24	33±2

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

MOPS buffer at pH 7 (adjusted with NaOH)

Table S8. Results obtained in the coupling reactions of **ON1a**; **X** = m^6g^6A with **ON2a** (average of, at least, two experiments).

Activators	pH	Time (h)	Average Yield ± Error (%)[a]
EDC/Sulfo-NHS	7	24	20±2

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

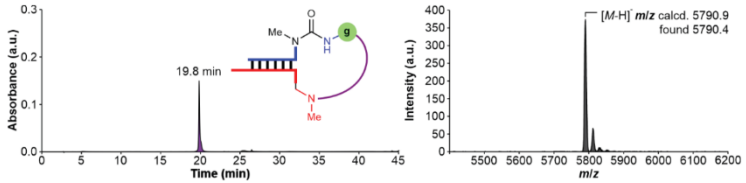


Figure S14. Left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product **ON3a**.

6.3 Screening of activators using ON1a (m^6g^6A) and ON2b (nm^5U)

Scheme S10. Coupling of **ON1a**; **X** = m^6g^6A with **ON2b**. The formed peptide bond is marked in purple.

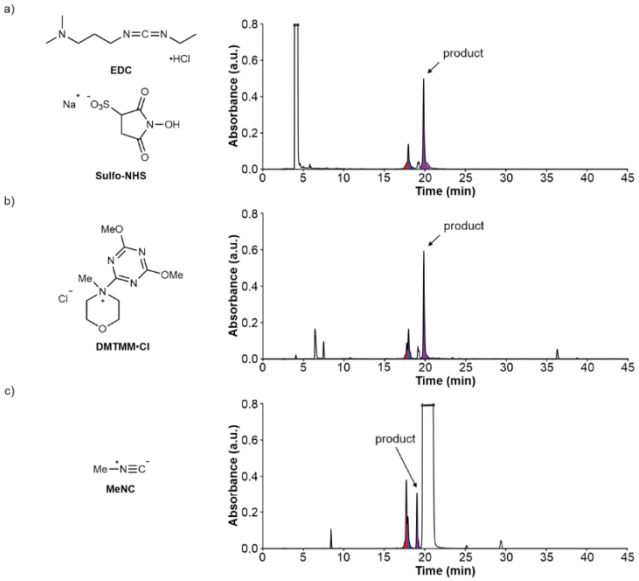


Figure S15. HPL-chromatograms of the reactions of **ON1a**; **X** = m^6g^6A with **ON2b** using: a) EDC/Sulfo-NHS; b) DMTMM•Cl and c) MeNC as activators. MES buffer (100 mM) at pH 6 in a) and b). DCI buffer (50 mM) at pH 6 in c).

Table S9. Results obtained in the coupling reactions of **ON1a**; **X** = m^6g^6A with **ON2b** (average of, at least, two experiments).

Activators	pH	Time (h)	Average Yield ± Error (%)[a]
EDC/Sulfo-NHS	6	24	64±2
DMTMM•Cl	6	24	66±2
MeNC	6	120	28±4

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

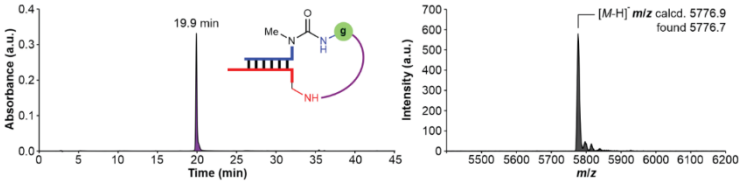


Figure S16. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product **ON3b**.

6.4 Screening of activators using ON1a (m^{6}g^{6}A) and ON2c (vmnm^{5}U)

Scheme S11. Coupling of **ON1a**; **X** = m^{6}g^{6}A with **ON2c**. The formed peptide bond is marked in purple.

Buffer at pH 6

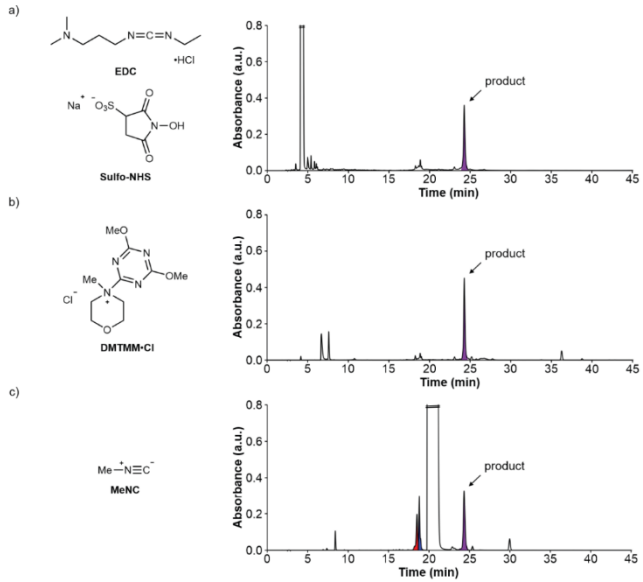


Figure S17. HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2c** using: a) EDC/Sulfo-NHS; b) DMTMM•Cl and c) MeNC as activators. MES buffer (100 mM) at pH 6 in a) and b). DCI buffer (50 mM) at pH 6 in c).

Table S10. Results obtained in the coupling reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2c** (average of, at least, two experiments).

Activators	pH	Time (h)	Average Yield ± Error (%)[a]
EDC/Sulfo-NHS	6	24	56±1
DMTMM•Cl	6	24	60±2
MeNC	6	120	50±5

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

MOPS buffer at pH 7 (adjusted with NaOH)

Table S11. Results obtained in the coupling reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2c** (average of, at least, two experiments).

Activators	pH	Time (h)	Average Yield ± Error (%)[a]
EDC/Sulfo-NHS	7	24	50±5
DMTMM•Cl	7	24	23±1

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

MOPS buffer at pH 8 (adjusted with NaOH)

Table S12. Results obtained in the coupling reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2c** (average of, at least, two experiments).

Activators	pH	Time (h)	Average Yield ± Error (%)[a]
EDC/Sulfo-NHS	8	24	34±1
DMTMM•Cl	8	24	5±2

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

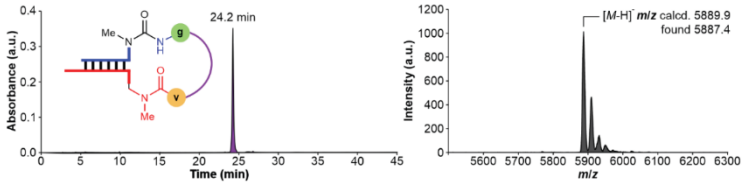


Figure S18. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product **ON3c**.

6.5 Coupling reactions of ON1j (m^{6}g^{6}A, amino nitrile) with ON2a-c

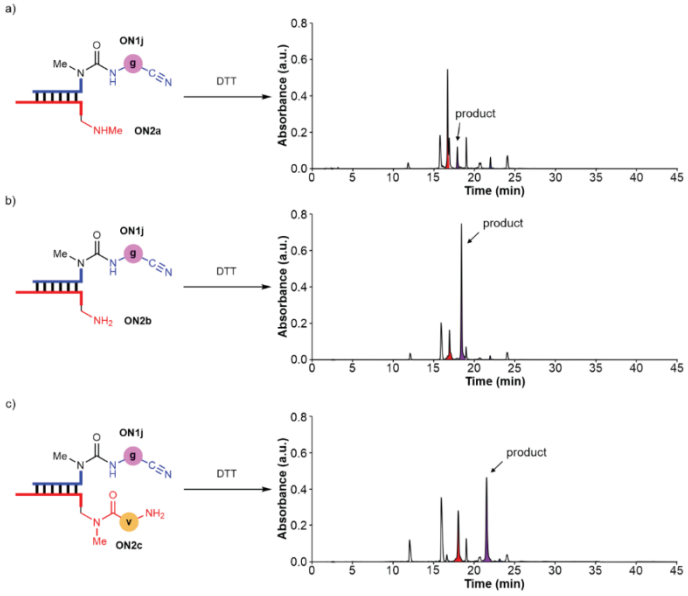


Figure S19. HPL-chromatograms of the reactions of **ON1j**; **X** = m^{6}g^{6}A (amino nitrile) with: a) **ON2a**; **X** = mnm^{5}U; b) **ON2b**; **X** = nm^{5}U and c) **ON2c**; **X** = vmnm^{5}U in boric acid buffer at pH 8 using DTT as activator.

Table S13. Results obtained in the coupling reactions of **ON1j**; **X** = $m^6\underline{g}^6A$ (amino nitrile) with **ON2a-c** using DTT as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1j; **X** = $m^6\underline{g}^6A$ (amino nitrile)	**ON2a**; **X** = mnm^5U	12±1
	ON2b; **X** = nm^5U	65±2
	ON2c; **X** = $\underline{v}mnm^5U$	42±1

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

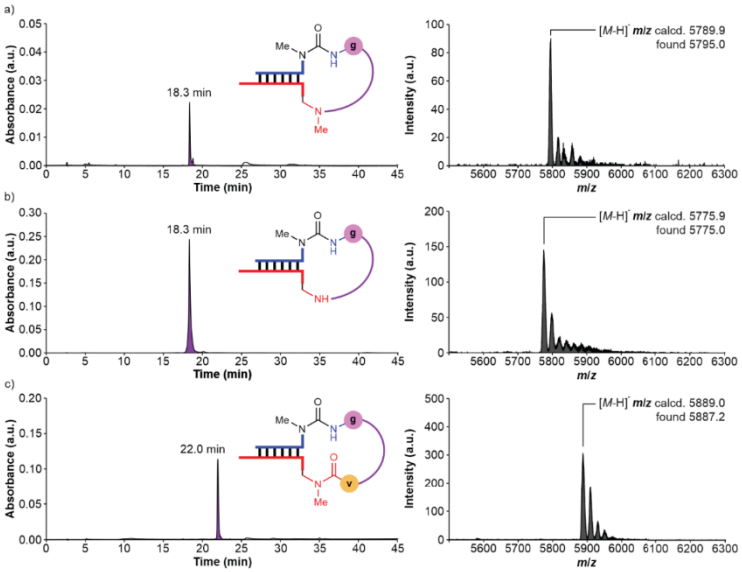


Figure S20. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON1j**; **X** = $m^6\underline{g}^6A$ (amino nitrile) with: a) **ON2a**; **X** = mnm^5U; b) **ON2b**; **X** = nm^5U and c) **ON2c**; **X** = $\underline{v}mnm^5U$.

6.6 Coupling reactions of ON1b-i (m^6aa^6A) with ON2a

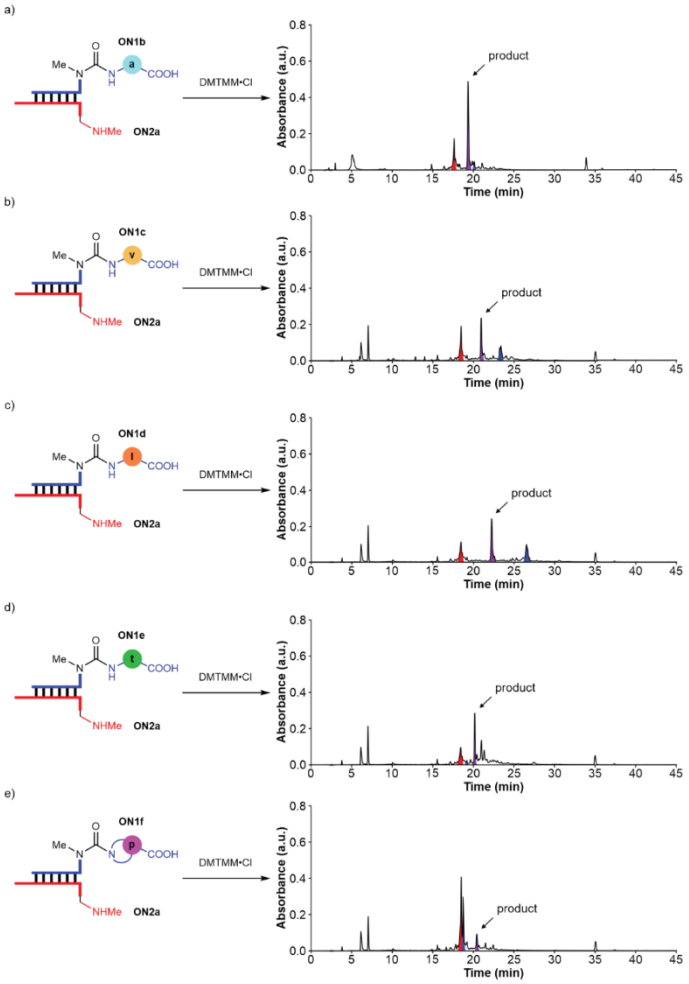


Figure S21. HPL-chromatograms of the reactions of **ON2a**; **X** = mnm^5U with: a) **ON1b**; **X** = $m^6\underline{a}^6A$; b) **ON1c**; **X** = $m^6\underline{v}^6A$; c) **ON1d**; **X** = $m^6\underline{l}^6A$; d) **ON1e**; **X** = $m^6\underline{t}^6A$ and e) **ON1f**; **X** = $m^6\underline{p}^6A$ in MES buffer at pH 6 using DMTMM•Cl as activator.

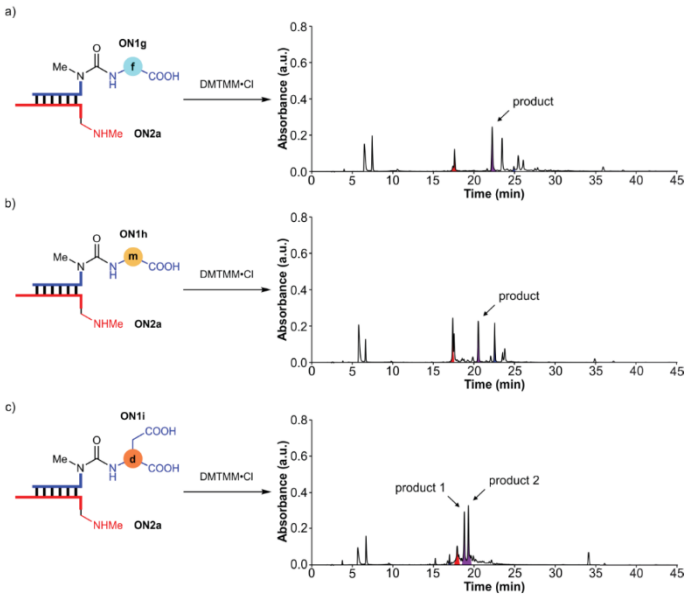


Figure S22. HPL-chromatograms of the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1g**; **X** = m^{6}f^{6}A; b) **ON1h**; **X** = m^{6}m^{6}A and c) **ON1i**; **X** = m^{6}d^{6}A in MES buffer at pH 6 using DMTMM•Cl as activator. For **ON1i**, the two peaks corresponded to the products of the reaction of the Asp α-COOH and of the side chain COOH. An assignment was not performed.

Table S14. Results obtained in the coupling reactions of **ON1b-i**; **X** = m^{6}aa^{6}A with **ON2a** using DMTMM•Cl as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1b; **X** = m^{6}a^{6}A	**ON2a**; **X** = mnm^{5}U	51±1
ON1c; **X** = m^{6}v^{6}A		21±1
ON1d; **X** = m^{6}t^{6}A		27±1
ON1e; **X** = m^{6}l^{6}A		18±5
ON1f; **X** = m^{6}p^{6}A		11±1
ON1g; **X** = m^{6}f^{6}A		27±1
ON1h; **X** = m^{6}m^{6}A		22±1
ON1i; **X** = m^{6}d^{6}A		28±4; 26±3[b]

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**. [b] For **ON1i**, the two yields describe the reaction of the Asp α-COOH and of the side chain COOH. An assignment was not performed.

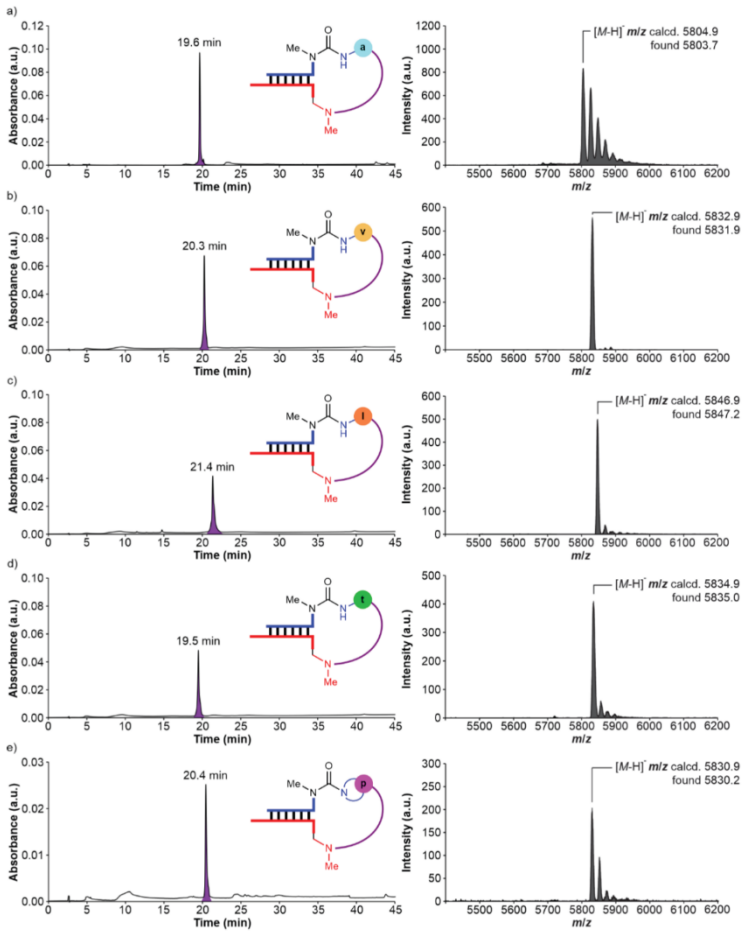


Figure S23. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1b**; **X** = m^{6}a^{6}A; b) **ON1c**; **X** = m^{6}v^{6}A; c) **ON1d**; **X** = m^{6}t^{6}A; d) **ON1e**; **X** = m^{6}l^{6}A and e) **ON1f**; **X** = m^{6}p^{6}A.

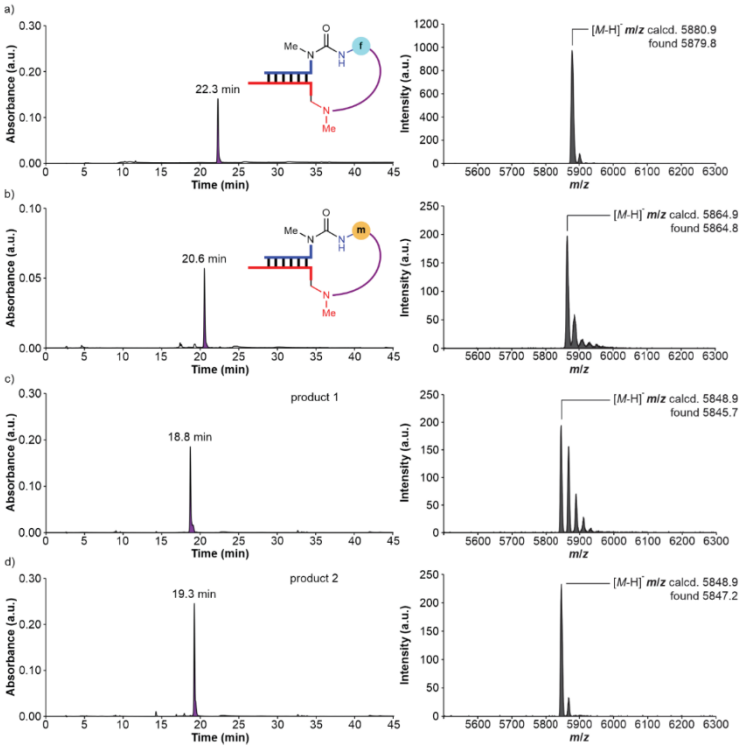


Figure S24. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1g**; **X** = m^{6}g^{6}A; b) **ON1h**; **X** = m^{6}m^{6}A; c) **ON1i**; **X** = m^{6}g^{6}A and d) **ON1i**; **X** = m^{6}g^{6}A.

6.7 Coupling reactions of ON1b-i (m^{6}aa^{6}A) with ON2c

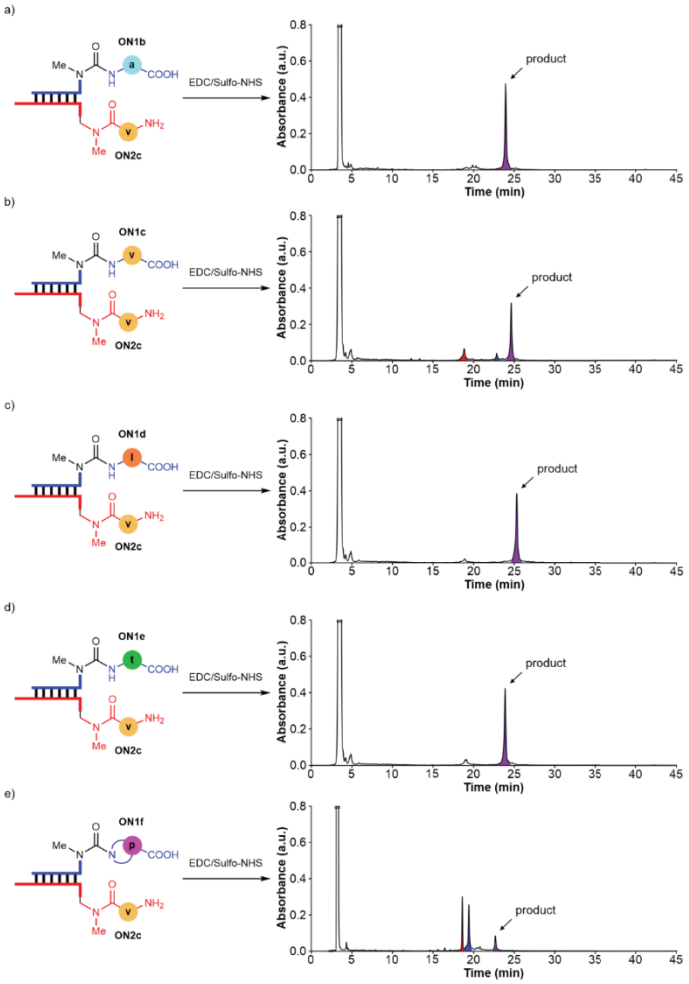


Figure S25. HPL-chromatograms of the reactions of **ON2c**; **X** = vmnm^{5}U with: a) **ON1b**; **X** = m^{6}a^{6}A; b) **ON1c**; **X** = m^{6}v^{6}A; c) **ON1d**; **X** = m^{6}l^{6}A; d) **ON1e**; **X** = m^{6}t^{6}A and e) **ON1f**; **X** = m^{6}p^{6}A in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

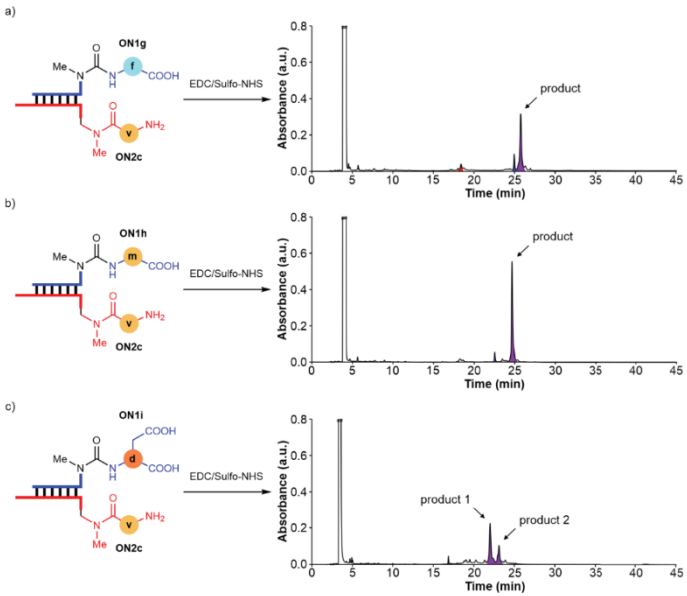


Figure S26. HPL-chromatograms of the reactions of **ON2c**; **X** = v̲mnm^{5}U with: a) **ON1g**; **X** = m^{6}f̲^{6}A; b) **ON1h**; **X** = m^{6}m̲^{6}A and c) **ON1i**; **X** = m^{6}d̲^{6}A in MES buffer at pH 6 using EDC/Sulfo-NHS as activator. For **ON1i**, the two peaks corresponded to the products of the reaction of the Asp *α*-COOH and of the side chain COOH. An assignment was not performed.

Table S15. Results obtained in the coupling reactions of **ON1b-i**; **X** = m^{6}a̲a̲^{6}A with **ON2c** using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1b; **X** = m^{6}a̲^{6}A	**ON2c**; **X** = v̲mnm^{5}U	76±2
ON1c; **X** = m^{6}v̲^{6}A		54±1
ON1d; **X** = m^{6}i̲^{6}A		77±1
ON1e; **X** = m^{6}t̲^{6}A		77±1
ON1f; **X** = m^{6}p̲^{6}A		18±4 (55±5)[b]
ON1g; **X** = m^{6}f̲^{6}A		50±1
ON1h; **X** = m^{6}m̲^{6}A		70±2
ON1i; **X** = m^{6}d̲^{6}A		34±1; 17±2[c]

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**. [b] Using DMTMM•Cl as activator. [c] For **ON1i**, the two yields describe the reaction of the Asp *α*-COOH and of the side chain COOH. An assignment was not performed.

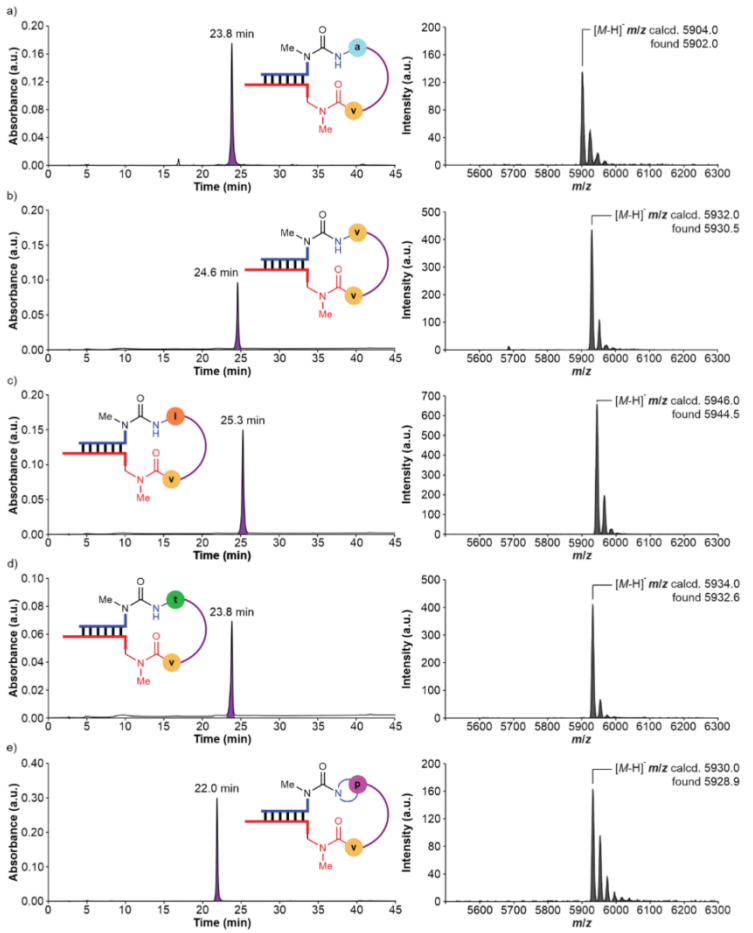


Figure S27. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2c**; **X** = v̲mnm^{5}U with: a) **ON1b**; **X** = m^{6}a̲^{6}A; b) **ON1c**; **X** = m^{6}v̲^{6}A; c) **ON1d**; **X** = m^{6}i̲^{6}A; d) **ON1e**; **X** = m^{6}t̲^{6}A and e) **ON1f**; **X** = m^{6}p̲^{6}A.

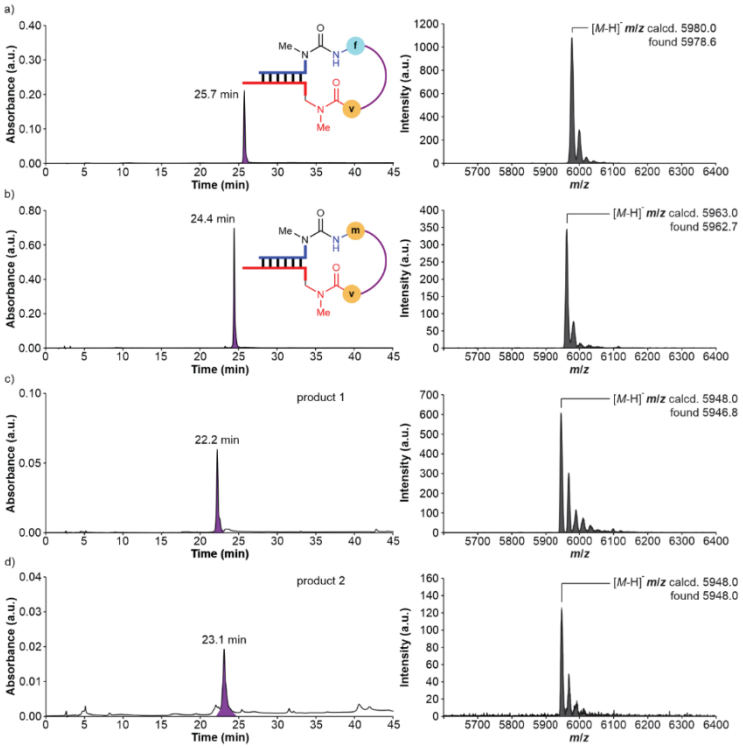


Figure S28. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2c**; **X** = $\underline{v}$mnm^{5}U with: a) **ON1g**; **X** = m^{6}f^{6}A; b) **ON1h**; **X** = m^6$\underline{m}$^{6}A; c) **ON1i**; **X** = m^{6}d^{6}A and d) **ON1i**; **X** = m^6$\underline{d}$^{6}A.

7. Synthesized peptide-oligonucleotides using solid support beads

7.1 Donor peptide-oligonucleotides with a complementary sequence

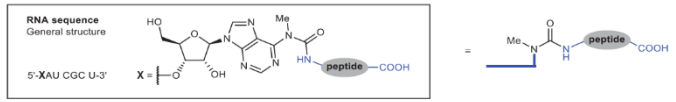


Figure S29. RNA sequence and general structure of peptide-modified carbamoyl adenosine derivatives.

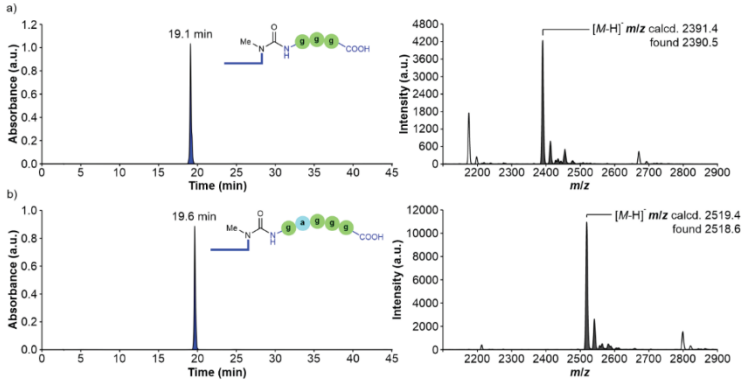


Figure S30. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the synthesized peptide-oligonucleotides: a) 5'-m^6($\underline{ggg}$)^{6}A-RNA-3' and b) 5'-m^6($\underline{gagag}$)^{6}A-RNA-3'.

7.2 Acceptor peptide-oligonucleotides with a complementary sequence

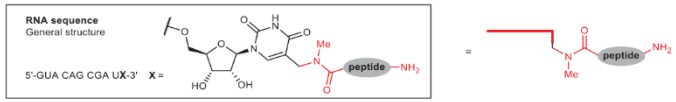


Figure S31. RNA sequence and general structure of peptide-modified methylaminomethyl uridine derivatives.

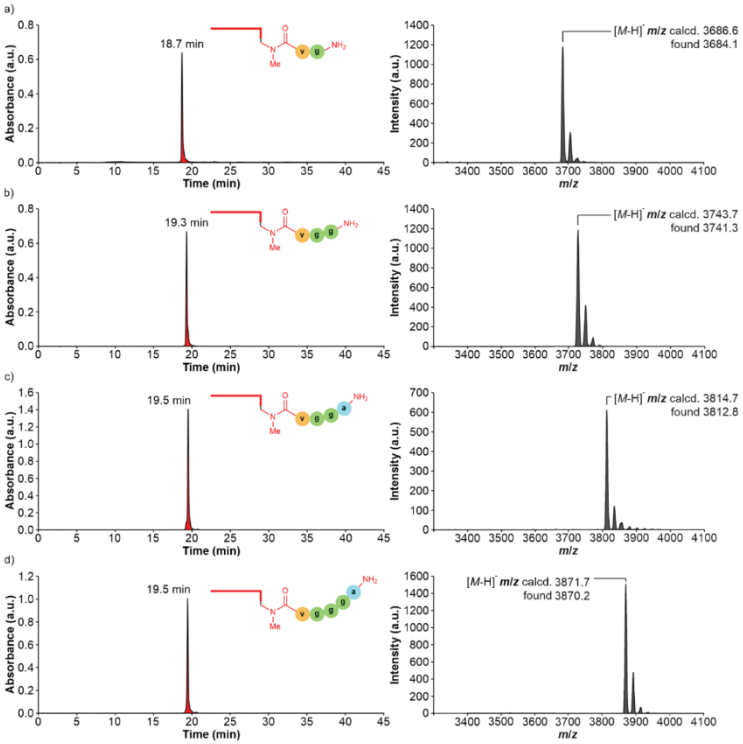


Figure S32. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the synthesized peptide-oligonucleotides: a) 3'-gvmnm^{5}U-RNA-5'; b) 3'-ggvmnm^{5}U-RNA-5'; c) 3'-aggvmnm^{5}U-RNA-5' and d) 3'-agggvmnm^{5}U-RNA-5'.

5'-GmUmAm CmAmGm CmGmAm Um**X**-3'; **X** = aggmnm^{5}U

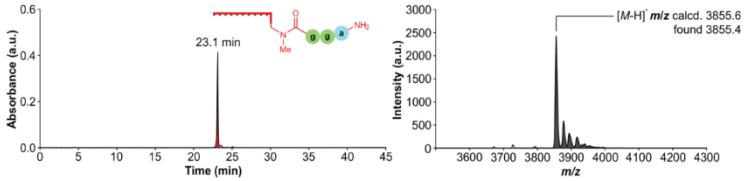


Figure S33. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the synthesized peptide-oligonucleotide 3'-aggmnm^{5}U-RNA-5' containing 2'-OMe nucleosides.

8. Coupling reactions between donor and acceptor peptide-oligonucleotides

The peptide coupling reactions were carried out under identical conditions to those described in Section 0.

8.1 Coupling reactions of donor peptide-oligonucleotides with ON2c

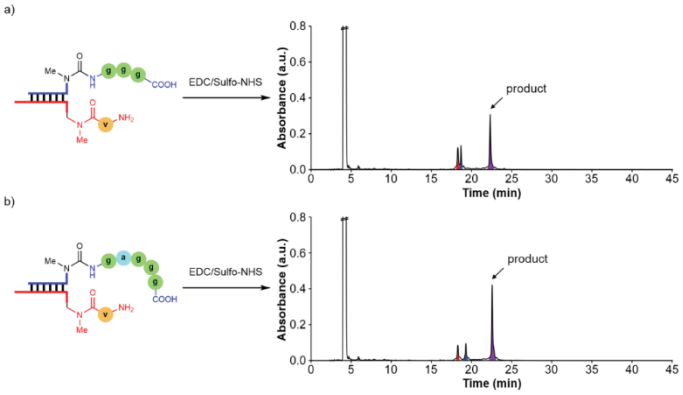


Figure S34. HPL-chromatograms of the reactions of **ON2c**; **X** = vmnm^{5}U with: a) 5'-m^6(ggg)^{6}A-RNA-3' and b) 5'-m^6(gaggg)^{6}A-RNA-3' in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S16. Results obtained in the coupling reactions of **ON2c**; **X** = v̲mnm^{5}U with peptide-modified donor oligonucleotides using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
5′-m^6(g̲gg)^{6}A-RNA-3′	**ON2c**; **X** = v̲mnm^{5}U	35±1
5′-m^6(g̲aggg)^{6}A-RNA-3′		43±1

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

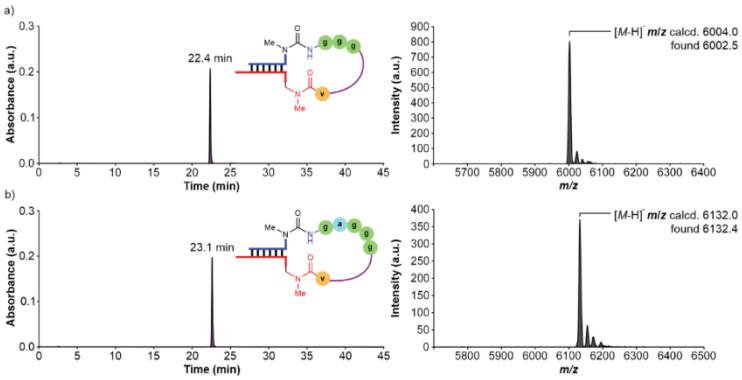


Figure S35. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2c**; **X** = v̲mnm^{5}U with: a) 5′-m^6(g̲gg)^{6}A-RNA-3′ and b) 5′-m^6(g̲aggg)^{6}A-RNA-3′.

8.2 Coupling reactions of ON1a (m^{6}g^{6}A) with acceptor peptide-oligonucleotides

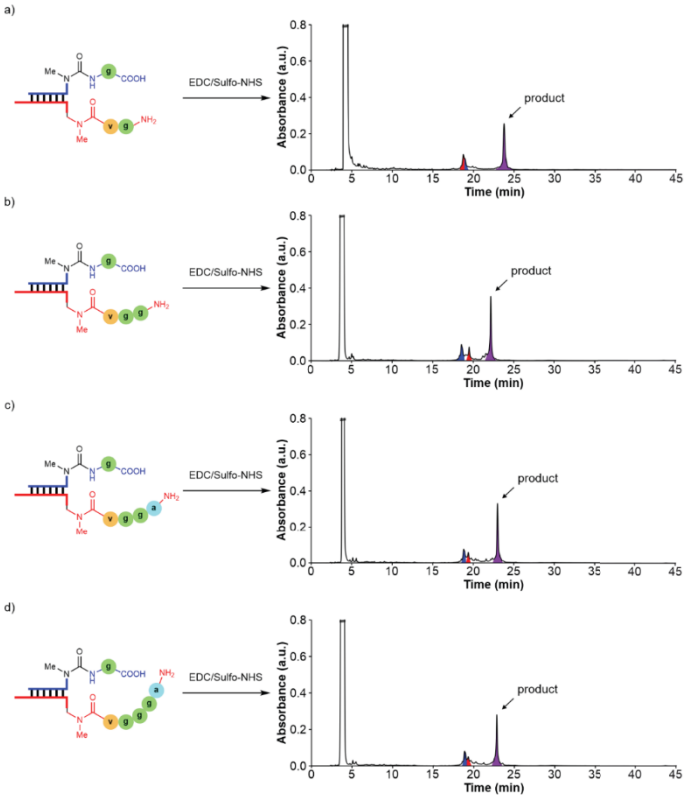


Figure S36. HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g̲^{6}A with: a) 3′-g̲vmnm^{5}U-RNA-5′; b) 3′-g̲gvmnm^{5}U-RNA-5′; c) 3′-aggvmnm^{5}U-RNA-5′ and d) 3′-agggvmnm^{5}U-RNA-5′ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S17. Results obtained in the coupling reactions of **ON1a**; **X** = m^{6}g̲^{6}A with peptide-modified acceptor oligonucleotides using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1a; **X** = m^{6}g̲^{6}A	3′-g̲vmnm^{5}U-RNA-5′	51±1
	3′-g̲gvmnm^{5}U-RNA-5′	46±4
	3′-aggvmnm^{5}U-RNA-5′	40±1
	3′-agggvmnm^{5}U-RNA-5′	40±3 (57±2)[b]

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**. [b] Using DMTMM•Cl as activator.

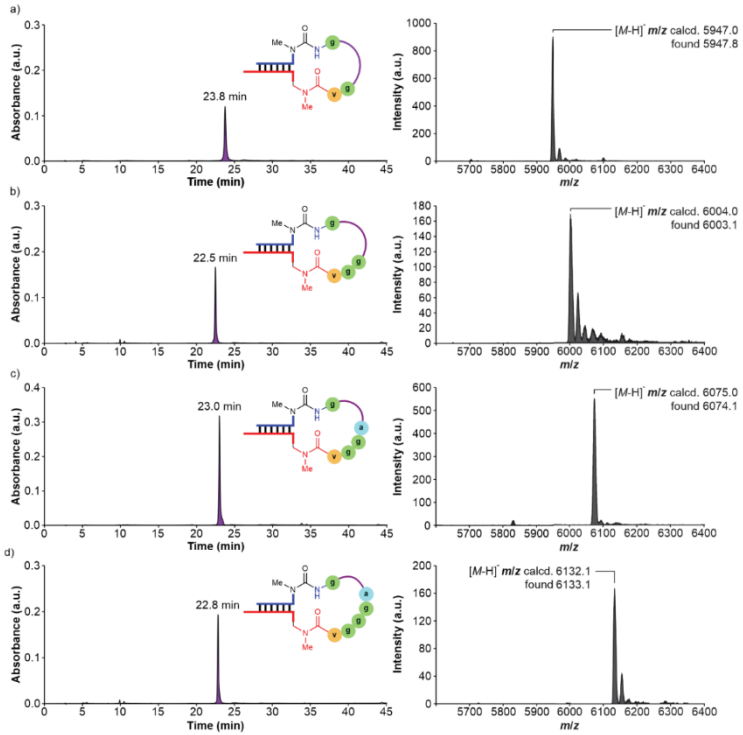


Figure S37. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON1a**; **X** = m^{6}g^{6}A with: a) 3'-gvmnm^{5}U-RNA-5'; b) 3'-ggvmnm^{5}U-RNA-5'; c) 3'-aggvmnm^{5}U-RNA-5' and d) 3'-agggvmnm^{5}U-RNA-5'.

8.3 Coupling reactions of donor and acceptor peptide-oligonucleotides

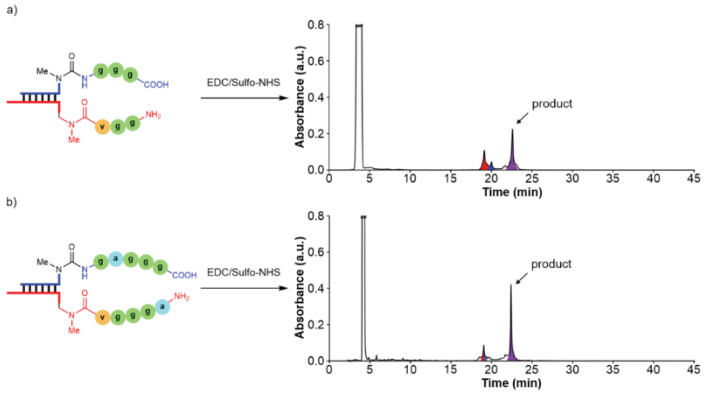


Figure S38. HPL-chromatograms of the reactions of: a) 5'-m^6(ggg)^{6}A-RNA-3' with 3'-ggvmnm^{5}U-RNA-5' and b) 5'-m^6(gaggg)^{6}A-RNA-3' with 3'-agggvmnm^{5}U-RNA-5' in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S18. Results obtained in the coupling reactions of peptide-modified donor and acceptor oligonucleotides using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
5'-m^6(ggg)^{6}A-RNA-3'	3'-ggvmnm^{5}U-RNA-5'	53±1
5'-m^6(gaggg)^{6}A-RNA-3'	3'-agggvmnm^{5}U-RNA-5'	56±3

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

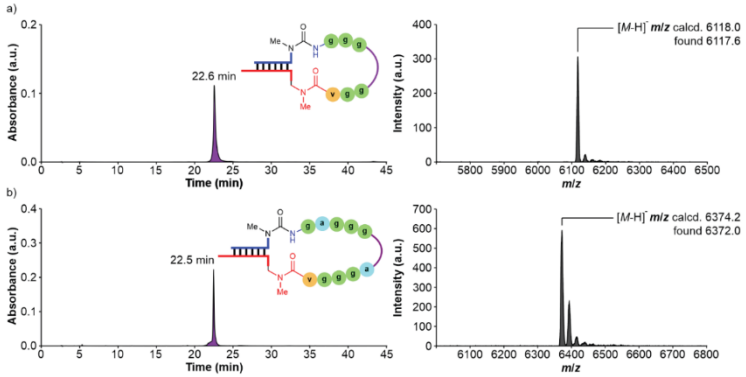


Figure S39. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of: a) 5'-m^6(ggg)^{6}A-RNA-3' with 3'-ggvmnm^{5}U-RNA-5' and b) 5'-m^6(gaggg)^{6}A-RNA-3' with 3'-agggvmnm^{5}U-RNA-5'.

9. Concentration of the product versus time in selected coupling reactions

The peptide coupling reactions were carried out under identical conditions to those described in Section 0 using DMTMM•Cl as activator.

The data (concentration of product *vs.* time) was fit to the corresponding theoretical kinetic model using the Parameter Estimation Module of COPASI software Version 4.29.[12] We introduced the theoretical kinetic model shown below:

Double strand → Hairpin-type Intermediate; k_{app}

The initial concentration of the double strand was refined as variable but constrained between 30 and 50 × 10^{-6} M. The fit of the data returned the rate constant value k_{app}. This fitting procedure is similar to that reported by others in the literature.[13]

In all cases, the fit of the experimental data was good based on the residual values, reported as sum of squared residuals (SSR), and the visual inspection of the curves.

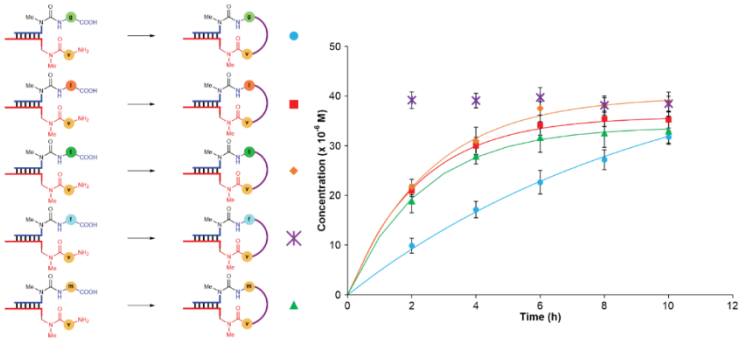


Figure S40. Concentration of the product (M) *vs.* time (h) in selected peptide coupling reactions using DMTMM•Cl as activator. Lines show fit of the data to the theoretical kinetic model. Error bars are the standard deviations.

Table S19. Calculated rate constant values for selected coupling reactions (average of, at least, two experiments).

Donor strand	Acceptor strand	k_{app} (h^{-1})[a]	SSR[b]
ON1a; **X** = $m^6\underline{g}^6A$		0.12±0.02	2.00 × 10^{-12}
ON1d; **X** = $m^6\underline{f}^6A$		0.42±0.02	8.20 × 10^{-13}
ON1e; **X** = $m^6\underline{f}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	0.39±0.04	2.50 × 10^{-12}
ON1g; **X** = $m^6\underline{f}^6A$		>1	n.d.
ON1h; **X** = $m^6\underline{m}^6A$		0.42±0.04	5.80 × 10^{-13}

[a] Errors are indicated as standard deviations. [b] SSR = Sum of squared residuals.

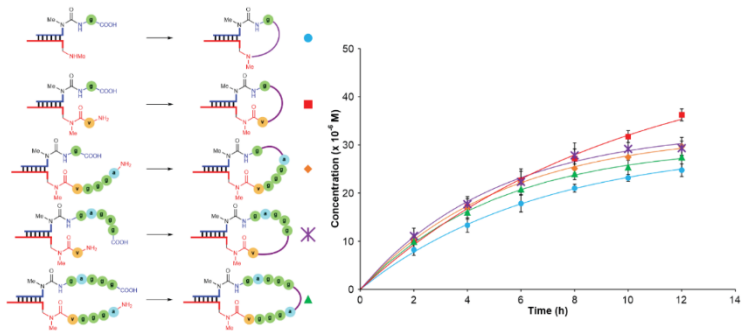


Figure S41. Concentration of the product (M) *vs.* time (h) in selected peptide coupling reactions using DMTMM•Cl as activator. Lines show fit of the data to the theoretical kinetic model. Error bars are the standard deviations.

Table S20. Calculated rate constant values for selected coupling reactions (average of, at least, two experiments).

Donor strand	Acceptor strand	k_{app} (h^{-1})[a]	SSR[b]
ON1a; **X** = $m^6\underline{g}^6A$	**ON2a**; **X** = mnm^5U	0.14±0.02	3.36 × 10^{-13}
ON1a; **X** = $m^6\underline{g}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	0.12±0.02	2.00 × 10^{-12}
ON1a; **X** = $m^6\underline{g}^6A$	3'-agggvmnm^5U-RNA-5'	0.18±0.02	5.59 × 10^{-13}
5'-m^6(gaggg)^{6}A-RNA-3'	**ON2c**; **X** = $\underline{v}mnm^5U$	0.19±0.02	3.82 × 10^{-12}
5'-m^6(gaggg)^{6}A-RNA-3'	3'-agggvmnm^5U-RNA-5'	0.19±0.01	4.45 × 10^{-13}

[a] Errors are indicated as standard deviations. [b] SSR = Sum of squared residuals.

10. Coupling reactions between oligonucleotides containing multiple donor or acceptor units

The peptide coupling reactions were carried out under identical conditions to those described in Section 0.

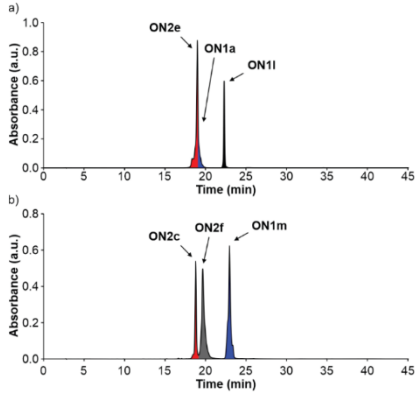


Figure S42. HPL-chromatograms of equimolar mixtures of: a) **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1l**; **X** = $m^6\underline{v}^6A$ and **ON2e**; $\mathbf{X^1}$ = $\underline{g}mnm^5U$ and $\mathbf{X^2}$ = nm^5U, and b) **ON1m**; $\mathbf{X^1}$ = $m^5\underline{v}^6A$ and $\mathbf{X^2}$ = $m^6\underline{g}^6A$, **ON2c**; **X** = $\underline{v}mnm^5U$ and **ON2f**; **X** = $\underline{v}mnm^5U$.

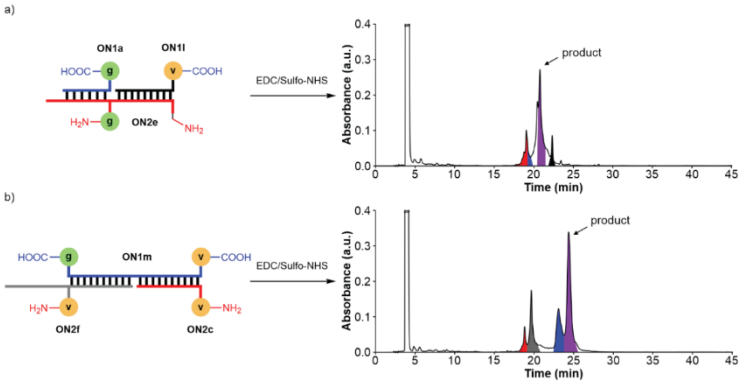


Figure S43. HPL-chromatograms of the reactions of: a) **ON1a**; **X** = m^6g^6A, **ON1l**; **X** = m^6v^6A and **ON2e**; $\mathbf{X^1}$ = $gmnm^5U$ and $\mathbf{X^2}$ = nm^5U and b) **ON1m**; $\mathbf{X^1}$ = m^6v^6A and $\mathbf{X^2}$ = m^6g^6A, **ON2c**; **X** = $vmnm^5U$ and **ON2f**; **X** = $vmnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator. The terminal functional groups of the ONs (urea and amide) are omitted for clarity.

Table S21. Results obtained in the coupling reactions of oligonucleotides containing multiple donor or acceptor units using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1a; **X** = m^6g^6A **ON1l**; **X** = m^6v^6A	**ON2e**; $\mathbf{X^1}$ = $gmnm^5U$ and $\mathbf{X^2}$ = nm^5U	35±2 (29±1)[b]
ON1m; $\mathbf{X^1}$ = m^6v^6A and $\mathbf{X^2}$ = m^6g^6A	**ON2c**; **X** = $vmnm^5U$ **ON2f**; **X** = $vmnm^5U$	35±3 (32±2)[b]

[a] Calculated yield from the chromatographic peak of the product based on the total area of the initial components (Figure S42). [b] Using DMTMM•Cl as activator.

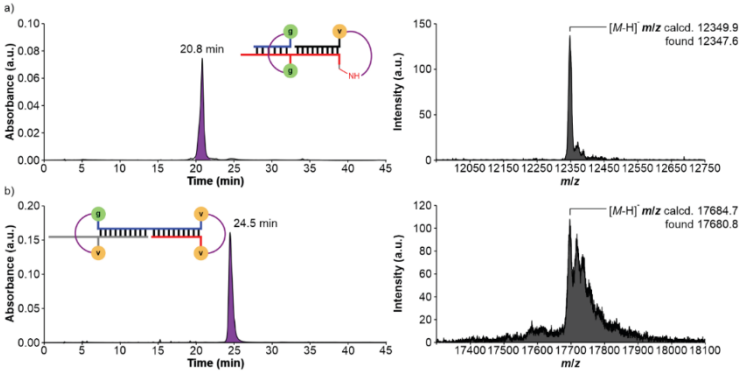


Figure S44. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of: a) **ON1a**; **X** = m^6g^6A, **ON1l**; **X** = m^6v^6A and **ON2e**; $\mathbf{X^1}$ = $gmnm^5U$ and $\mathbf{X^2}$ = nm^5U and b) **ON1m**; $\mathbf{X^1}$ = m^6v^6A and $\mathbf{X^2}$ = m^6g^6A, **ON2c**; **X** = $vmnm^5U$ and **ON2f**; **X** = $vmnm^5U$.

11. Coupling reactions between ON2c and donor oligonucleotides with non-complementary sequences

The peptide coupling reactions were carried out under identical conditions to those described in Section 0.

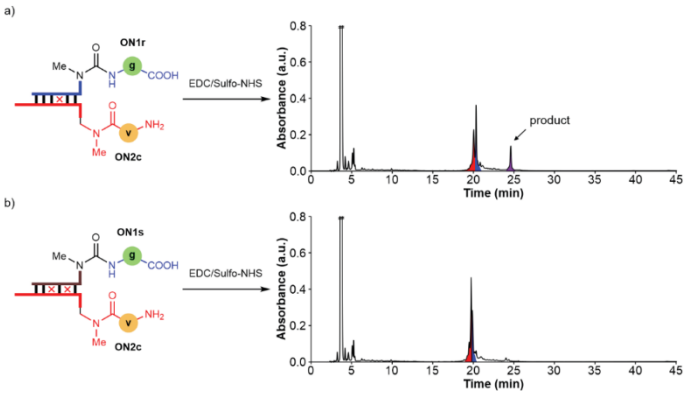


Figure S45. HPL-chromatograms of the reactions of **ON2c**; **X** = $vmnm^5U$ with: a) **ON1r**; **X** = m^6g^6A and b) **ON1s**; **X** = m^6g^6A in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S22. Results obtained in the coupling reactions of **ON2c**; **X** = $vmnm^5U$ with **ON1r**; **X** = m^6g^6A or **ON1s**; **X** = m^6g^6A using EDC/Sulfo-NHS as activator.

Donor strand	Acceptor strand	Yield (%)[a]
ON1r; **X** = m^6g^6A	**ON2c**; **X** = $vmnm^5U$	~14 (~35)[b]
ON1s; **X** = m^6g^6A		< 3 (~12)[b]

[a] Estimated yield from the chromatographic peak of the product using the calibration curve of **CON3**. Note that we assumed that the formed product features an extinction coefficient similar to that of **CON3**. [b] Using DMTMM•Cl as activator.

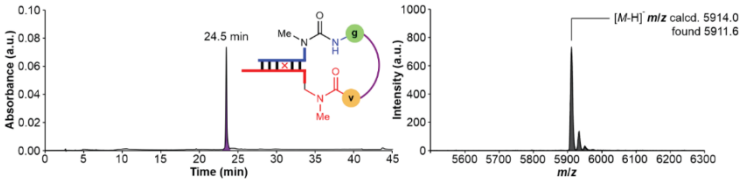
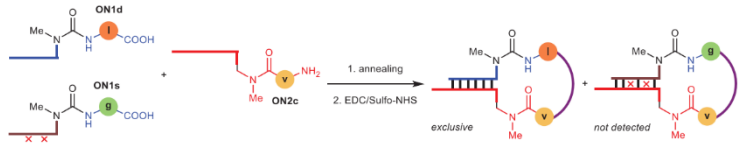


Figure S46. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product from the reaction of **ON2c**; **X** = $vmnm^5U$ with **ON1r**; **X** = m^6g^6A.



Scheme S12. Annealing and coupling reaction of **ON1d**; **X** = m^6i^6A, **ON1s**; **X** = m^6g^6A and **ON2c**; **X** = $vmnm^5U$. The formed peptide bonds are marked in purple.

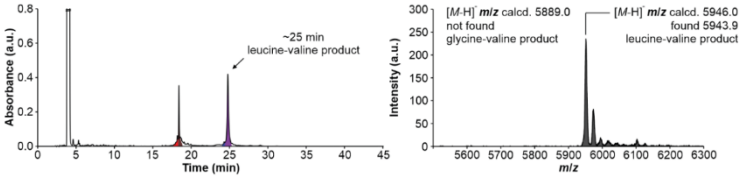


Figure S47. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the reaction of **ON1d**; **X** = $m^6\underline{I}^6A$, **ON1s**; **X** = $m^6\underline{g}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S23. Results obtained in the coupling reaction of **ON1d**; **X** = $m^6\underline{I}^6A$, **ON1s**; **X** = $m^6\underline{g}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$ using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error of <u>lv</u>-peptide (%)[a]	Yield of <u>gv</u>-peptide (%)
ON1d; **X** = $m^6\underline{I}^6A$ **ON1s**; **X** = $m^6\underline{g}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	65±2	not detected

[a] Calculated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

12. Coupling reactions between ON2c and donor oligonucleotides with different lengths

The peptide coupling reactions were carried out under identical conditions to those described in Section 0.

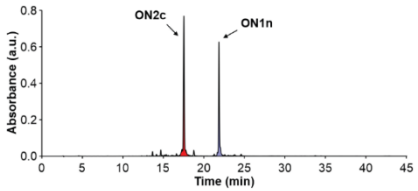


Figure S48. HPL-chromatogram of an equimolar mixture of **ON1n**; **X** = $m^6\underline{v}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$.

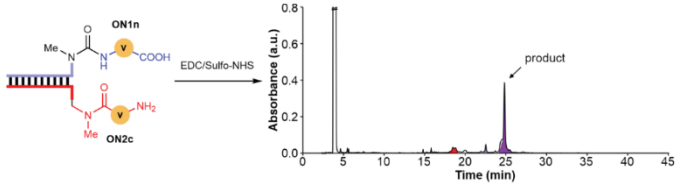


Figure S49. HPL-chromatogram of the reaction of **ON1n**; **X** = $m^6\underline{v}^6A$ with **ON2c**; **X** = $\underline{v}mnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S24. Result obtained in the coupling reaction of **ON1n**; **X** = $m^6\underline{v}^6A$ with **ON2c**; **X** = $\underline{v}mnm^5U$ using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1n; **X** = $m^6\underline{v}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	49±1

[a] Calculated yield from the chromatographic peak of the product based on the total area of the initial components (Figure S48).

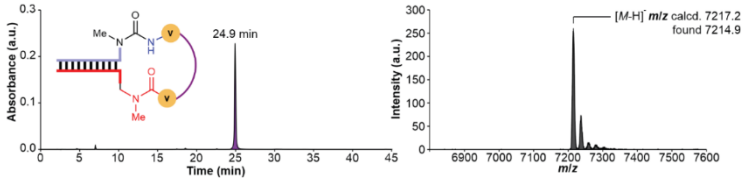
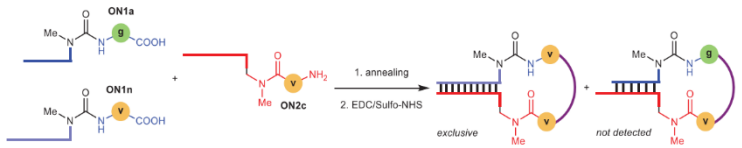


Figure S50. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product from the reaction of **ON1n**; **X** = $m^6\underline{v}^6A$ with **ON2c**; **X** = $\underline{v}mnm^5U$.



Scheme S13. Annealing and coupling reaction of **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1n**; **X** = $m^6\underline{v}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$. The formed peptide bonds are marked in purple.

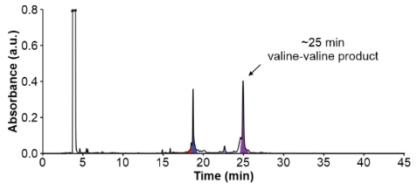


Figure S51. HPL-chromatogram of the reaction of **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1n**; **X** = $m^6\underline{v}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S25. Results obtained in the coupling reaction of **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1n**; **X** = $m^6\underline{v}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$ using EDC/Sulfo-NHS as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error of <u>vv</u>-peptide (%)[a]	Yield of <u>gv</u>-peptide (%)[a]
ON1a; **X** = $m^6\underline{g}^6A$ **ON1n**; **X** = $m^6\underline{v}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	49±2	not detected

[a] Calculated yield from the chromatographic peak of the product based on the total area of the initial components (Figure S48).

13. Stability of selected acceptor oligonucleotides (ON2)

The oligonucleotide (0.5 nmol) was added to an Eppendorf tube. Buffer, NaCl and water were added to the ON's solution and the reaction was heated in a Thermocycler.

Concentration of the components in the reaction mixture: 10-50 μM of oligonucleotide, 100 mM of buffer and 100 mM of NaCl (see figure footnotes for details).

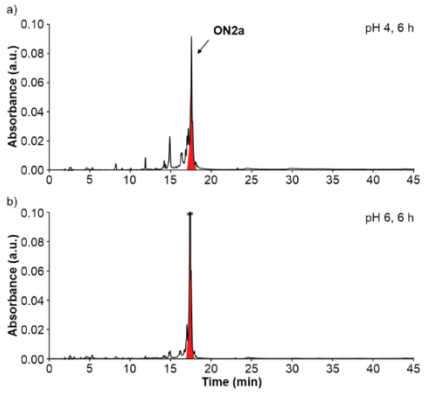


Figure S52. HPL-chromatograms of the stability of **ON2a**; **X** = mnm^{5}U in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 6 h at 90°C.

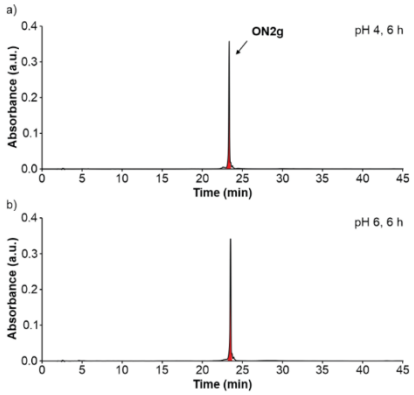


Figure S53. HPL-chromatograms of the stability of **ON2g**; **X** = mnm^{5}U in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 6 h at 90°C.

Table S26. Results obtained in the stability of **ON2a** and **ON2g** (average of, at least, two experiments).[a]

		Average Amount ± Error (%)	
pH	**Time (h)**	**ON2a**	**ON2g**
4	6	40±3	>95
6	6	70±5	>95

[a] Calculated amounts from the chromatographic peaks using the corresponding calibration curves.

14. Cleavage of urea in selected oligonucleotides and cyclic peptide products

The cleavage reactions were carried out under identical conditions to those described in Section 13.

14.1 Cleavage reactions of ON1c (m^{6}v^{6}A) and ON1k (v^{6}A) at pH 5

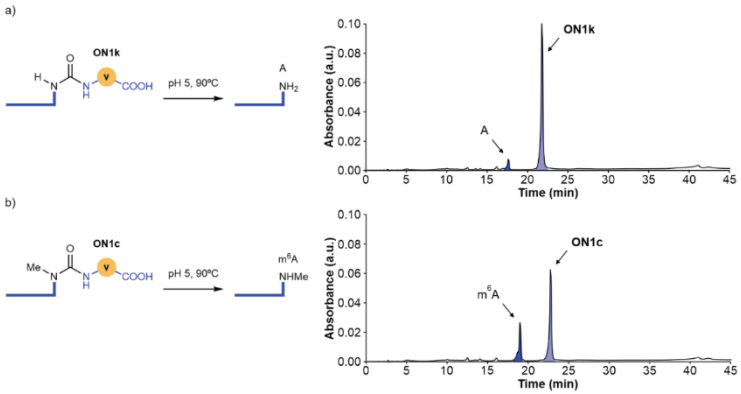


Figure S54. HPL-chromatograms of the cleavage reactions of: a) **ON1k**; **X** = v^{6}A and b) **ON1c**; **X** = m^{6}v^{6}A in acetate buffer at pH 5 after 12 h at 90°C.

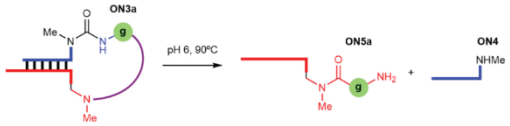
These experiments indicated that the urea cleavage reaction of the unmethylated aa^{6}A-RNA donor strand **ON1k** was slower than that of the methylated version, m^{6}aa^{6}A-RNA **ON1c**.

Table S27. Results obtained in the cleavage reactions of **ON1c** and **ON1k** (average of, at least, two experiments).[a]

		Average Amount ± Error (%)	
pH	**Time (h)**	**ON1k**	**A-strand**
		85±3	10±1
5	12	**ON1c**	**m^{6}A-strand**
		65±1	20±1

[a] Calculated amounts from the chromatographic peaks using the corresponding calibration curves.

14.2 Cleavage reaction of ON3a (m^{6}g^{6}A coupled with mnm^{5}U)



Scheme S14. Cleavage of urea in **ON3a**. The peptide bond is marked in purple.

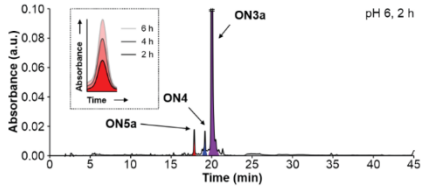


Figure S55. HPL-chromatogram of the cleavage reaction of **ON3a** in MES buffer at pH 6 after 2 h at 90°C. Inset shows the selected region of the HPL-chromatograms after 2, 4 and 6 h.

Table S28. Results obtained in the cleavage reaction of **ON3a** (average of, at least, two experiments).[a]

		Average Amount ± Error (%)		
pH	**Time (h)**	**ON3a**	**ON4 (m^{6}A)**	**ON5a (gmnm^{5}U)**
6	6	75±2	15±1	15±1 (t_R = 17.5 min)

[a] Calculated amounts from the chromatographic peaks using the corresponding calibration curves.

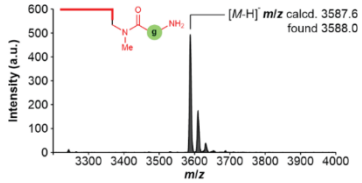


Figure S56. MALDI-TOF mass spectrum (negative mode) of the isolated **ON5a** (gmnm^{5}U).

Additional experiments at pH 4 and pH 6

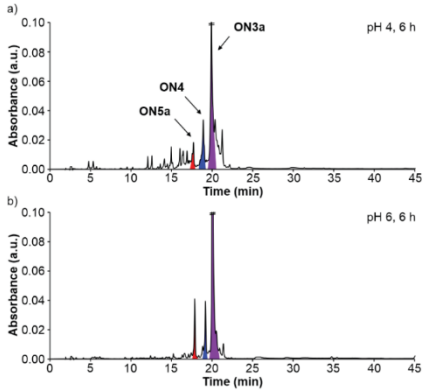


Figure S57. HPL-chromatograms of the cleavage reactions of **ON3a** in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 6 h at 90°C.

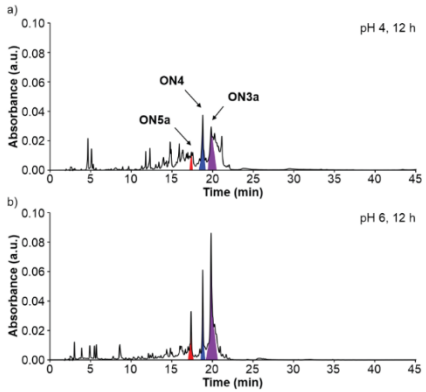


Figure S58. HPL-chromatograms of the cleavage reactions of **ON3a** in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 12 h at 90°C.

Table S29. Results obtained in the cleavage reaction of **ON3a** (average of, at least, two experiments).[a]

		Average Amount ± Error (%)
pH	**Time (h)**	**ON5a (gmnm^{5}U)**
4	6	10±2
	12	n.d.
6	6	15±1
	12	10±1

[a] Calculated amounts from the chromatographic peak using the calibration curve of **CON2**. n.d. = not determined.

14.3 Cleavage reactions of ON3c (m^{6}g^{6}A coupled with vmnm^{5}U)

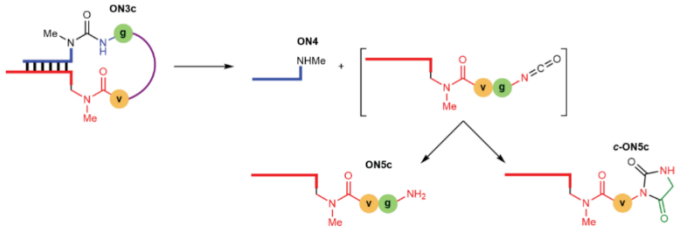


Figure S59. Cleavage of urea in **ON3c**. The peptide bond is marked in purple.

Cleavage reactions at 60°C

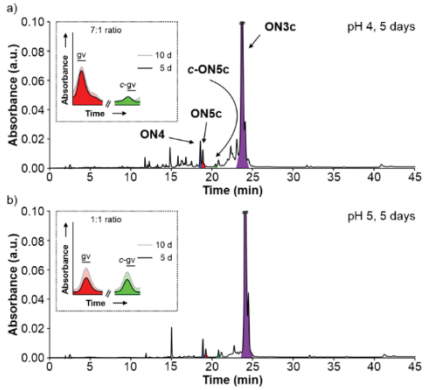


Figure S60. HPL-chromatograms of the cleavage reactions of **ON3c** in acetate buffer at: a) pH 4 and b) pH 5 after 5 days at 60°C. Inset shows the selected region of the HPL-chromatograms after 5 and 10 days.

Table S30. Results obtained in the cleavage reactions of **ON3c** at 60°C (average of, at least, two experiments).[a]

		Average Amount ± Error (%)				
pH	**Time (days)**	**ON3c**	**ON4 (m^{6}A)**	**ON5c (gvmnm^{5}U)**	***c*-ON5c (*c*-gvmnm^{5}U)**	**Ratio (ON5c/*c*-ON5c)**
4	10	50±2	10.5±1	9±1 (t_R = 19.5 min)	1.5±0.5 (t_R = 21.0 min)	~7:1
5	10	80±3	6±1	3±1	3±1	~1:1

[a] Calculated amounts from the chromatographic peaks using the corresponding calibration curves.

Cleavage reactions at 90°C

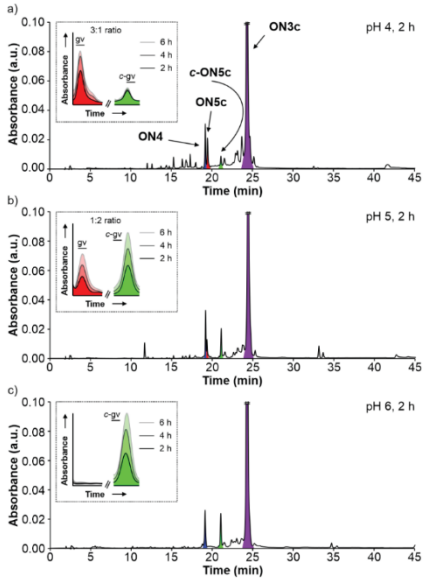


Figure S61. HPL-chromatograms of the cleavage reactions of **ON3c** in: a) acetate buffer at pH 4; b) acetate buffer at pH 5 and c) MES buffer at pH 6 after 2 h at 90°C. Inset shows the selected region of the HPL-chromatograms after 2, 4 and 6 h.

Table S31. Results obtained in the cleavage reactions of **ON3c** at 90°C (average of, at least, two experiments).[a]

		Average Amount ± Error (%)				
pH	**Time (h)**	**ON3c**	**ON4 (m^{6}A)**	**ON5c (gvmnm^{5}U)**	***c*-ON5c (*c*-gvmnm^{5}U)**	**Ratio (ON5c/*c*-ON5c)**
4	6	30±3	20±2	15±2 (t_R = 19.5 min)	5±1 (t_R = 21.0 min)	~3:1
5	6	55±3	25±2	8±1	17±3	~1:2
6	6	60±2	25±1	-	25±1	-

[a] Calculated amounts from the chromatographic peaks using the corresponding calibration curves.

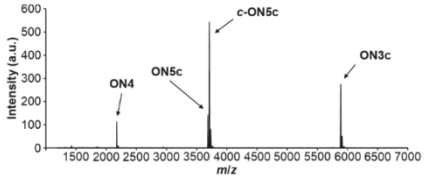


Figure S62. MALDI-TOF mass spectrum (negative mode) of the cleavage reaction of **ON3c** in acetate buffer at pH 5 after 2 h at 90°C. A similar MALDI-TOF mass spectrum was obtained at pH 4. The indicated peaks correspond to the $[M\text{-}H]^-$ ions.

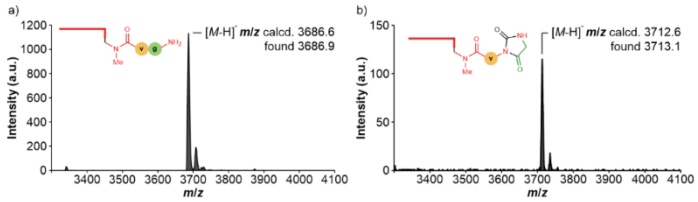
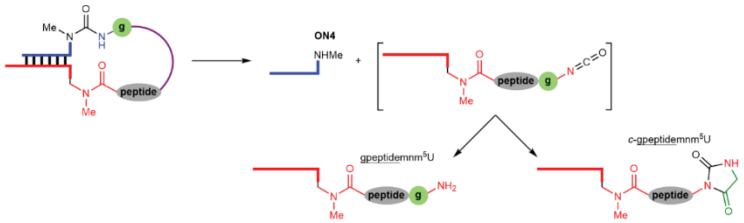


Figure S63. MALDI-TOF mass spectrum (negative mode) of the isolated: a) **ON5c** (gvmnm^{5}U) and b) ***c*-ON5c** (*c*-gvmnm^{5}U).

14.4 Cleavage reactions of peptide-oligonucleotides at pH 4



Scheme S15. Cleavage of urea in gpeptide-oligonucleotides. The peptide bond is marked in purple.

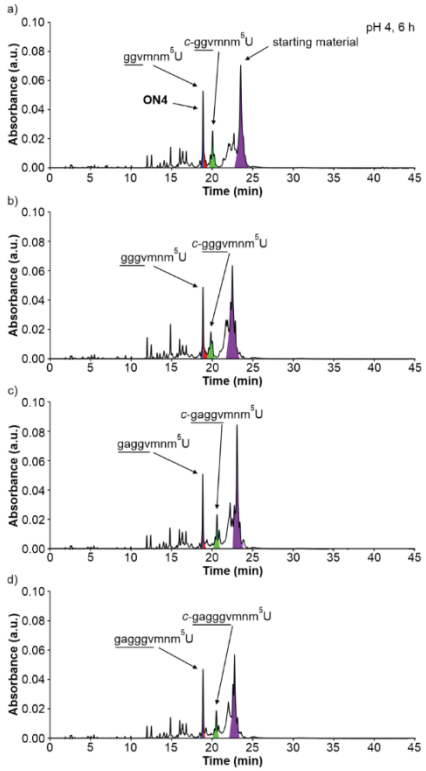


Figure S64. HPL-chromatograms of the cleavage reactions of peptide-oligonucleotides (Section 8.2) in acetate buffer at pH 4 to give: a) ggvmnm^{5}U; b) gggvmnm^{5}U; c) gaggvmnm^{5}U and d) gagggvmnm^{5}U oligonucleotides, together with hydantoin side products, after 6 h at 90°C.

The 3'-H_2N-peptidemnm^{5}U-RNA-5' and m^{6}A products overlap in the HPL-chromatograms. Therefore, they were isolated as a mixture in a single fraction.

Table S32. Results obtained in the cleavage reactions of peptide-oligonucleotides (Section 8.2).[a]

3'-H_2N-peptidemnm^{5}U-RNA-5'	Amount (%)
3'-ggvmnm^{5}U-RNA-5'	~12
3'-gggvmnm^{5}U-RNA-5'	~10
3'-gaggvmnm^{5}U-RNA-5'	~10
3'-gagggvmnm^{5}U-RNA-5'	~10

[a] Estimated amounts assuming that the 3'-H_2N-peptidemnm^{5}U-RNA-5' products and the hydantoin counterparts were formed in a similar extent.

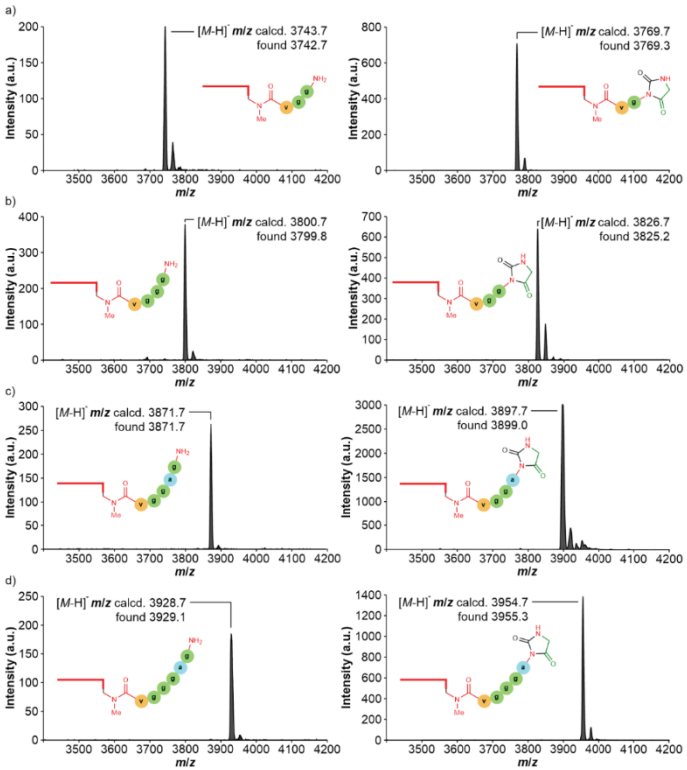


Figure S65. MALDI-TOF mass spectra (negative mode) of the isolated: a) ggvmnm^{5}U; b) gggvmnm^{5}U; c) gaggvmnm^{5}U and d) gagggvmnm^{5}U oligonucleotides (left) and hydantoin side products (right). Note that the analyzed 3'-H_2N-peptidemnm^{5}U-RNA-5' samples (left) contained the m^{6}A product (*m*/*z* region not shown).

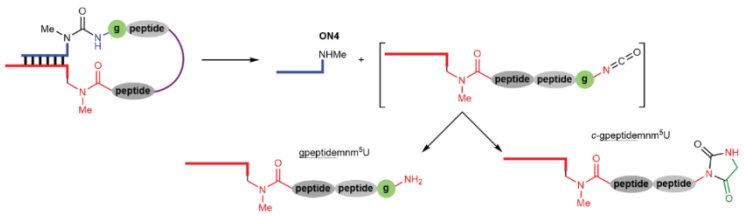


Figure S66. Cleavage of urea in gpeptide-peptidemnm^{5}U-oligonucleotides. The peptide bond is marked in purple.

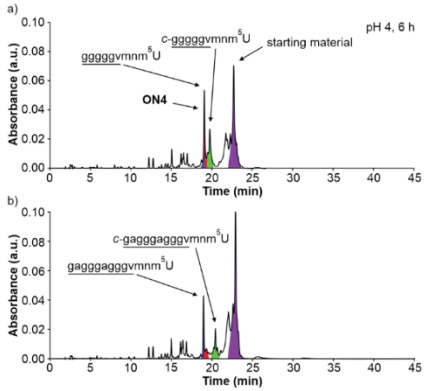


Figure S67. HPL-chromatograms of the cleavage reactions of peptide-oligonucleotides (Section 8.3) in acetate buffer at pH 4 to give: a) gggggvmnm^{5}U and b) gagggagggvmnm^{5}U oligonucleotides, together with hydantoin side products, after 6 h at 90°C.

The 3'-H_2N-peptidemnm^{6}U-RNA-5' and m^{6}A products overlap in the HPL-chromatograms. In addition, the 3'-H_2N-peptidemnm^{5}U-RNA-5' in a) overlaps with the hydantoin side product.

Table S33. Results obtained in the cleavage reactions of peptide-oligonucleotides (Section 8.3).[a]

3'-H_2N-peptidemnm^{5}U-RNA-5'	Amount (%)
3'-gggggvmnm^{5}U-RNA-5'	~10
3'-gagggagggvmnm^{5}U-RNA-5'	~9

[a] Estimated amounts assuming that the 3'-H_2N-peptidemnm^{5}U-RNA-5' products and the hydantoin counterparts were formed in a similar extent.

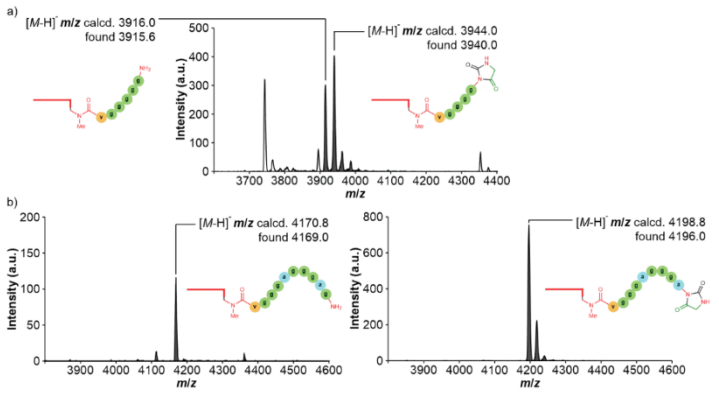


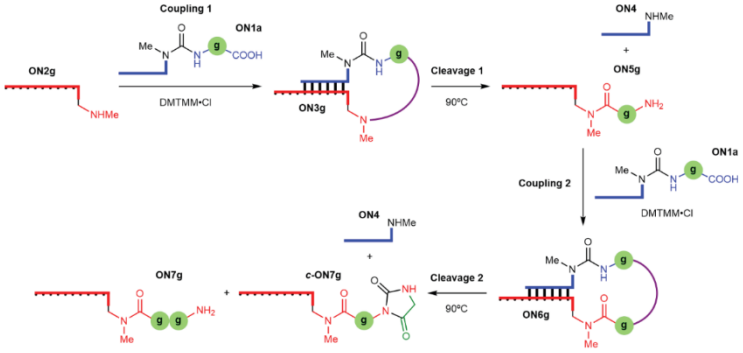
Figure S68. MALDI-TOF mass spectra (negative mode) of the isolated: a) ggggg**v**mnm^{5}U and b) gagggagggvmnm^{5}U. Hydantoin side products are also shown. Note that the analyzed 3'-H_2N-peptidemnm^{5}U-RNA-5' samples contained the m^{6}A product (*m/z* region not shown).

15. Coupling and cleavage reactions between donor and acceptor oligonucleotides containing 2'-OMe nucleosides

The peptide coupling and urea cleavage reactions were carried out under identical conditions to those described in Section 0 and Section 13, respectively.

15.1 Coupling and cleavage reactions of ON1a (m^{6}g^{6}A) with ON2g

Each coupling reaction was performed using 1 equiv. of **ON1a** with respect to the acceptor oligonucleotide, **ON2g** or **ON5g**.



Scheme S16. Coupling and cleavage of **ON1a**; **X** = m^{6}g^{6}A with **ON2g**. The formed peptide bond is marked in purple.

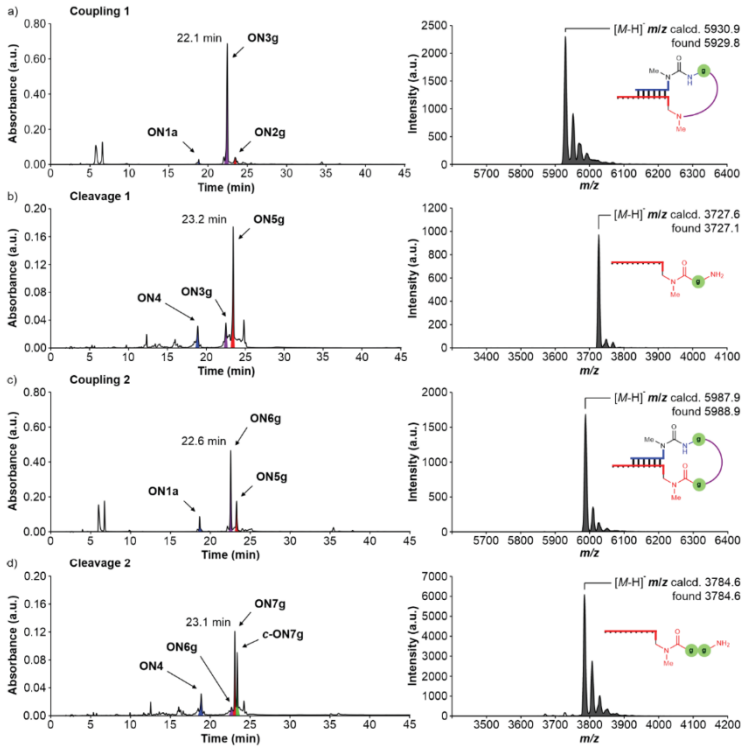


Figure S69. left) HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2g**: a) coupling 1; b) cleavage 1; c) coupling 2 and d) cleavage 2. The product of each step was separated by HPLC and added into the next reaction. right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions a)-d).

Table S34. Results obtained in the coupling and cleavage reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2g**.

Steps	Activators	pH	T (°C)	Time (h)	Yield (%)[a]
Coupling 1 (**ON3g**)	EDC/Sulfo-NHS	6	25	24	~39
	DMTMM•Cl	6	25	24	~69
Cleavage 1 (**ON5g**)	-	4	90	24	60
	-	6	90	24	46
Coupling 2 (**ON6g**)	DMTMM•Cl	6	25	24	~42
Cleavage 2 (**ON7g**)	-	4	90	24	34

[a] Calculated/estimated amounts from the chromatographic peaks using the corresponding calibration curves.

One pot reaction

The one pot reaction was performed with 15 nmol of **ON2g** as starting acceptor strand. 15 nmol of donor strand **ON1a** or **ON1g** were added for each coupling reaction. After each coupling reaction and the second cleavage, the crude was filtered using an Amicon® ultra centrifugal filter (3 kDa Nominal Molecular Weight Cut-Off) to remove the remaining activator and exchange the buffer solution. The volume of the solution was maintained constant throughout the five reaction steps. 20 µL of the crude (1 nmol) were analyzed by HPLC after the second coupling, the second cleavage and the third coupling reactions.

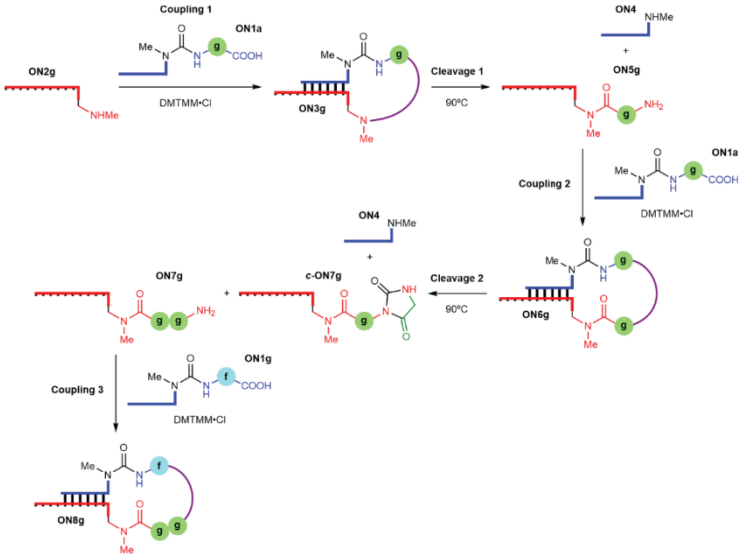


Figure S70. Coupling and cleavage of **ON1a**; **X** = $m^6\underline{g}^6A$ and **ON1g**; **X** = $m^{6\underline{f}6}A$ with **ON2g**. The formed peptide bond is marked in purple.

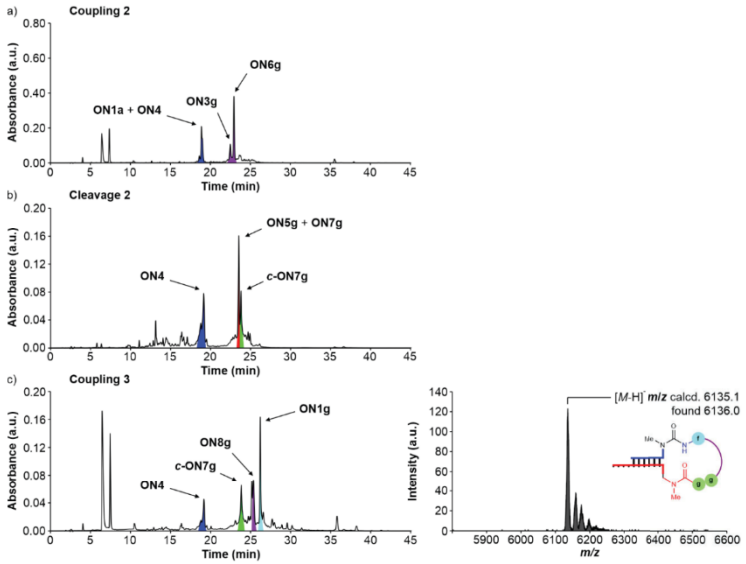


Figure S71. HPL-chromatograms of the one pot reaction of **ON1a**; **X** = $m^6\underline{g}^6A$ and **ON1g**; **X** = $m^{6\underline{f}6}A$ with **ON2g**: a) coupling 2; b) cleavage 2 and c) coupling 3.

Table S35. Results obtained in the one pot reaction of **ON1a**; **X** = $m^6\underline{g}^6A$ and **ON1g**; **X** = $m^{6\underline{f}6}A$ with **ON2g**.

Steps	Activators	pH	T (°C)	Time (h)	Yield (%)[a]
Coupling 2 (**ON6g**)	DMTMM•Cl	6	25	24	~36 in three steps
Cleavage 2 (**ON5g** + **ON7g**)	-	4	90	24	23 in four steps
Coupling 3 (**ON8g**)	DMTMM•Cl	6	25	24	~10 in five steps

[a] Calculated/estimated amounts from the chromatographic peaks using the corresponding calibration curves.

15.2 Coupling and cleavage reactions of ON1o (m^6g^6Am) with ON2h

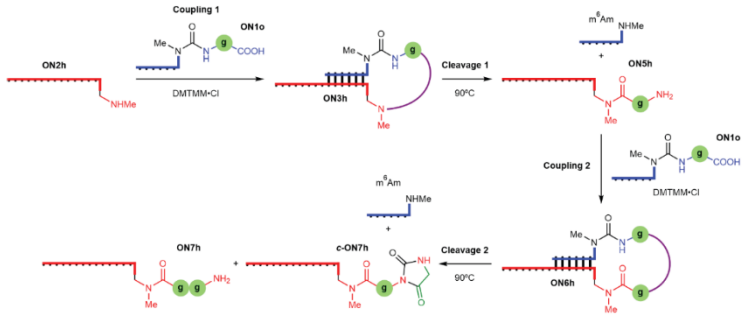


Figure S72. Coupling and cleavage of **ON1o**; **X** = $m^6\underline{g}^6Am$ with **ON2h**. The formed peptide bond is marked in purple.

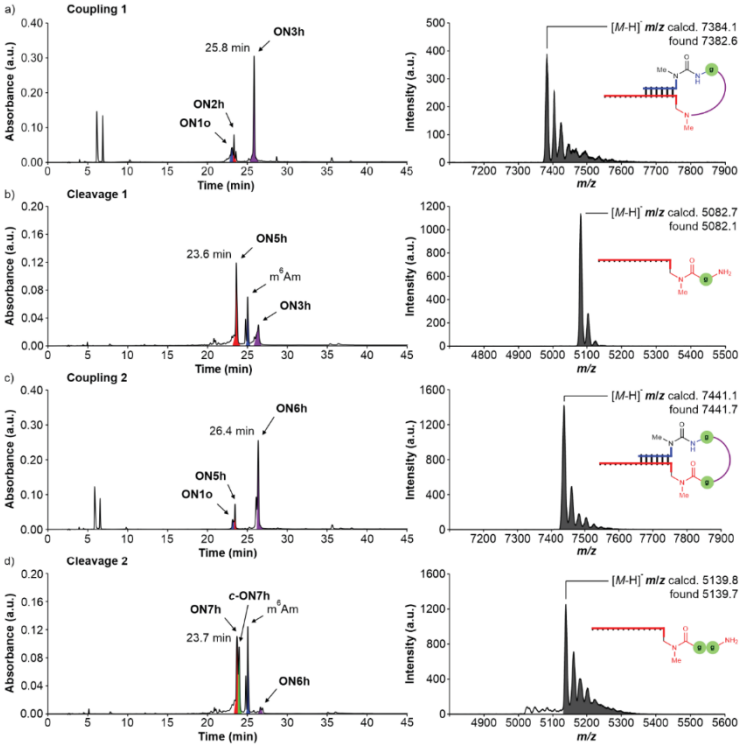


Figure S73. left) HPL-chromatograms of the reactions of **ON1o**; **X** = m^{6}g^{6}Am with **ON2h**: a) coupling 1; b) cleavage 1; c) coupling 2 and d) cleavage 2. The product of each step was separated by HPLC and added into the next reaction. right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions a)-d).

Table S36. Results obtained in the coupling and cleavage reactions of **ON1o**; **X** = m^{6}g^{6}Am with **ON2h**.

Steps	Activators	pH	T (°C)	Time (h)	Yield (%)[a]
Coupling 1 (**ON3h**)	DMTMM•Cl	6	25	24	46
Cleavage 1 (**ON5h**)	-	4	90	48	30
Coupling 2 (**ON6h**)	DMTMM•Cl	6	25	24	41
Cleavage 2 (**ON7h**)	-	4	90	48	28

[a] Calculated amounts from the chromatographic peaks using the corresponding calibration curves.

15.3 Coupling and cleavage reactions of donor and acceptor-peptide oligonucleotides

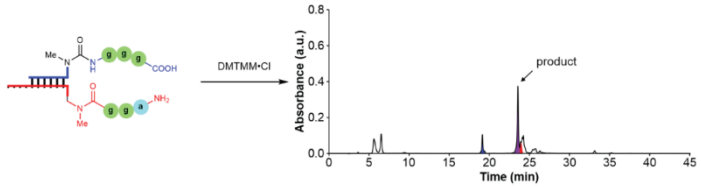


Figure S74. HPL-chromatogram of the reaction of 5'-m^6(ggg)^{6}A-RNA-3' with 3'-aggmnm^{5}U-RNA-5' containing 2'-OMe nucleosides in MES buffer at pH 6 using DMTMM•Cl as activator.

Table S37. Result obtained in the coupling reaction of peptide-modified donor and acceptor oligonucleotides using DMTMM•Cl as activator.

Donor strand	Acceptor strand	Yield (%)[a]
5'-m^6(ggg)^{6}A-RNA-3'	3'-aggmnm^{5}U-RNA-5'	~50

[a] Estimated yield from the chromatographic peak of the product using the calibration curve of **CON3**.

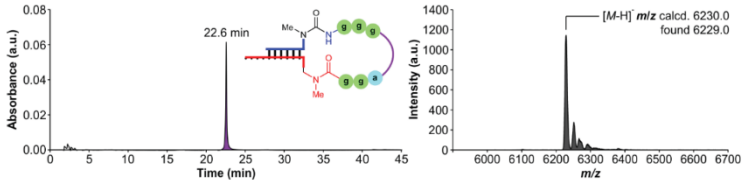
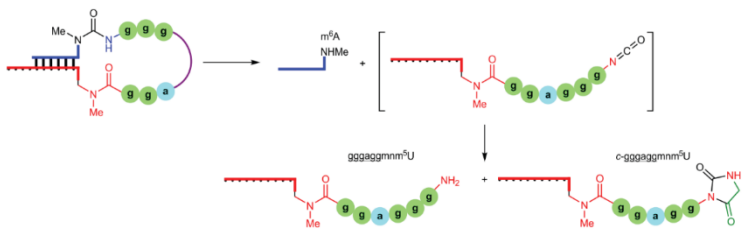


Figure S75. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product from the reaction of 5'-m^6(ggg)^{6}A-RNA-3' with 3'-aggmnm^{5}U-RNA-5' containing 2'-OMe nucleosides.



Scheme S17. Cleavage of urea in peptide-peptidemnm^{5}U-oligonucleotide. The peptide bond is marked in purple.

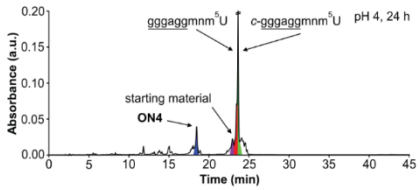


Figure S76. HPL-chromatogram of the cleavage reaction of peptide-oligonucleotide in acetate buffer at pH 4 to give gggaggmnm^{5}U and *c*-gggaggmnm^{5}U oligonucleotides after 24 h at 90°C.

The 3'-H$_2$N-peptidemnm^{5}U-RNA-5' and hydantoin side products overlap in the HPL-chromatogram.

Table S38. Result obtained in the cleavage reaction of peptide-oligonucleotide.[a]

Product oligonucleotides containing 2'-OMe nucleosides	Amount (%)
3'-gggaggmnm^{5}U-RNA-5' and 3'-*c*-gggaggmnm^{5}U-RNA-5'	~85 (t_R = 23.6 min)

[a] Estimated amount from the chromatographic peak using the calibration curve of **CON2**.

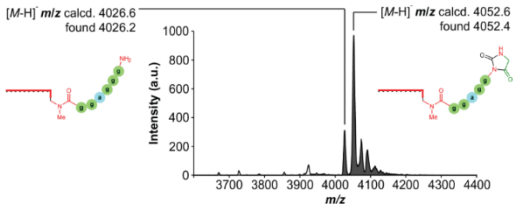


Figure S77. MALDI-TOF mass spectra (negative mode) of the isolated gggaggmnm^{5}U and hydantoin side product.

15.4 Coupling reactions between ON2g and donor oligonucleotides of different length

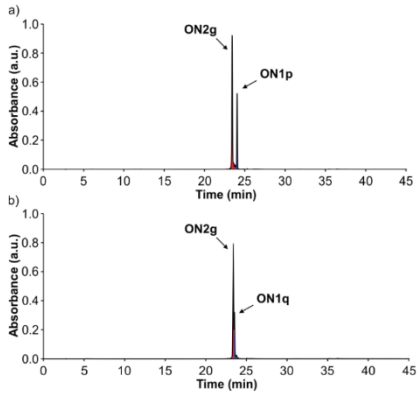
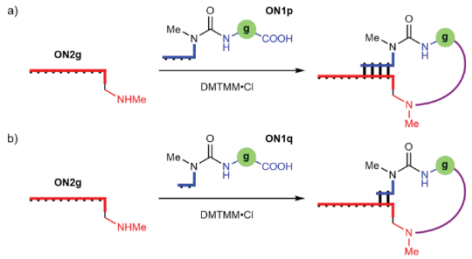


Figure S78. HPL-chromatograms of an equimolar mixture of **ON2g**; **X** = mnm^{5}U with: a) **ON1p**; **X** = m^{6}g^{6}Am and b) **ON1q**; **X** = m^{6}g^{6}Am.



Scheme S18. Coupling of **ON2g**; **X** = mnm^{5}U with: a) **ON1p**; **X** = m^{6}g^{6}Am and b) **ON1q**; **X** = m^{6}g^{6}Am. The formed peptide bond is marked in purple.

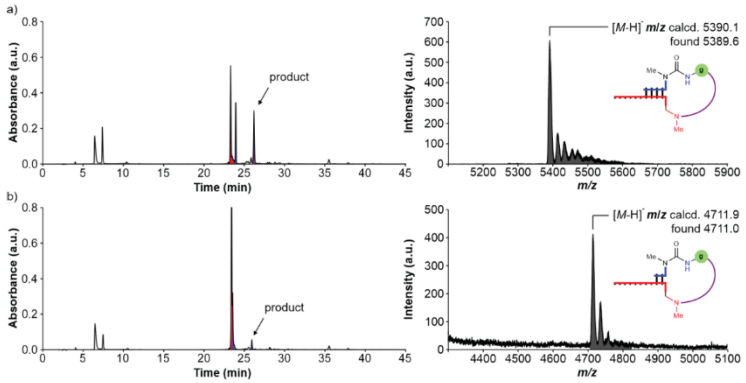


Figure S79. left) HPL-chromatograms of the reactions of: a) **ON1p**; **X** = m^{6}g^{6}Am and b) **ON1q**; **X** = m^{6}g^{6}Am with **ON2g**; **X** = mnm^{5}U in MES buffer at pH 6 using DMTMM•Cl as activator. The reaction b) was carried out at 0°C using 1 M NaCl. right) MALDI-TOF mass spectra (negative mode) of the isolated products.

Table S39. Results obtained in the coupling reactions of **ON2g**; **X** = mnm^{5}U with **ON1p**; **X** = m^{6}g^{6}Am or **ON1q**; **X** = m^{6}g^{6}Am using DMTMM•Cl as activator (average of, at least, two experiments).

Donor strand	Acceptor strand	Average Yield ± Error (%)[a]
ON1p; **X** = m^{6}g^{6}Am	**ON2g**; **X** = mnm^{5}U	19±2 (t_R = 26.2 min)
ON1q; **X** = m^{6}g^{6}Am		5±1[b] (t_R = 26.0 min)

[a] Calculated yield from the chromatographic peak of the product based on the total area of the initial components (Figure S78). [b] Using 1 M NaCl at 0°C.

16. Determination of melting temperatures by UV spectroscopic experiments

The UV melting curves were measured on a JASCO V-650 spectrometer at 260 nm using 10 mm QS cuvettes with a scanning rate of 1°C·min^{-1}. The obtained UV spectroscopic data were fit to the corresponding function to determine the melting temperature/s.

For double strands of non-self-complementary oligonucleotides, the data were fit to a two-state melting model, *i.e.* double strand – random coil equilibrium, using a mono-sigmoidal Boltzmann function.[14] On the contrary, the data were fit to a three-state melting model, *i.e.* double strand – hairpin – random coil equilibria, for single strands of self-complementary oligonucleotides using a double-sigmoidal Boltzmann function.[15,16]

For the experiments, we prepared aqueous solutions containing equimolar amounts of the oligonucleotides (5 μM), 10 mM phosphate buffer at pH 7 and 150 mM NaCl. The oligonucleotides were annealed by heating to 95°C for 4 min and, subsequently, by cooling down slowly to 5°C before the variable-temperature UV spectroscopic experiment.

16.1 Melting temperature of a double strand from canonical oligonucleotides

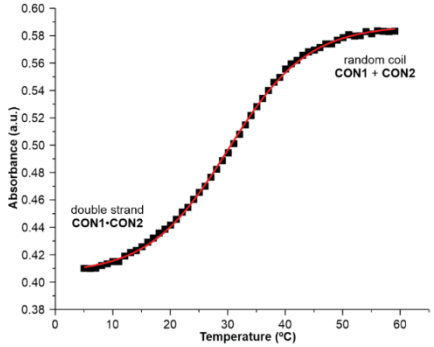


Figure S80. Melting curve of **CON1** and **CON2**. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.1°C.

16.2 Melting temperatures of double strands from donor and acceptor oligonucleotides

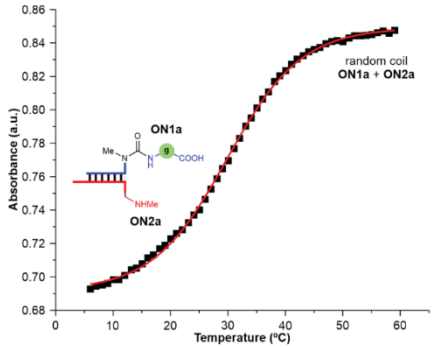


Figure S81. Melting curve of **ON1a**; **X** = m^{6}g^{6}A and **ON2a**; **X** = mnm^{5}U. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.4°C.

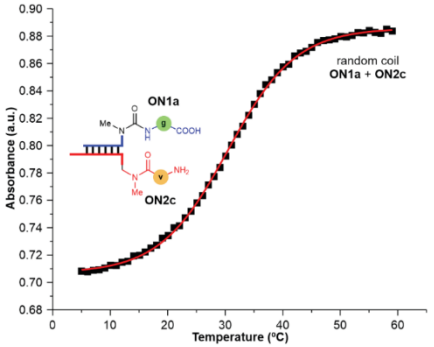


Figure S82. Melting curve of **ON1a**; **X** = m^{6}g^{6}A and **ON2c**; **X** = vmnm^{5}U. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.5°C.

The melting temperatures of the double strands containing modified A and U bases, **ON1a**; **X** = m^{6}g^{6}A, **ON2a**; **X** = mnm^{5}U and **ON2c**; **X** = vmnm^{5}U, were very similar to those determined for canonical oligonucleotides, **CON1** and **CON2**.

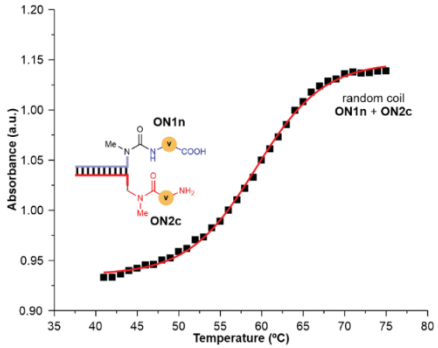


Figure S83. Melting curve of **ON1n**; **X** = m^{5}v^{6}A and **ON2c**; **X** = vmnm^{5}U. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 59.2°C.

16.3 Melting temperatures of double strands from donor and acceptor peptide-oligonucleotides

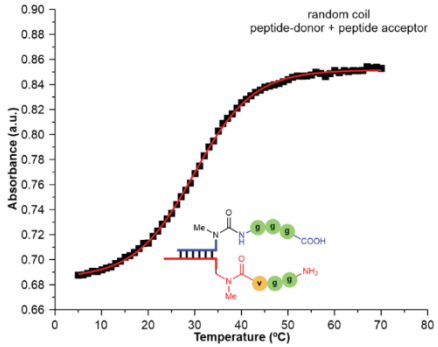


Figure S84. Melting curve of 5'-m^6(ggg)^{6}A-RNA-3' with 3'-ggvmnm^{5}U-RNA-5'. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.0°C.

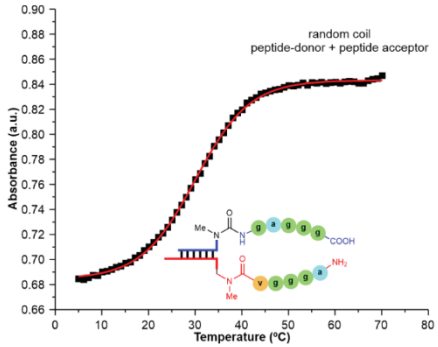


Figure S85. Melting curve of 5'-m^6(gaggg)^{6}A-RNA-3' with 3'-agggvmnm^{5}U-RNA-5'. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 29.8°C.

16.4 Melting temperatures of selected cyclic peptide products

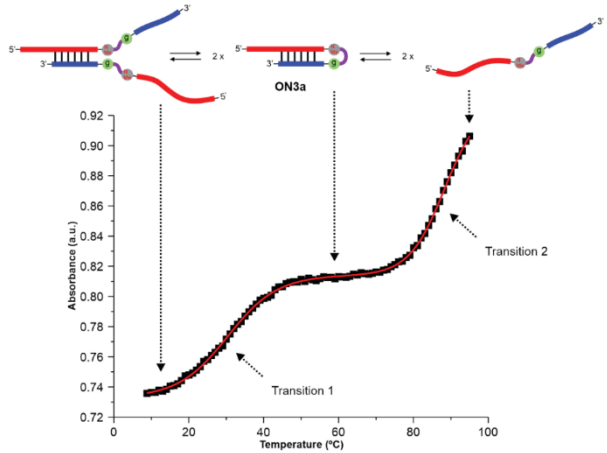


Figure S86. Melting curve of **ON3a**. Line shows the fit of the data to a three-state melting model using a double-sigmoidal Boltzmann function. T_{m1} = 30.8°C and T_{m2} = 87.5 °C. Top panel shows representation of the three states involved in the two transitions.

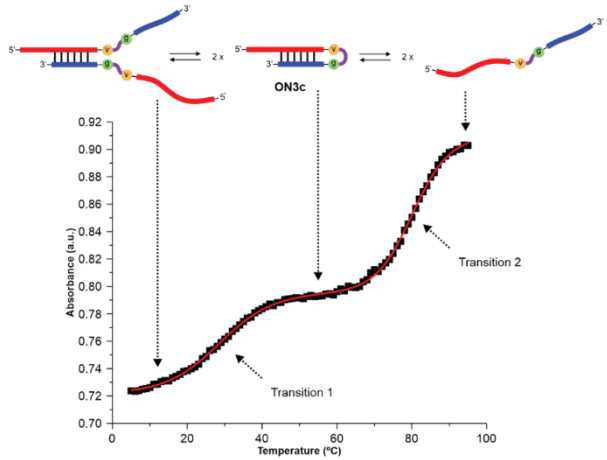


Figure S87. Melting curve of **ON3c**. Line shows the fit of the data to a three-state melting model using a double-sigmoidal Boltzmann function. T_{m1} = 28.4°C and T_{m2} = 80.1°C. Top panel shows representation of the three states involved in the two transitions.

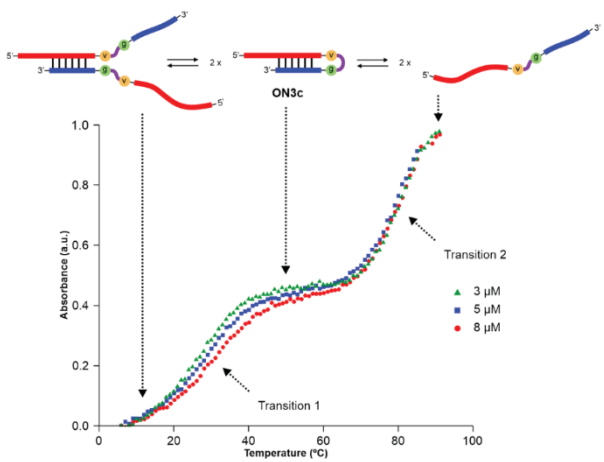
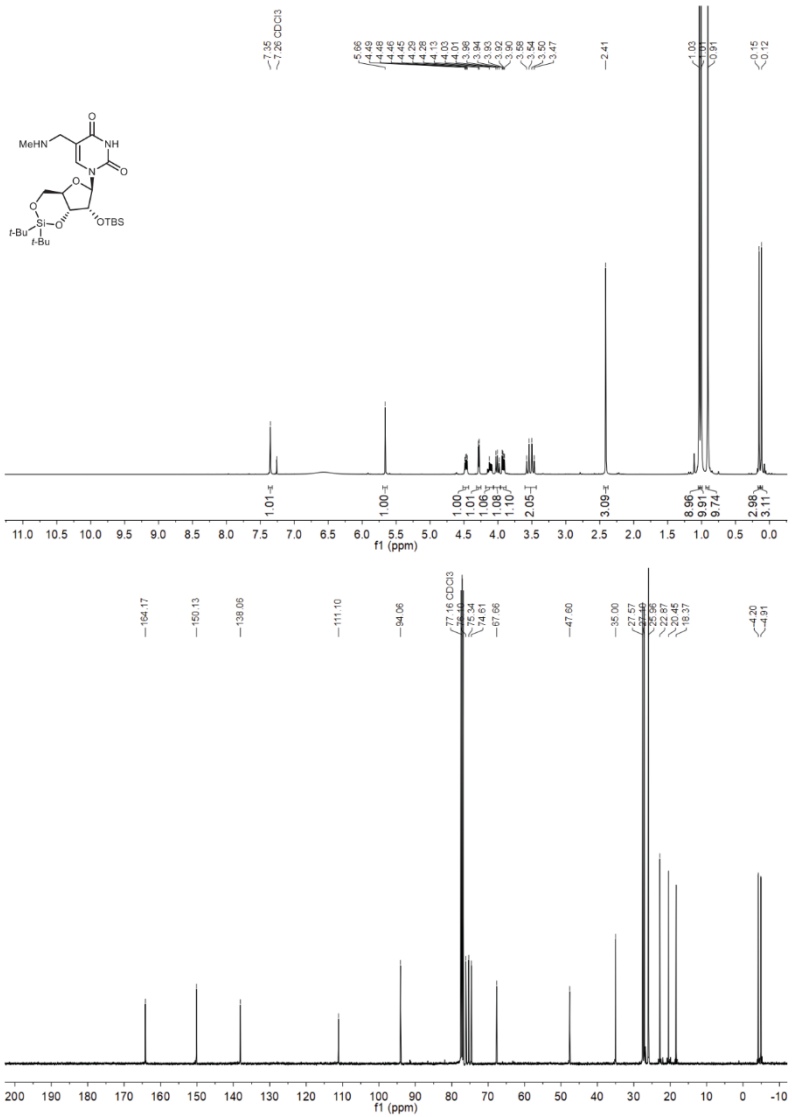


Figure S88. Normalized melting curves of **ON3c** at 3, 5 and 8 μM concentration. Top panel shows representation of the three states involved in the two transitions.

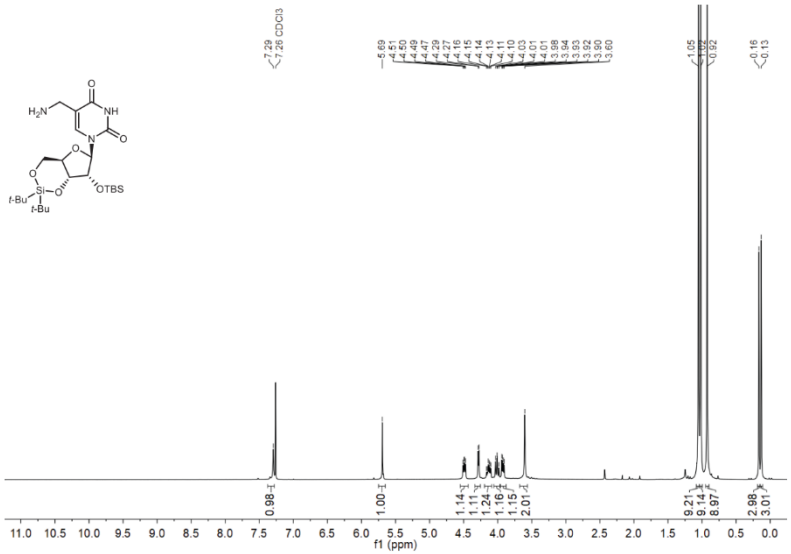
We observed a two-step melting profile in the experiments performed with the RNA oligonucleotides **ON3a** and **ON3c** (**Figure S86**, **Figure S87** and **Figure S88**). At low temperature (transition 1), the double strand (duplex) is transformed into the hairpin. At high temperature (transition 2), the hairpin is converted into the random coil. The intermolecular and intramolecular dissociation of the base pairs, *i.e.* breaking of hydrogen-bonding and π-stacking interactions, is induced by the increase in temperature over the course of the experiments.[15,16]

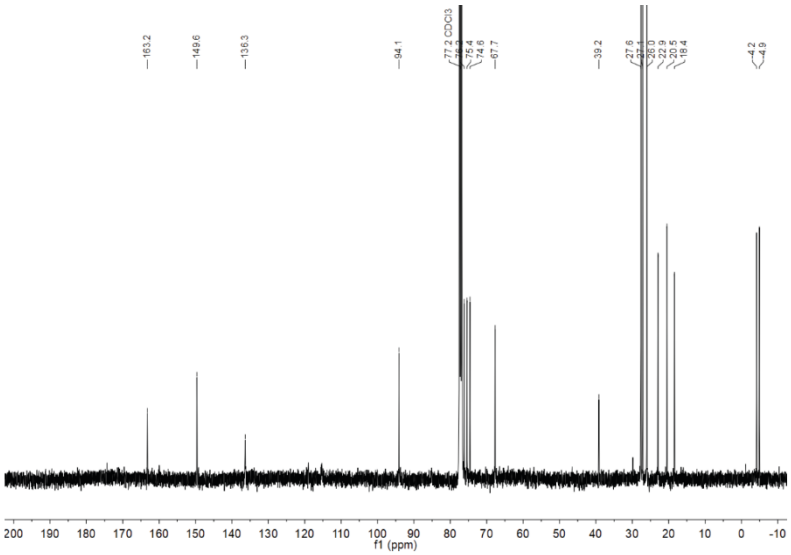
17. NMR spectra of synthesized compounds

^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 3a

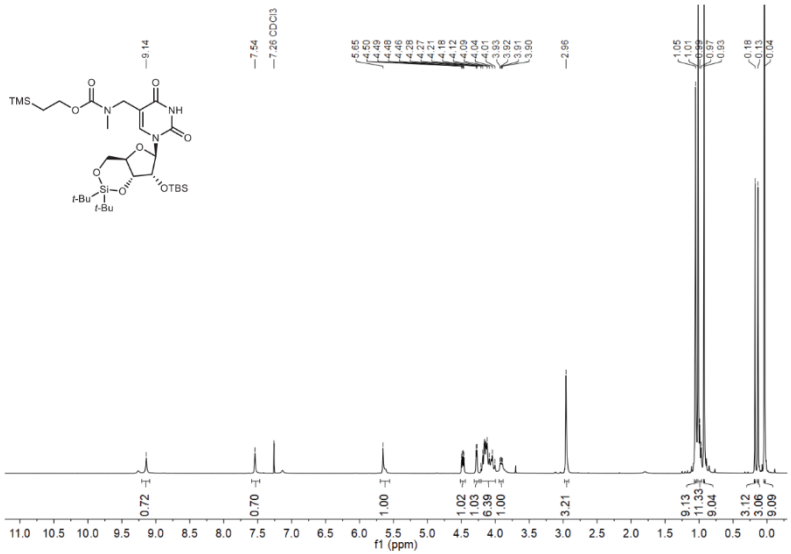


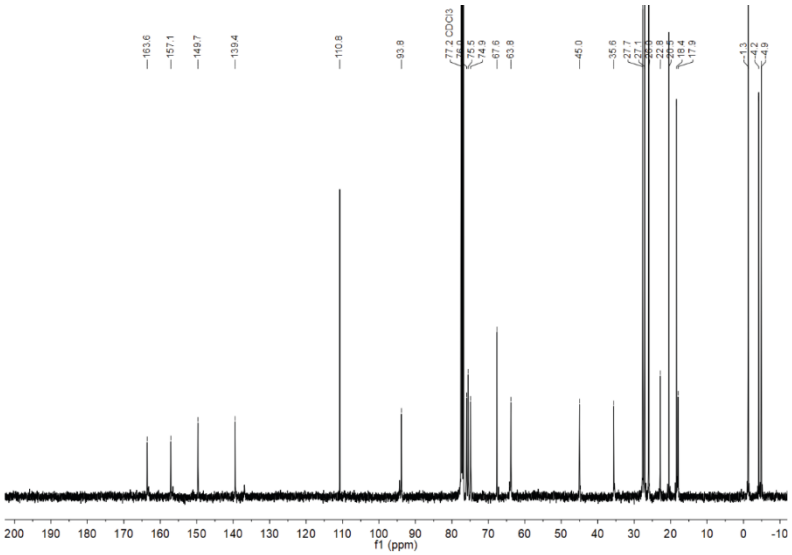
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 3b



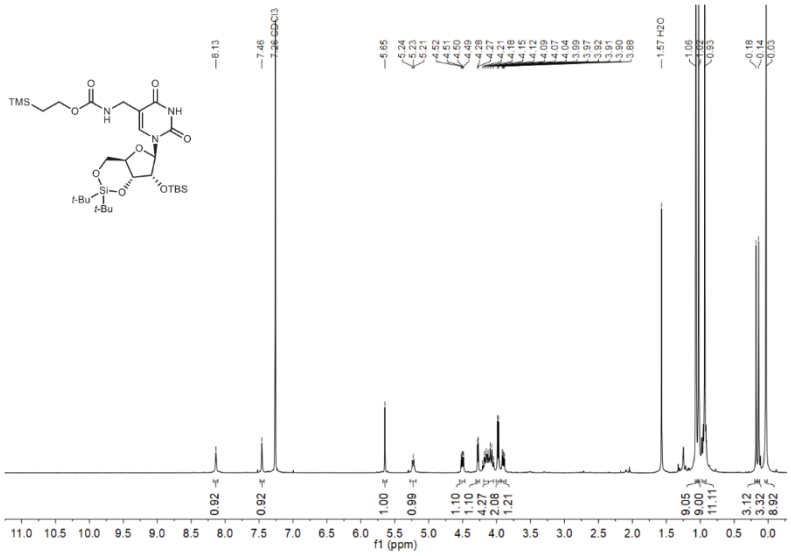


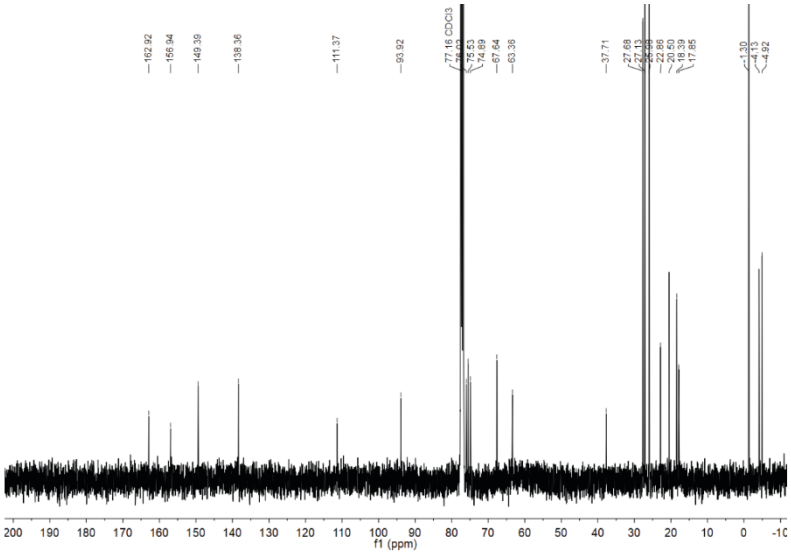
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4a



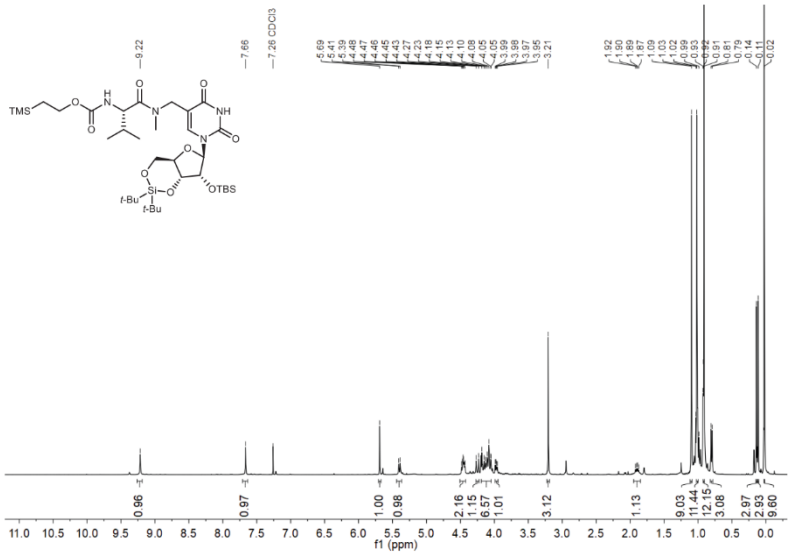


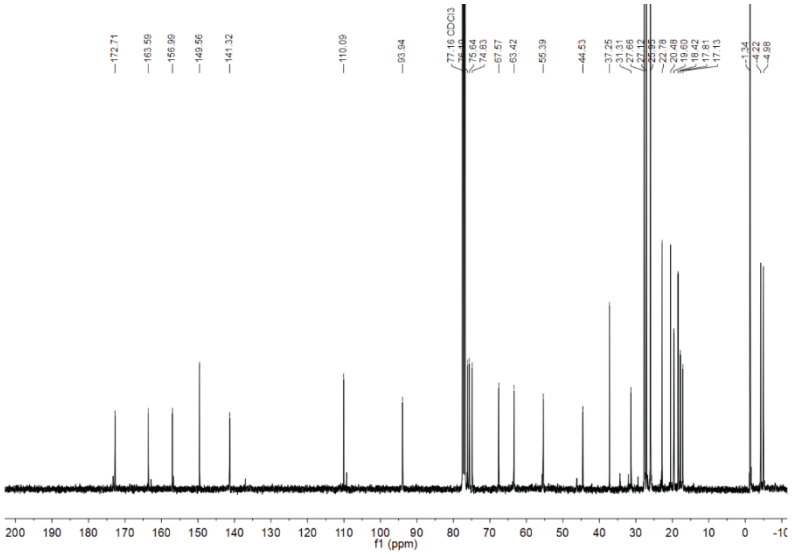
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4b



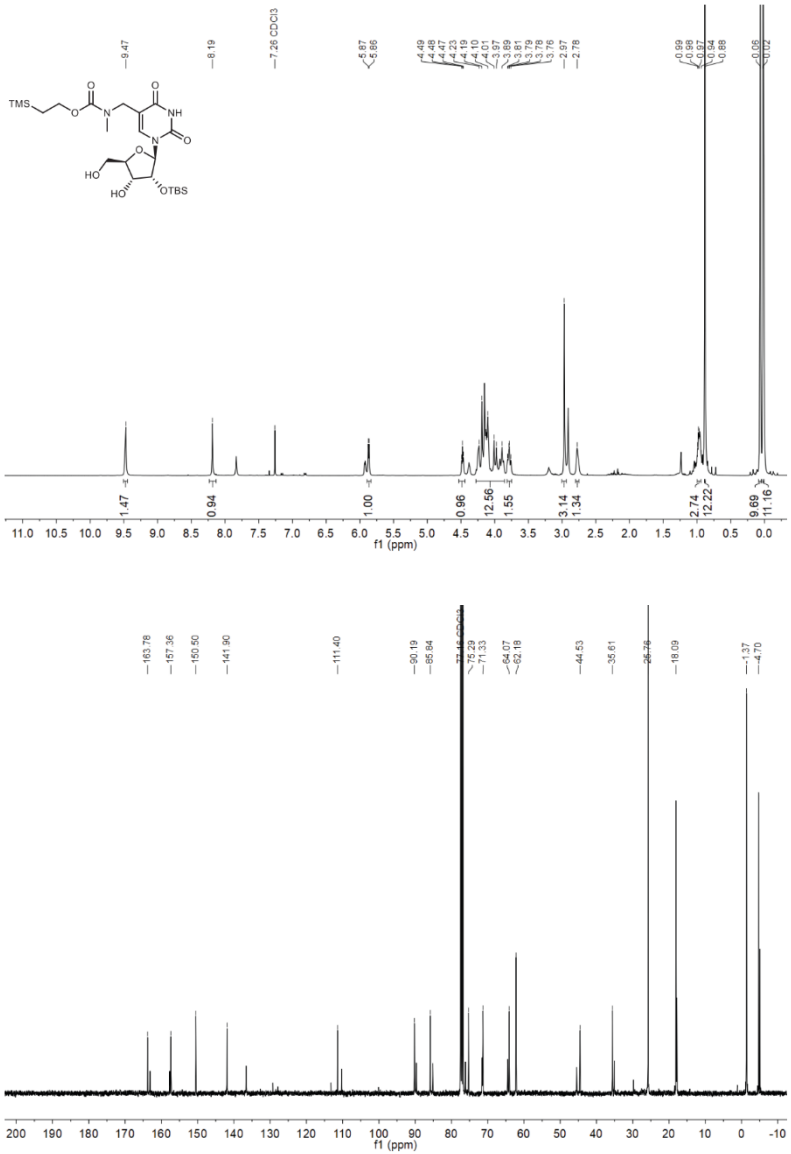


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4c

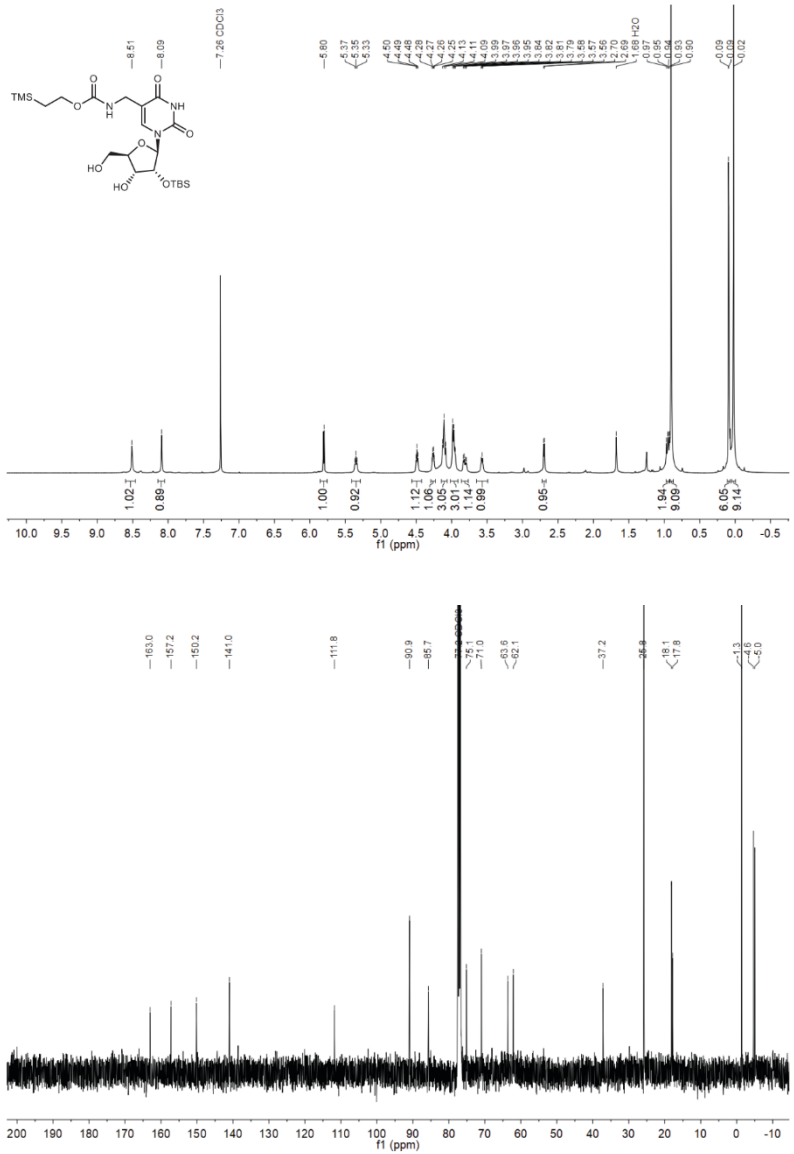




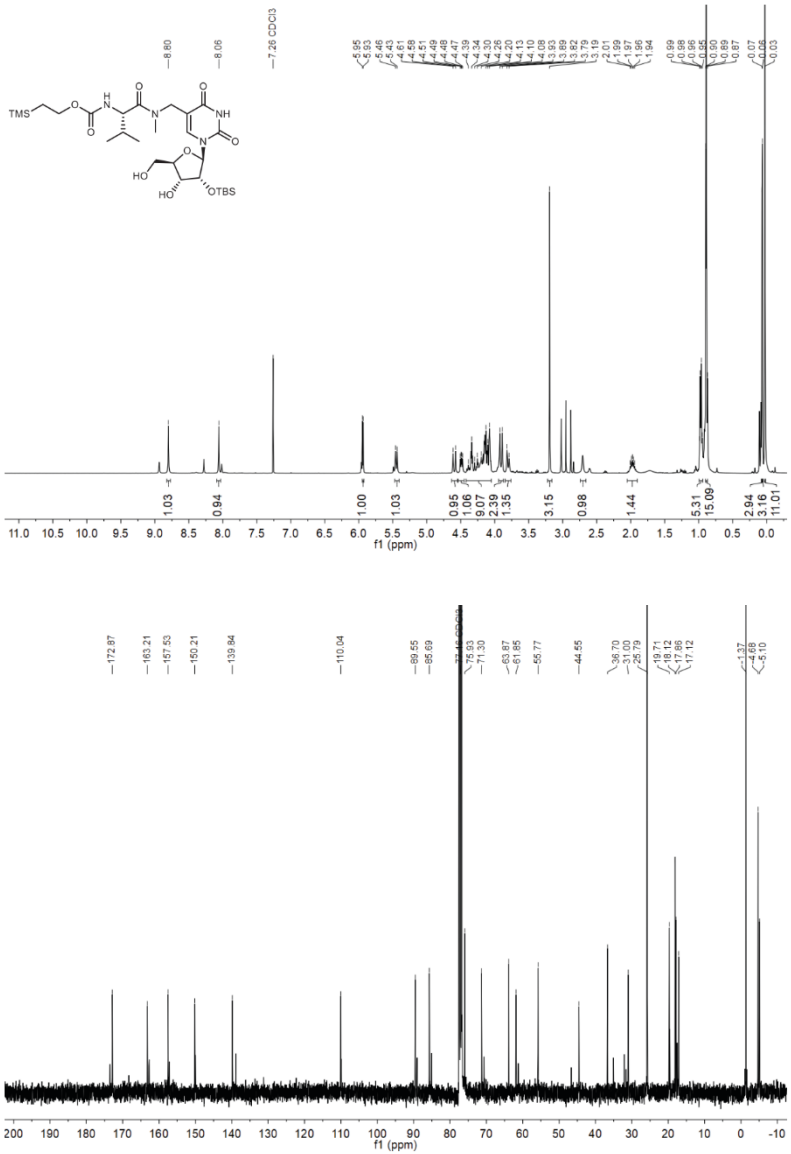
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5a



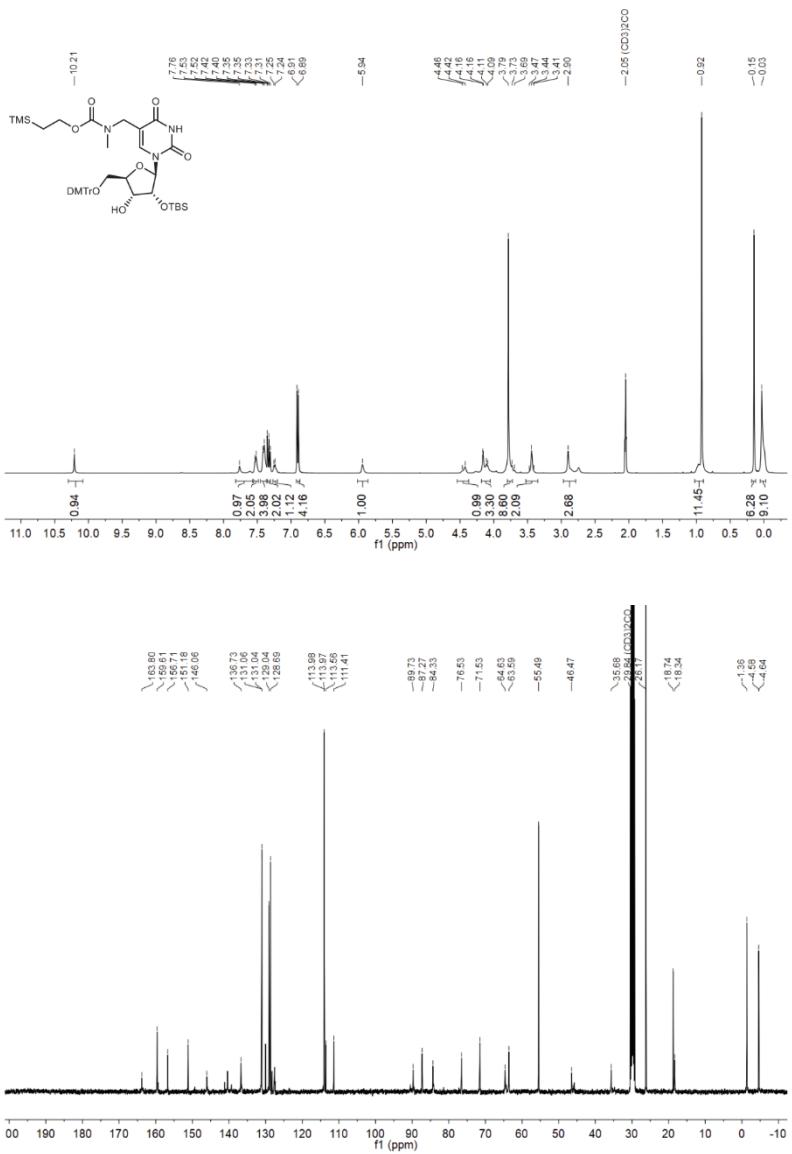
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5b



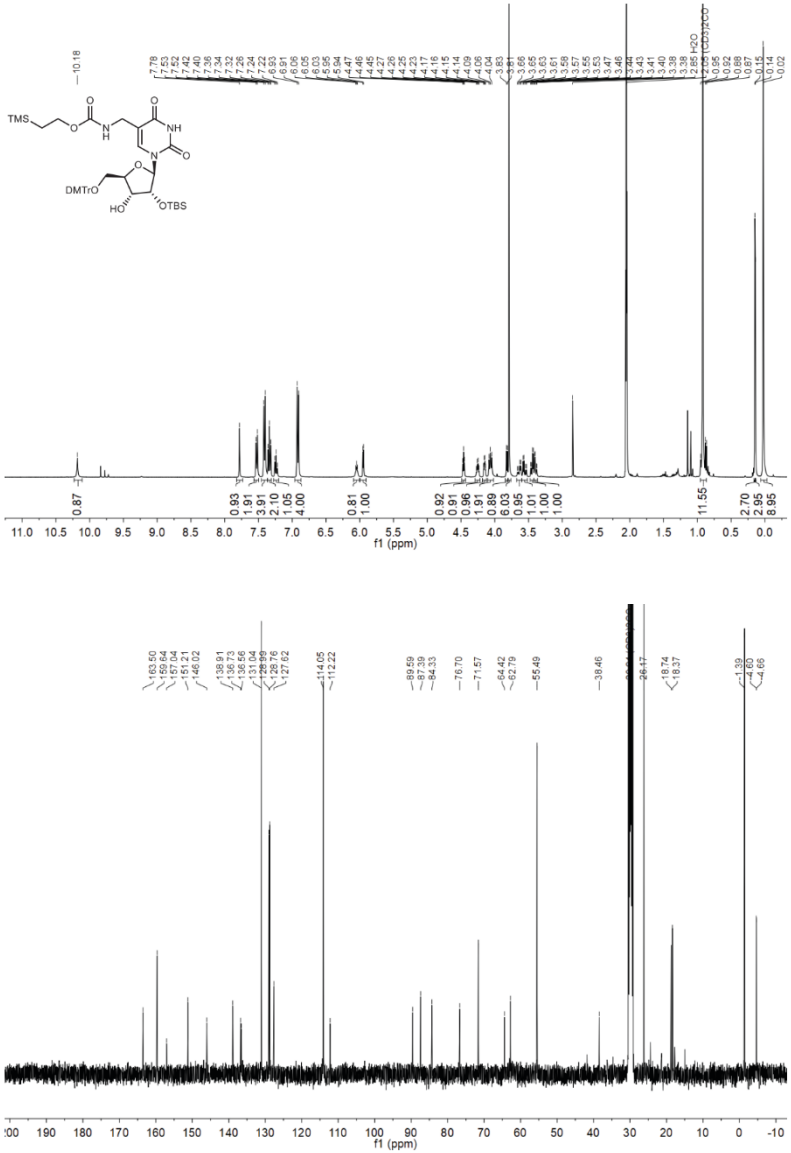
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5c



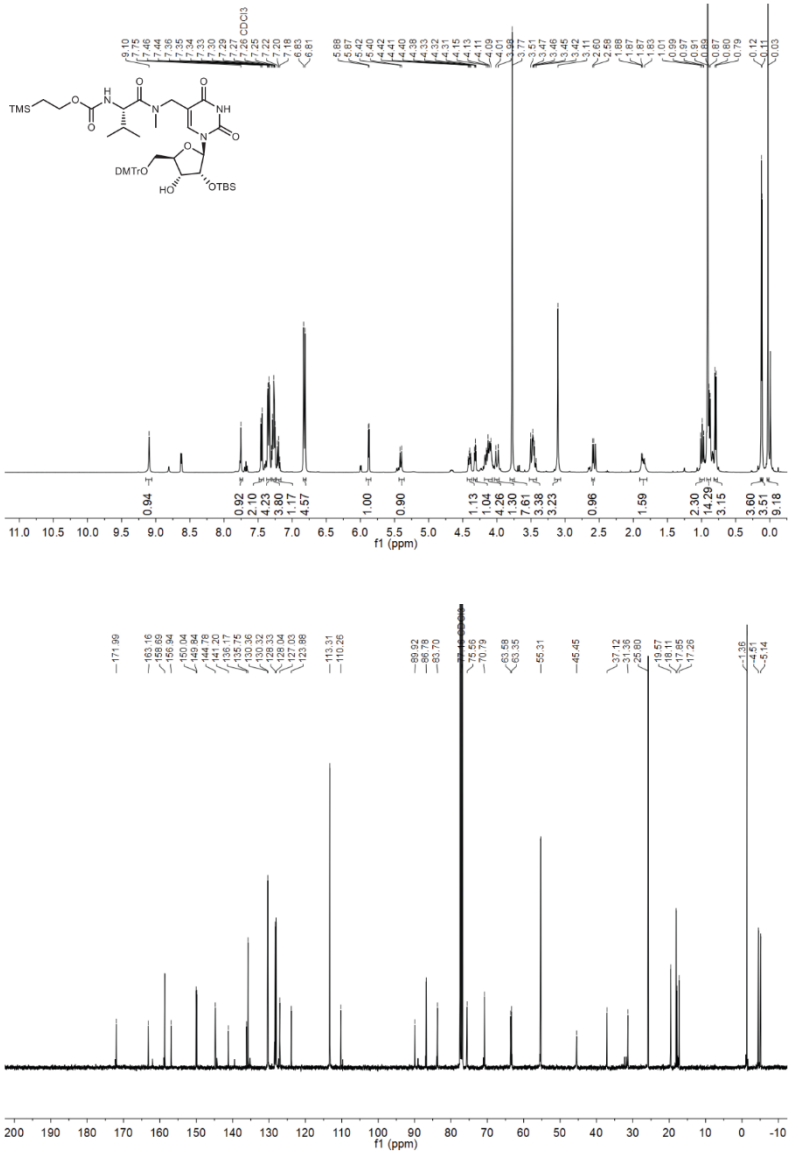
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6a



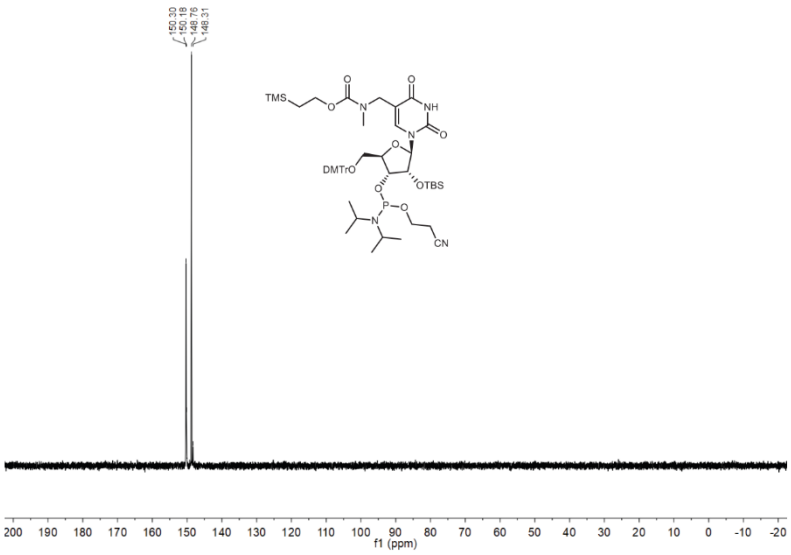
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6b



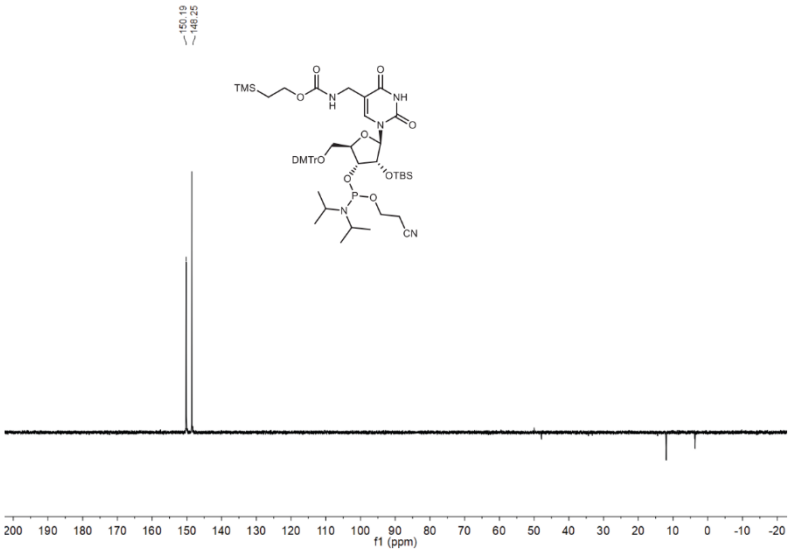
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6c



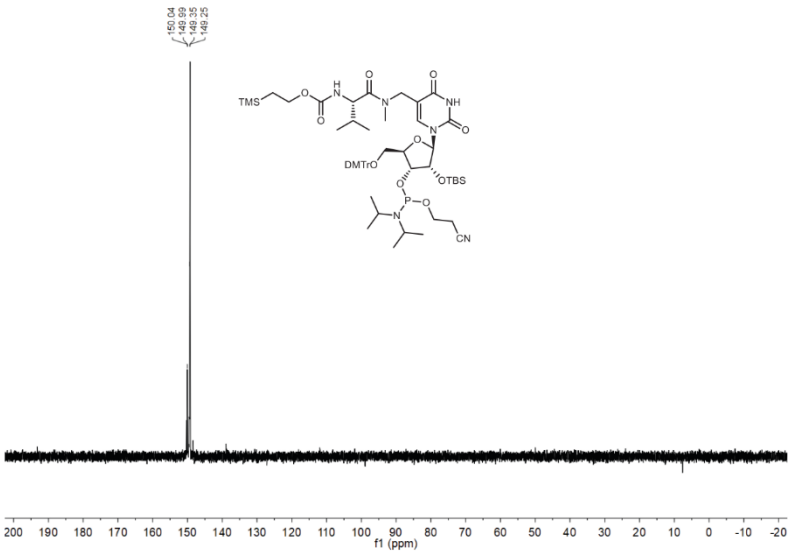
$^{31}P\{^{1}H\}$ NMR spectrum of compound 7a



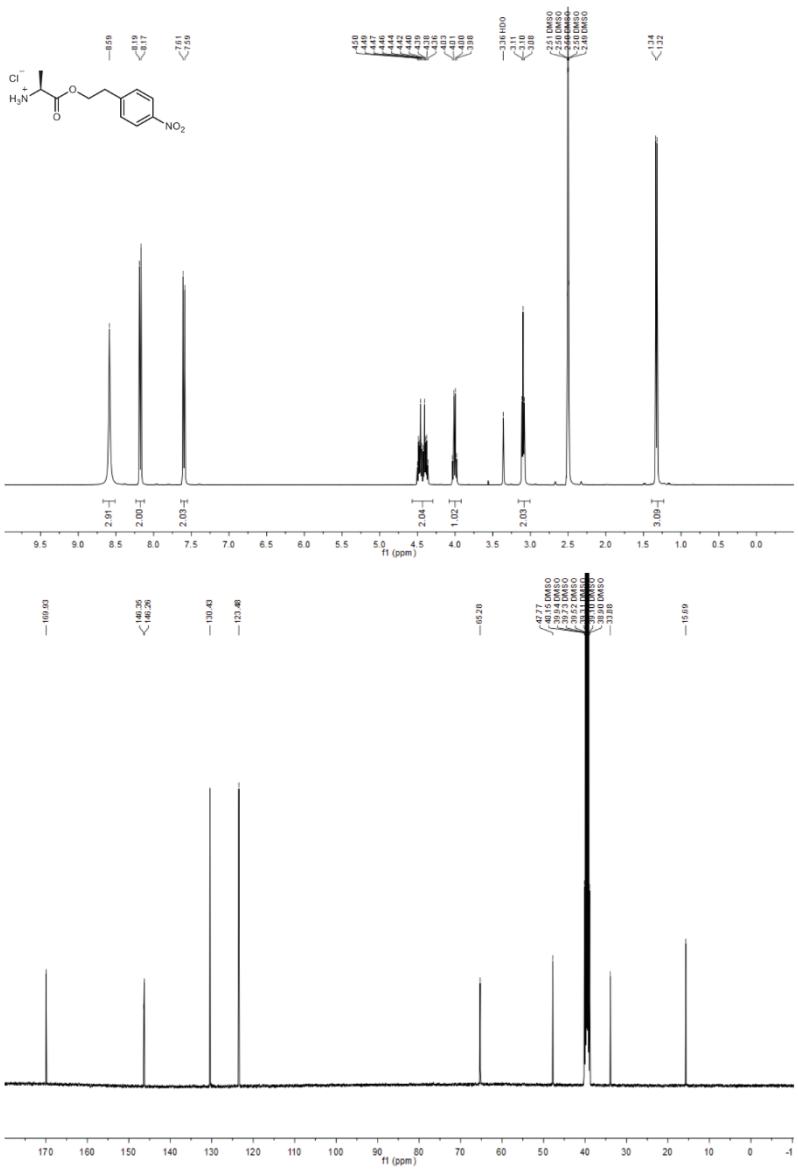
$^{31}P\{^{1}H\}$ NMR spectrum of compound 7b



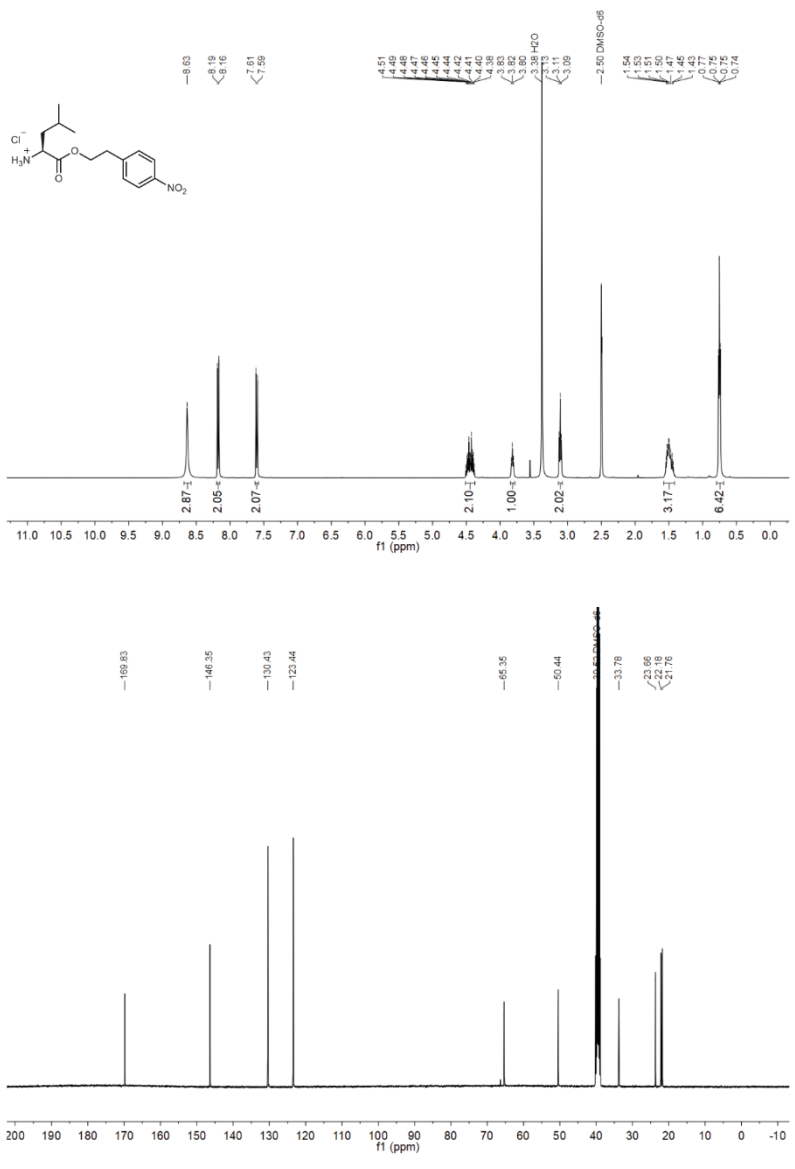
$^{31}P\{^{1}H\}$ NMR spectrum of compound 7c



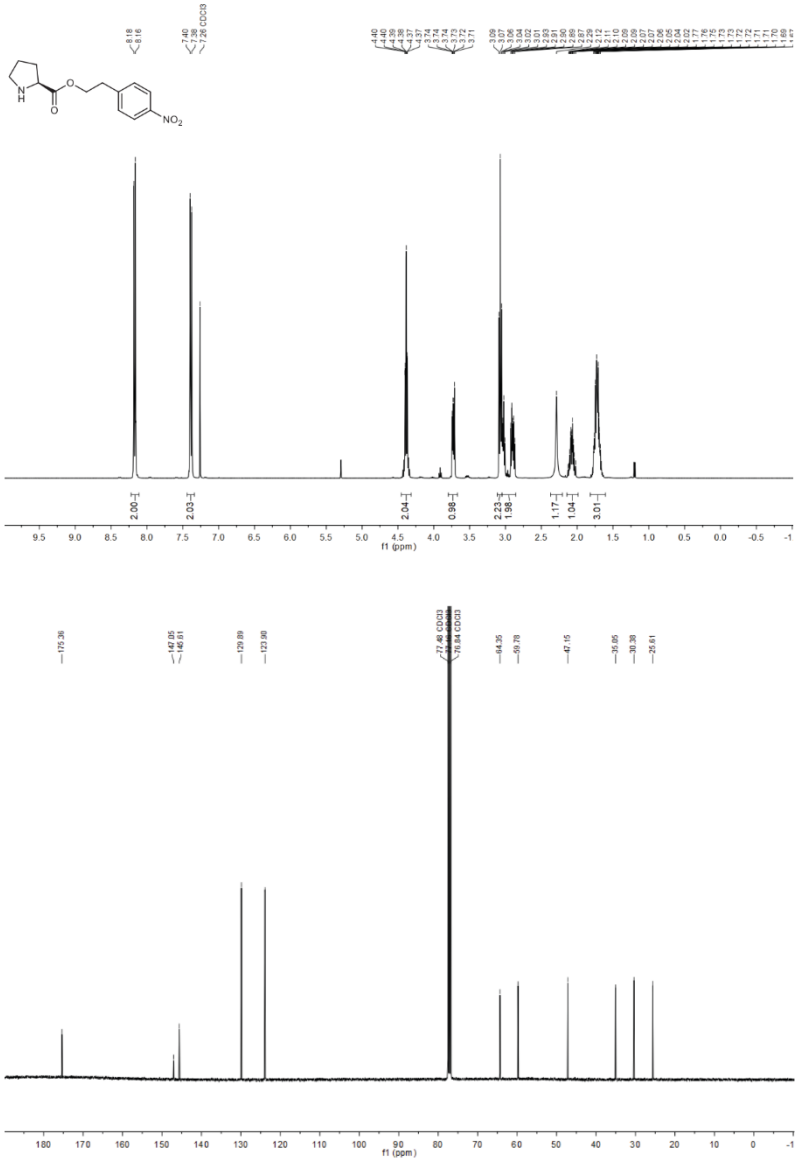
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Ala-Onpe•HCl



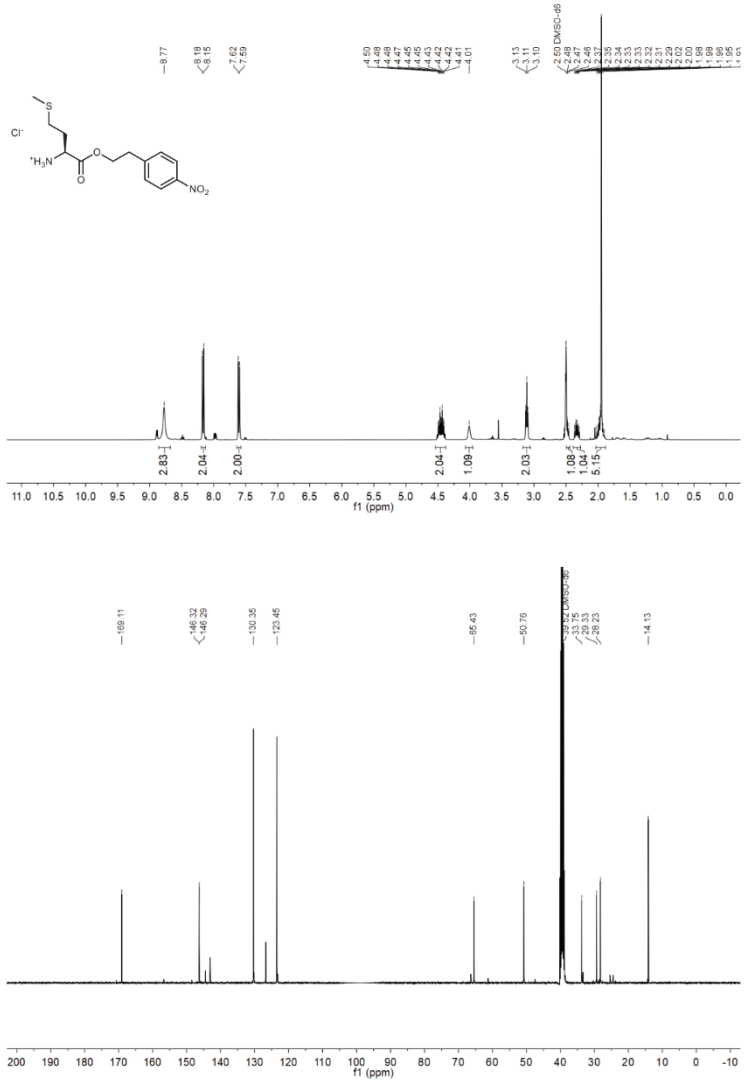
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Leu-Onpe•HCl



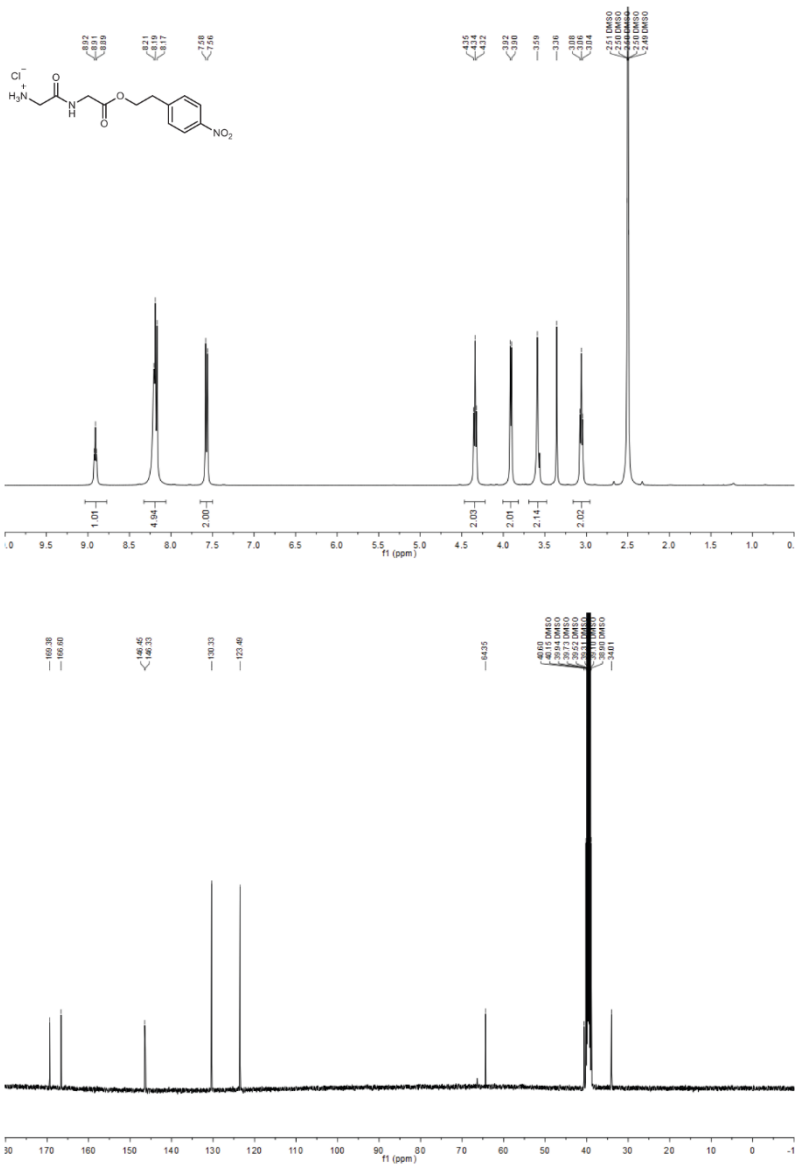
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Pro-Onpe



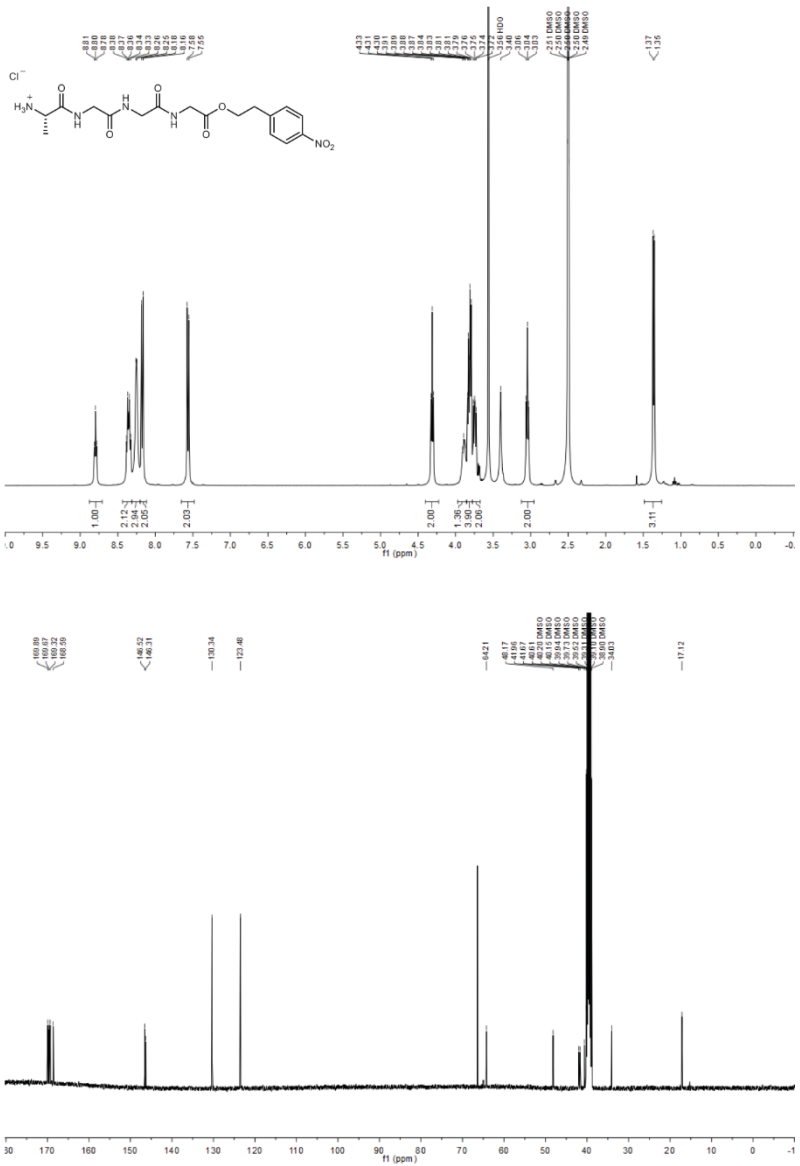
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Met-Onpe•HCl



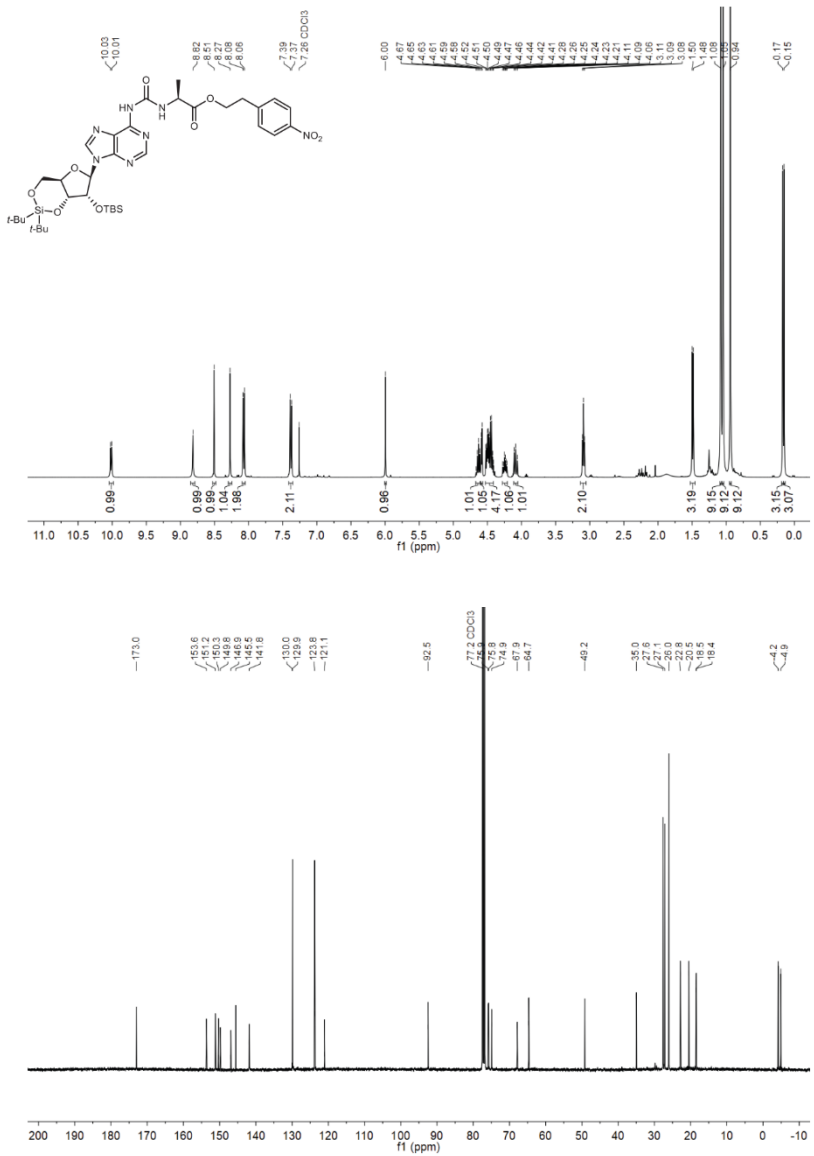
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Gly-Gly-Onpe•HCl



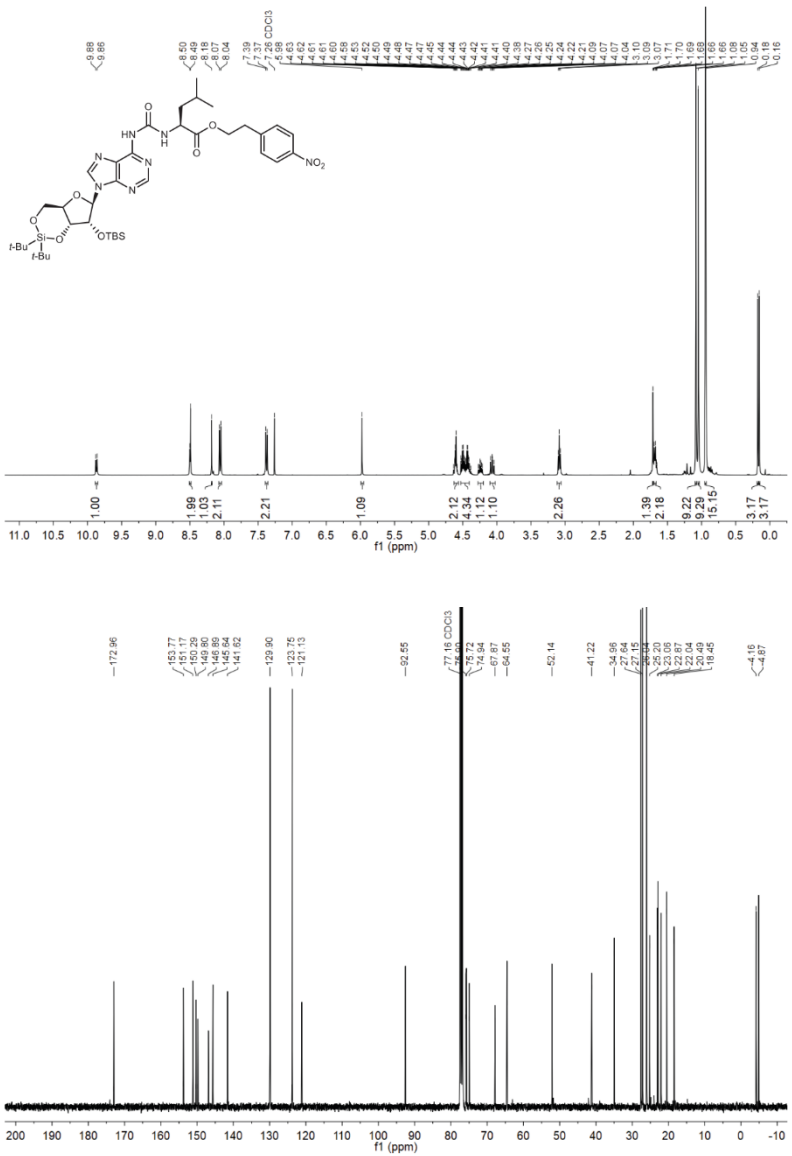
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Ala-Gly-Gly-Gly-Onpe•HCl



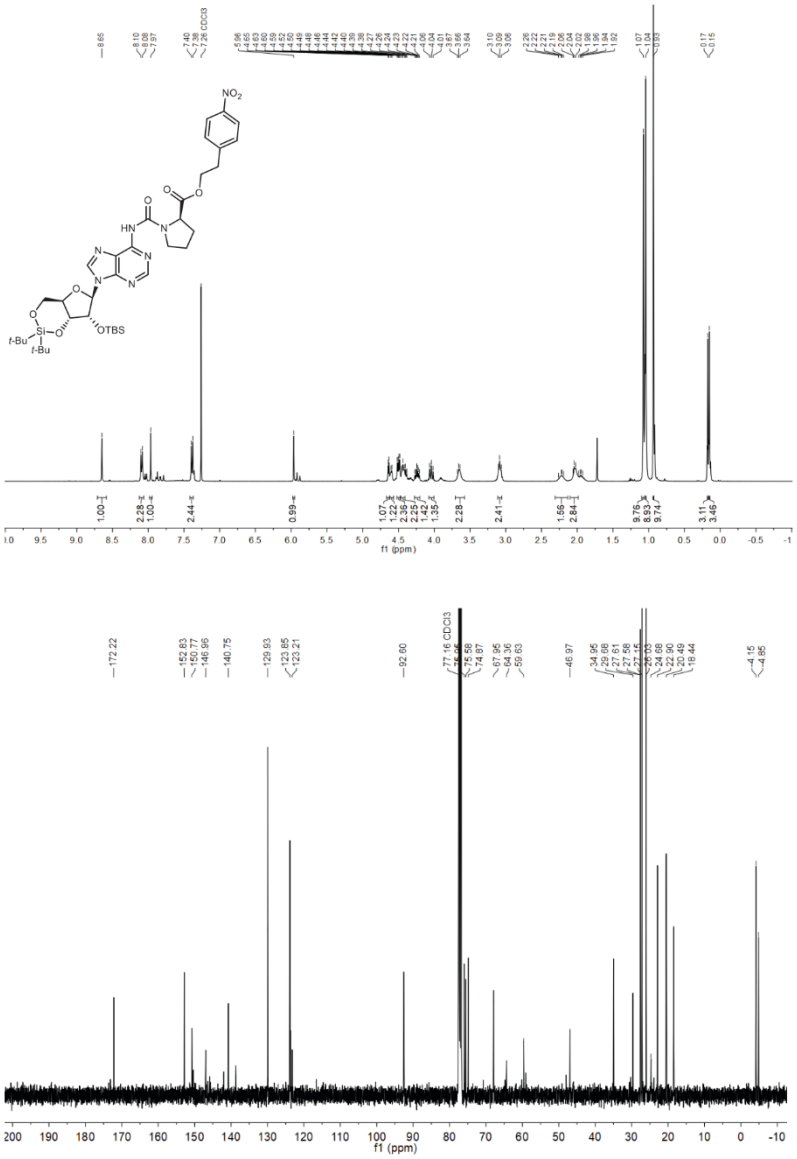
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10b



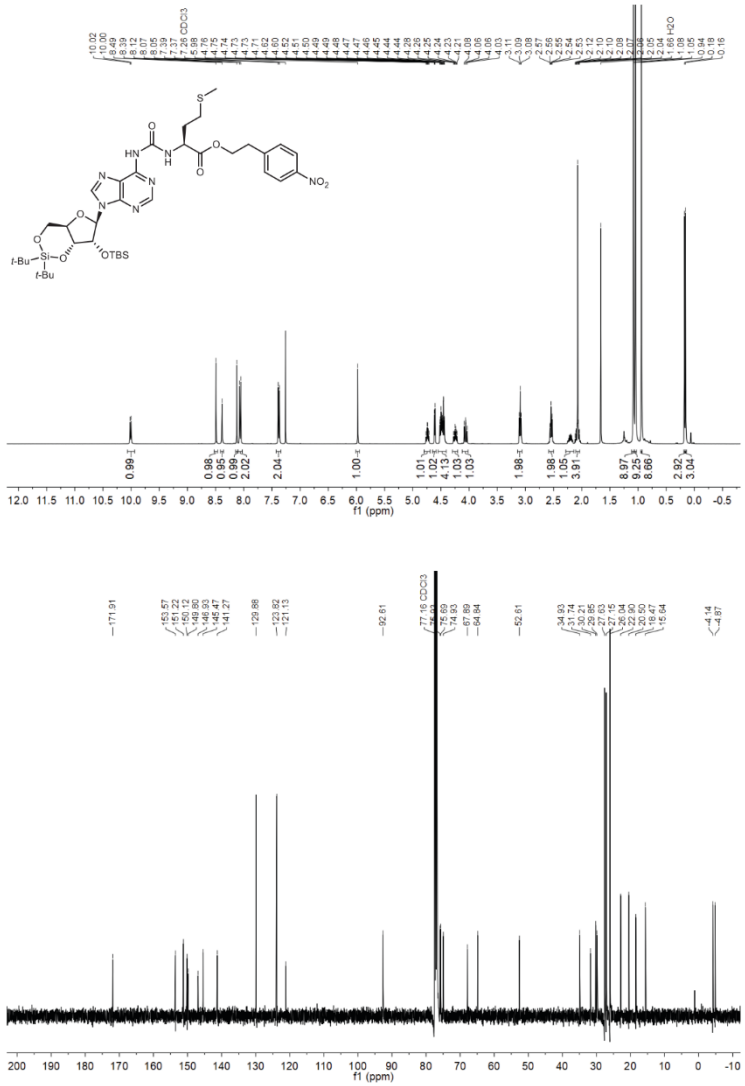
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10d



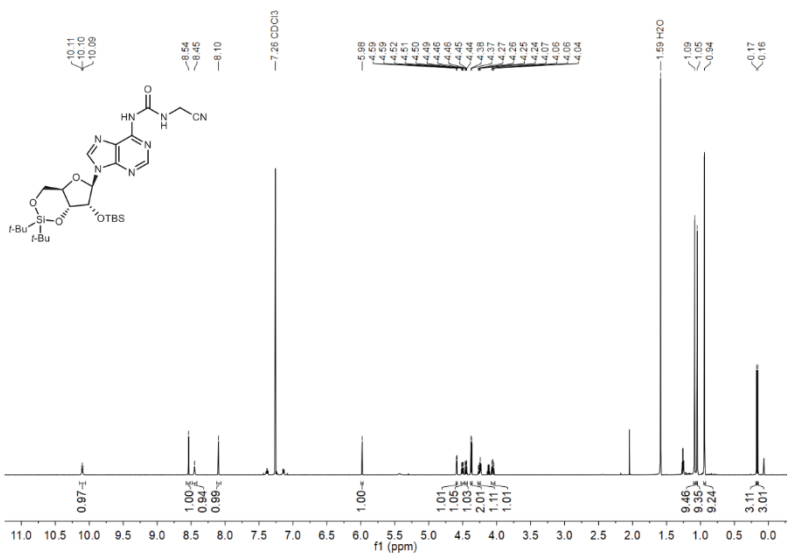
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10f

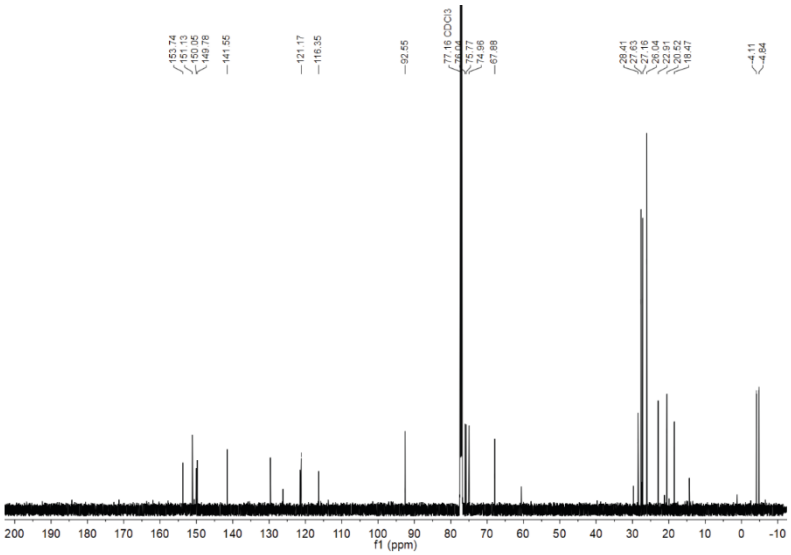


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10h

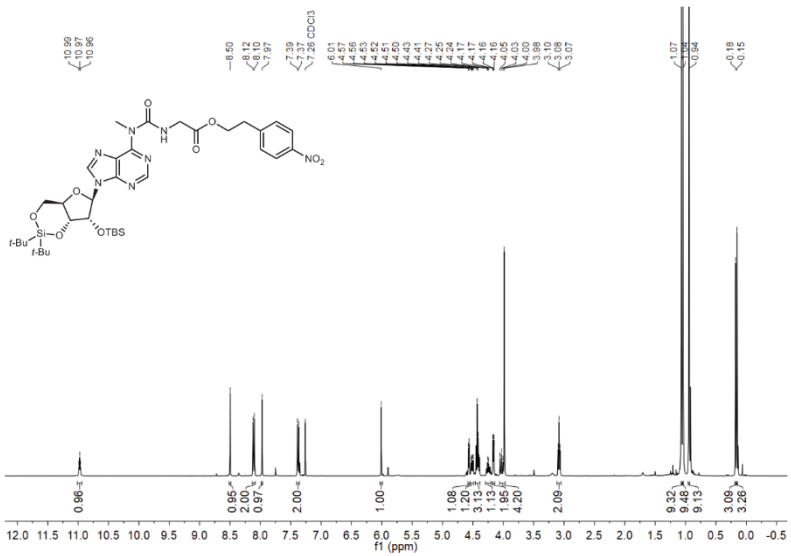


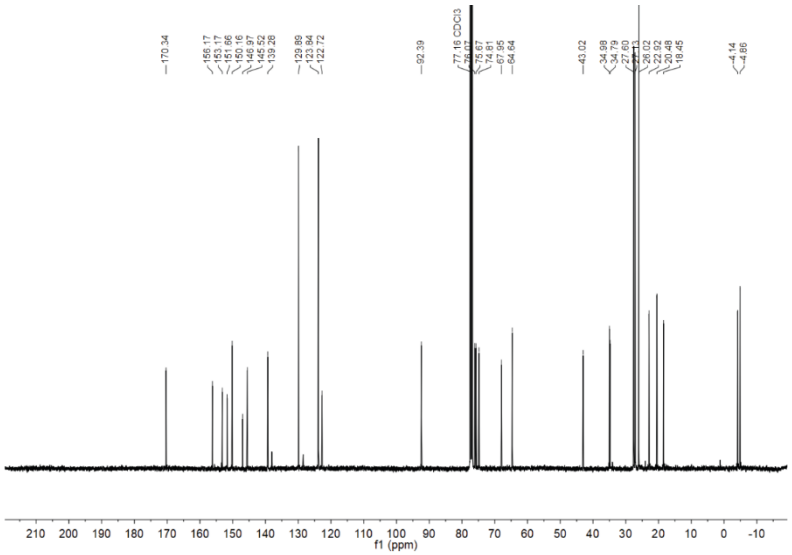
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10j



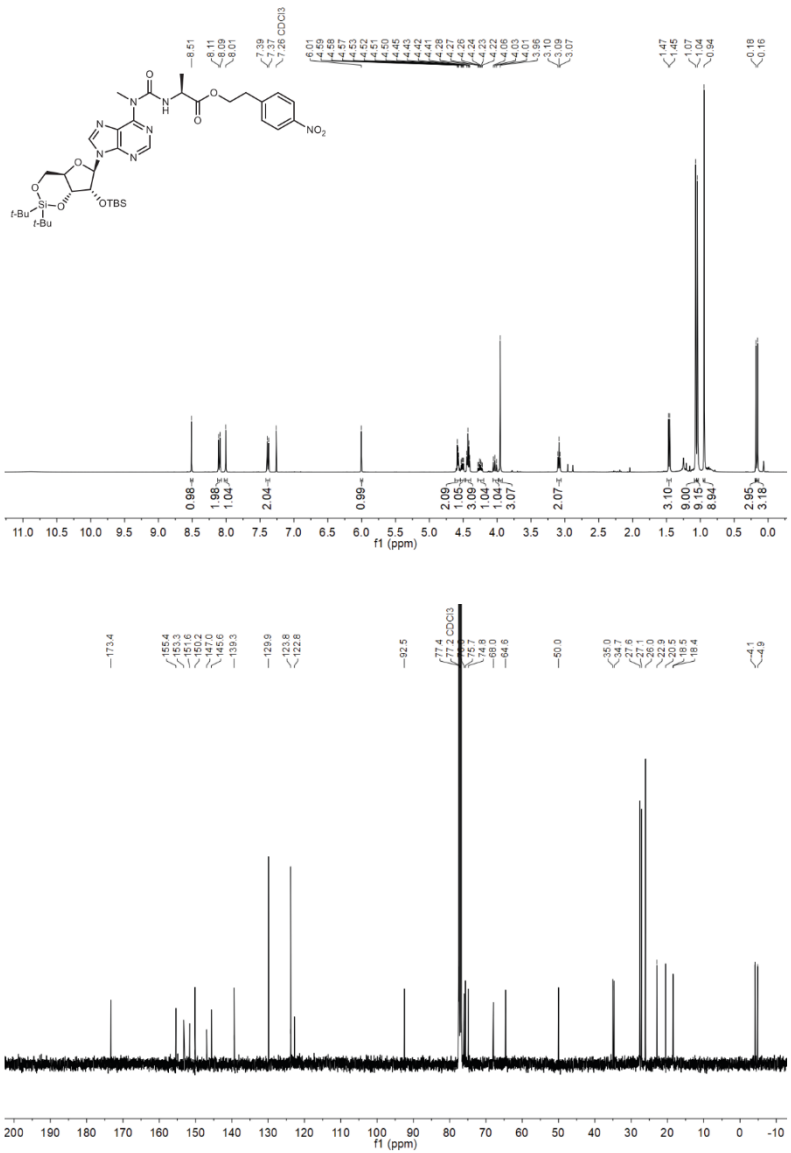


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11a

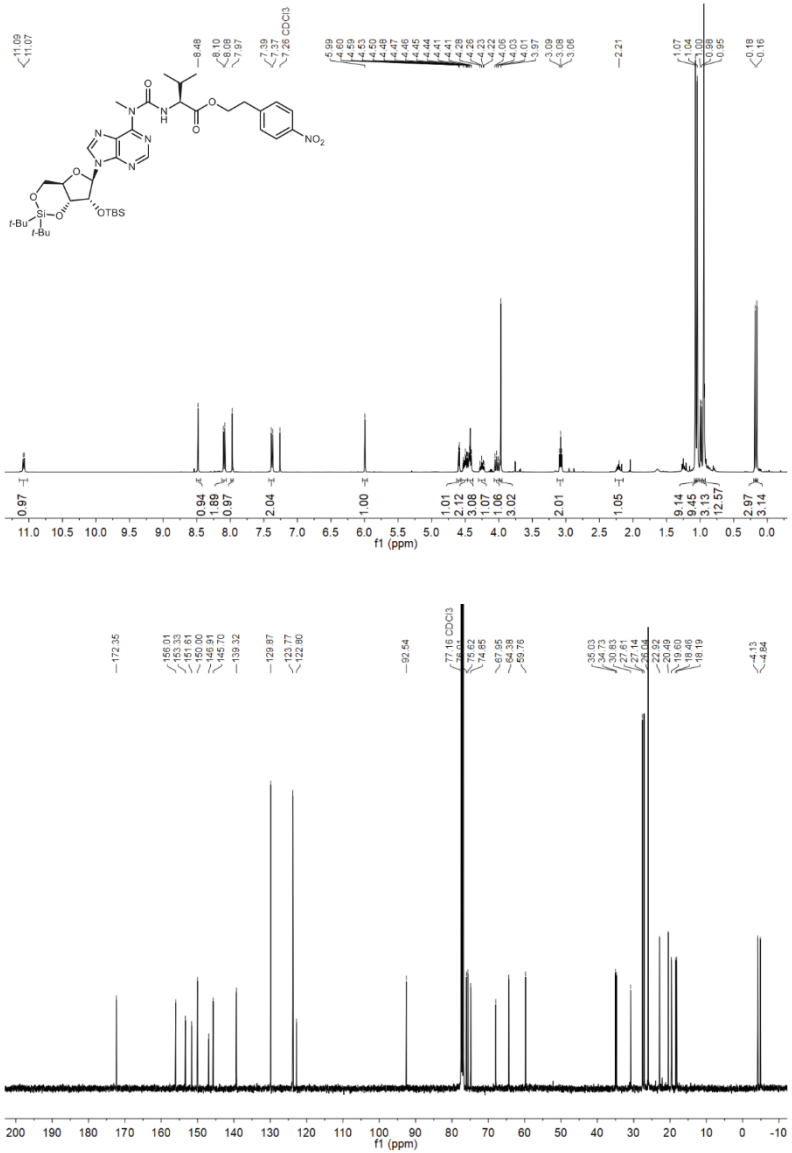




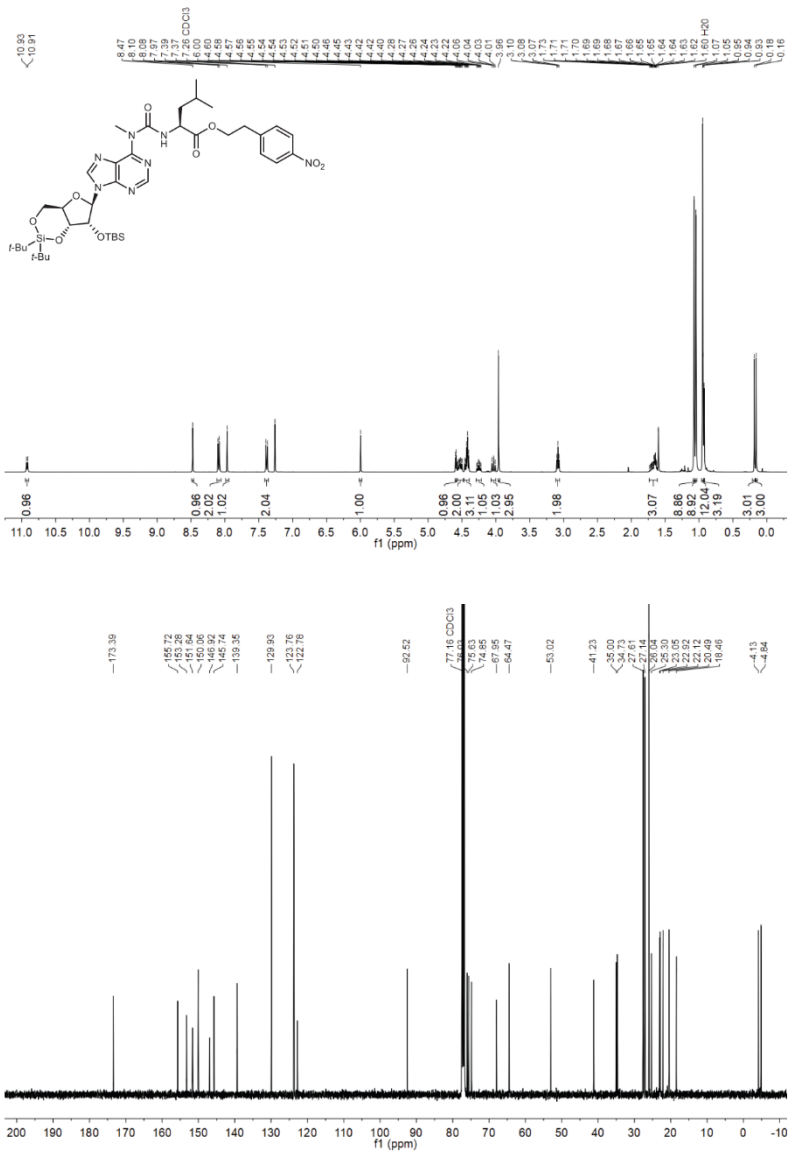
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11b



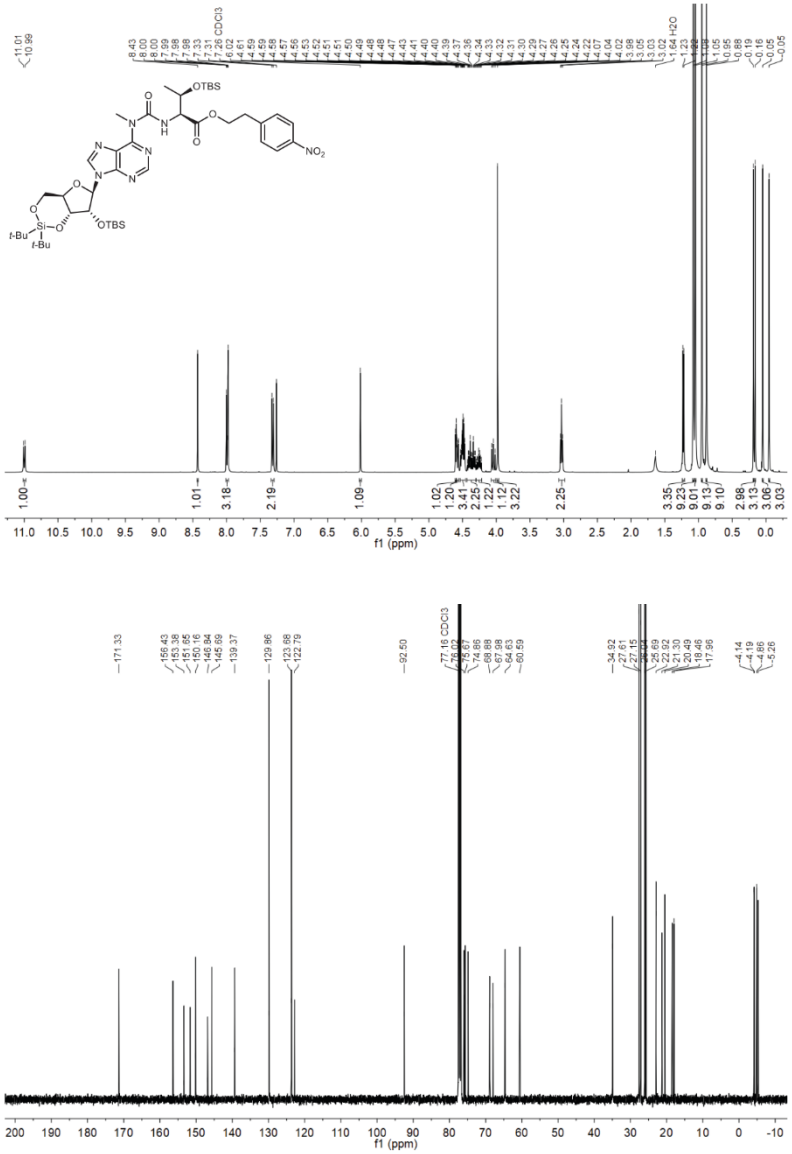
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11c



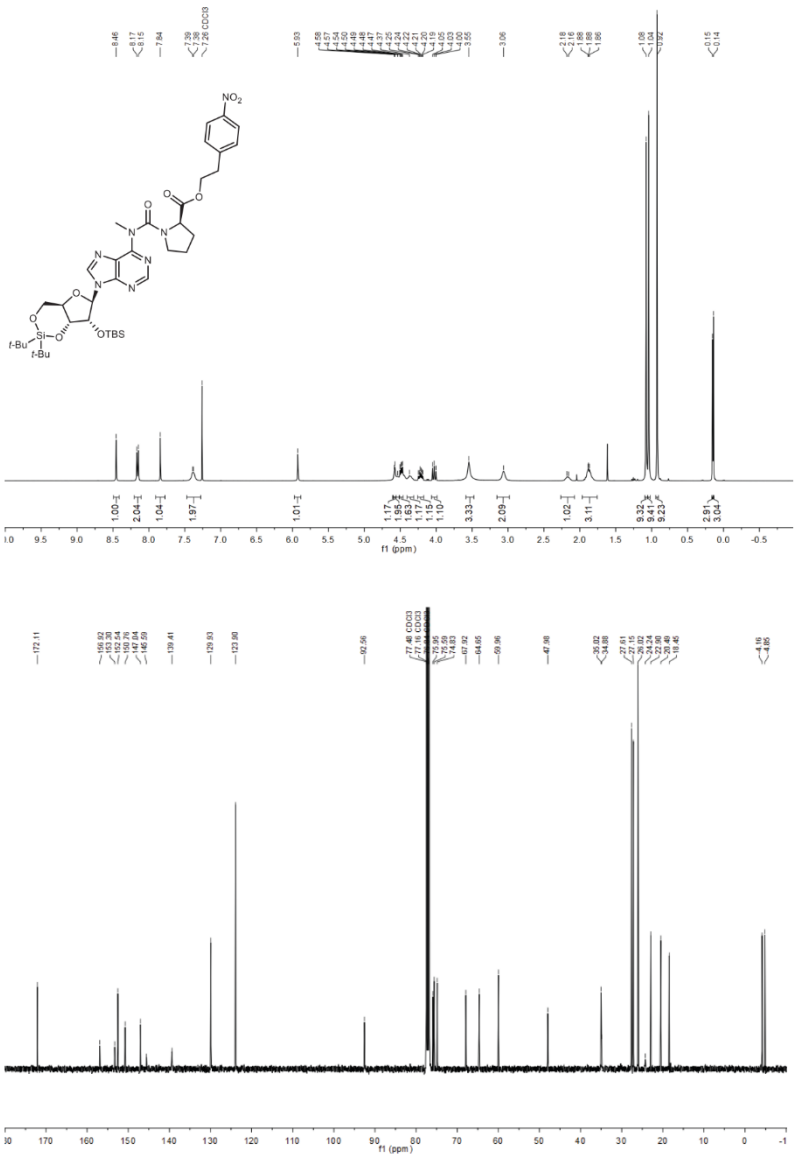
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11d



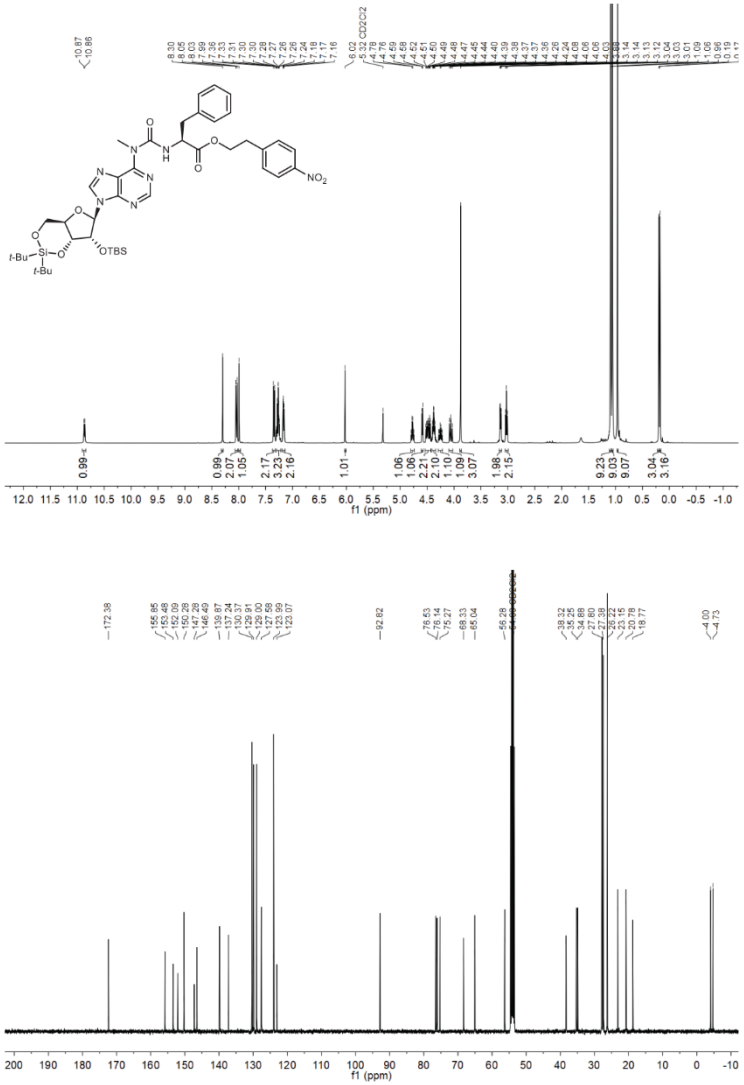
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11e



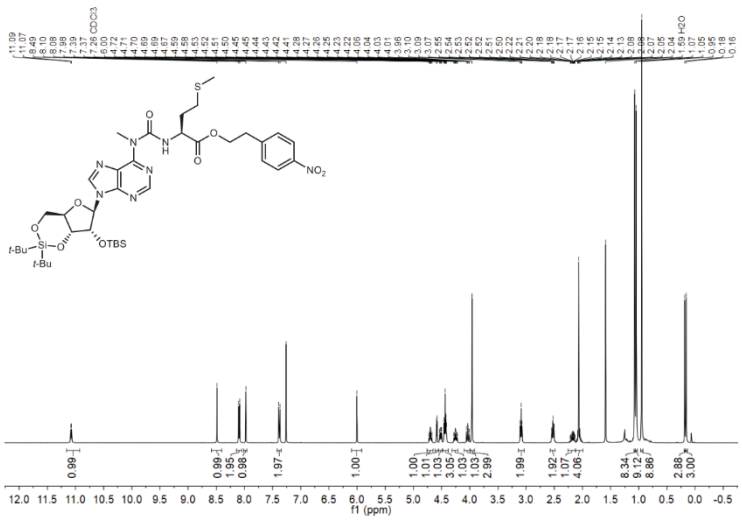
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11f

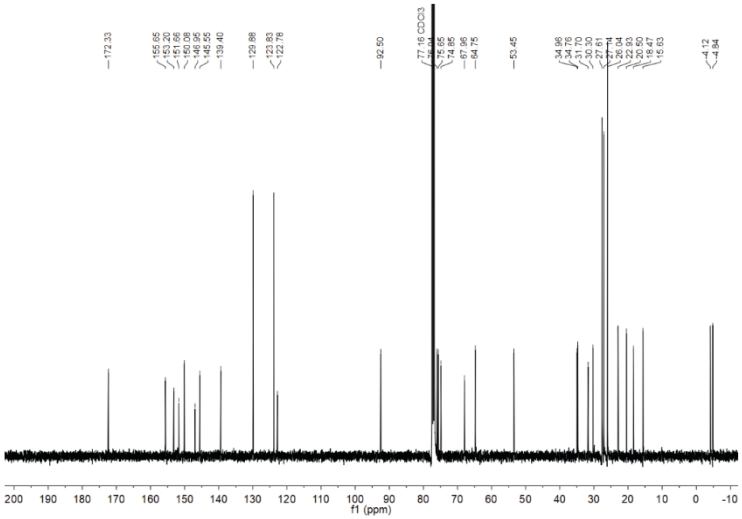


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11g

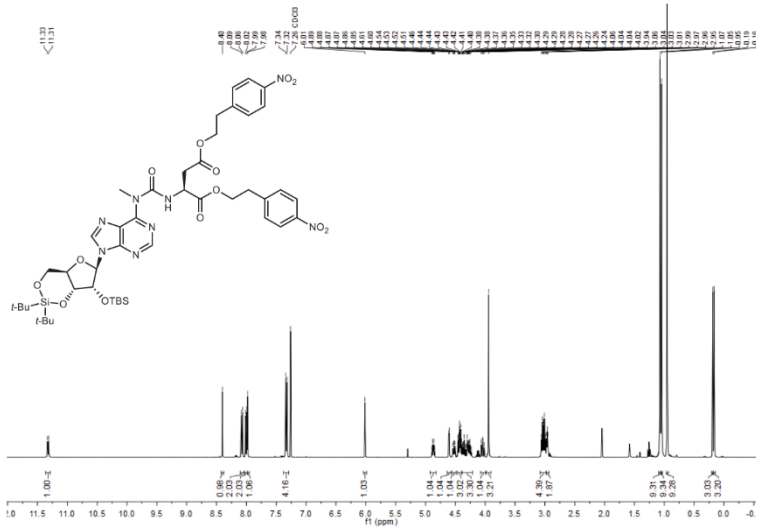


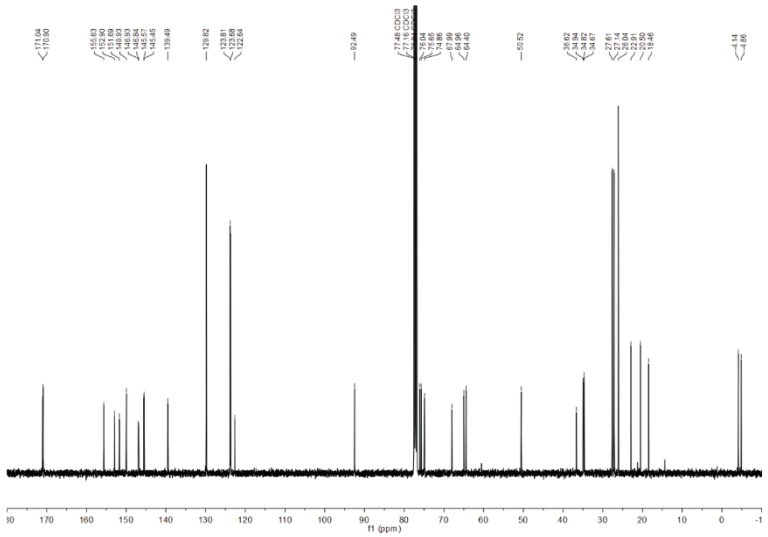
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11h



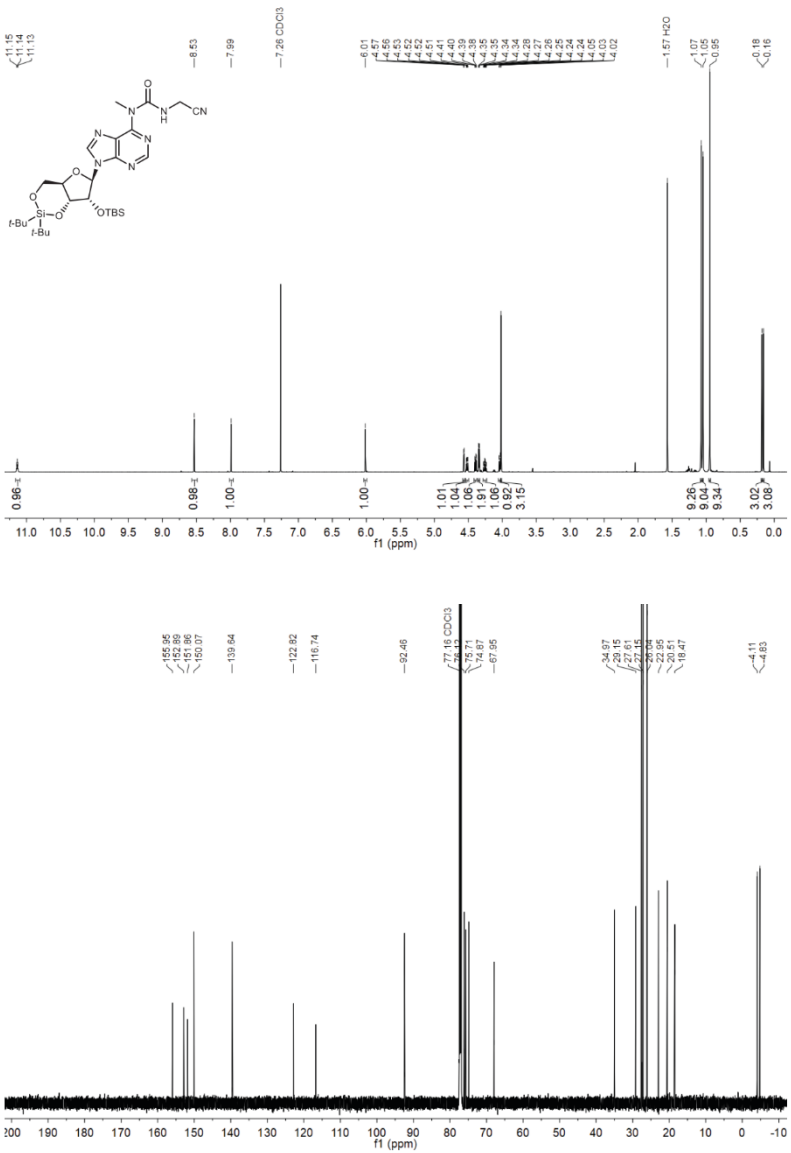


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11i

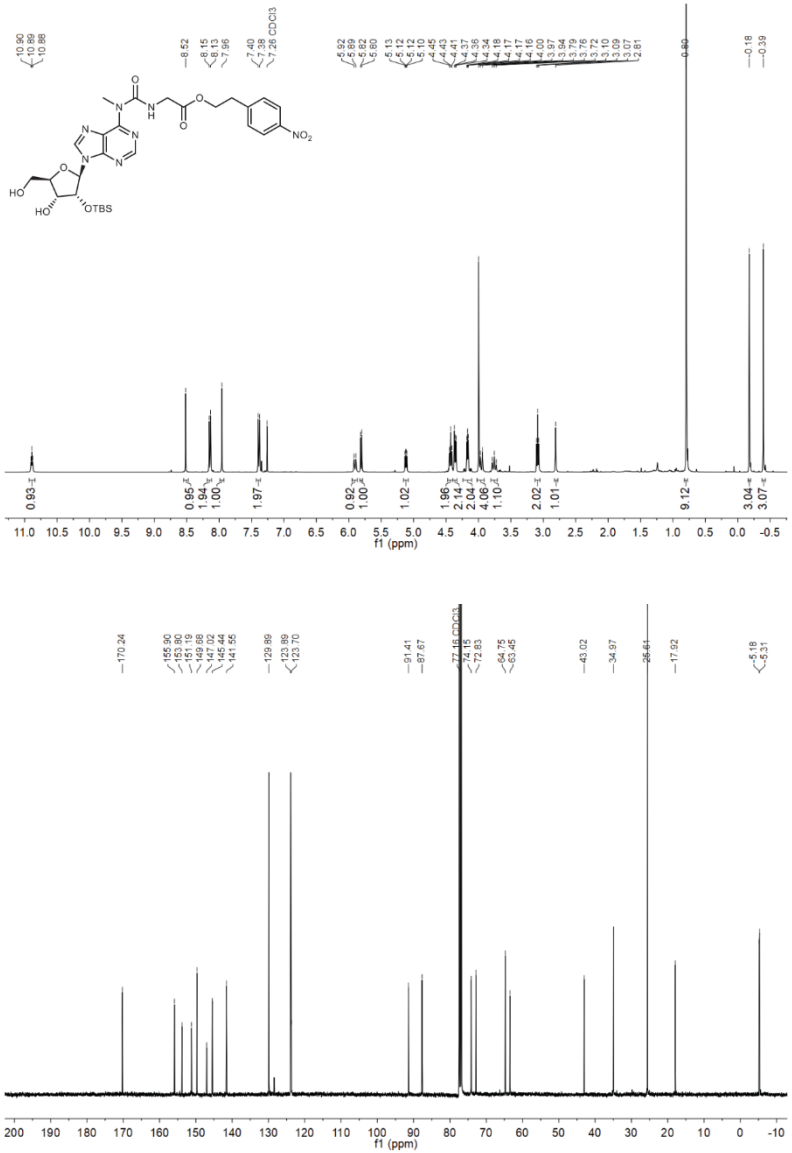




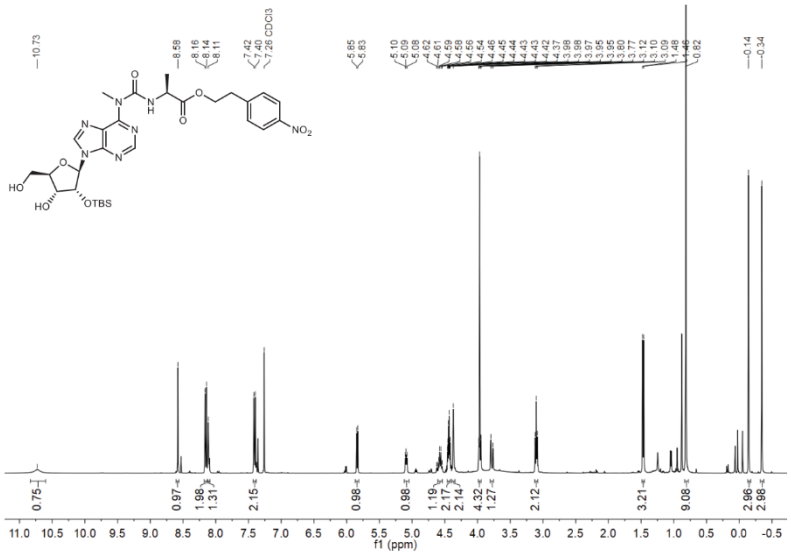
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11j

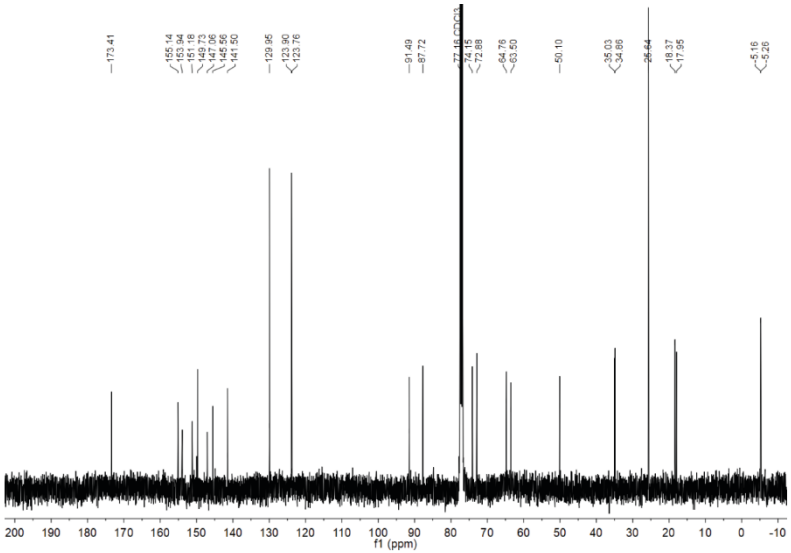


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12a

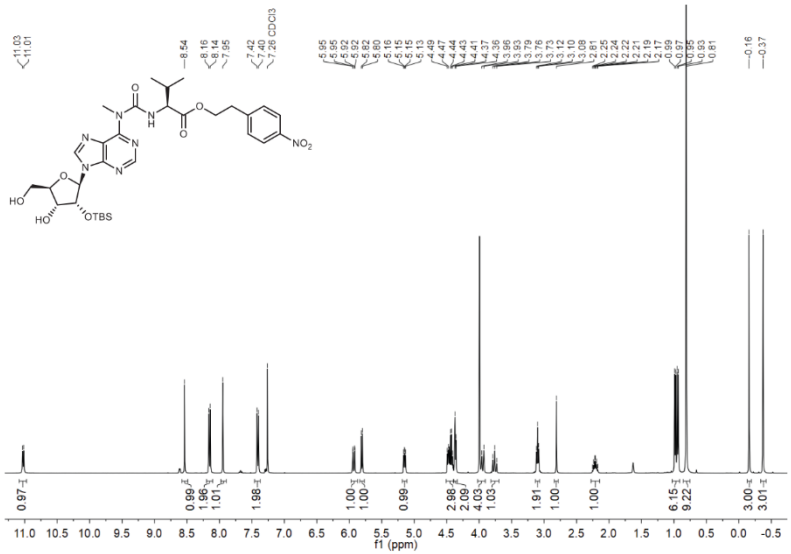


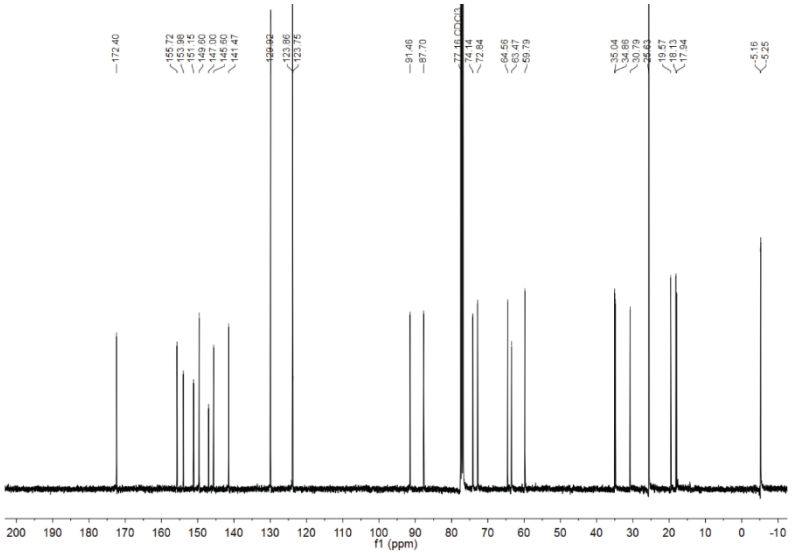
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12b



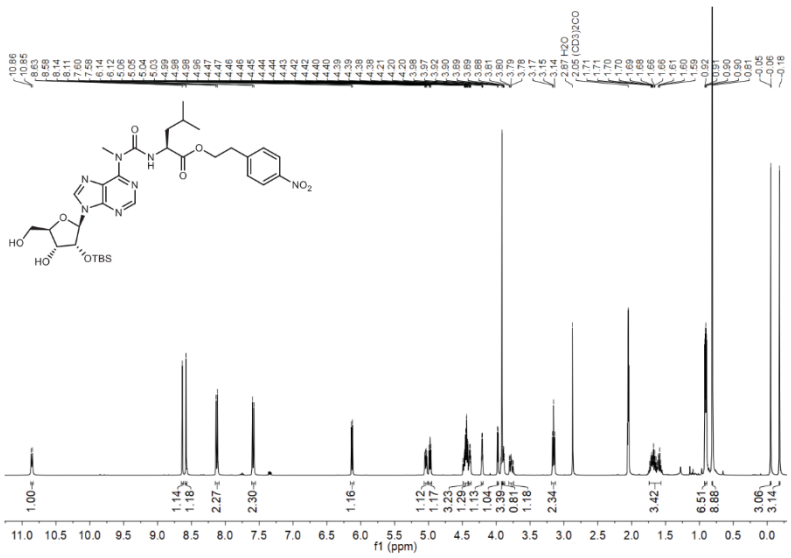


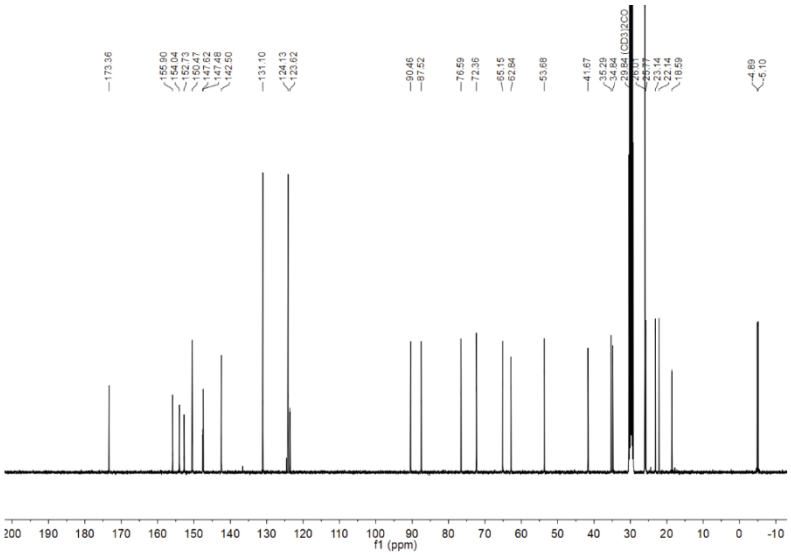
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12c



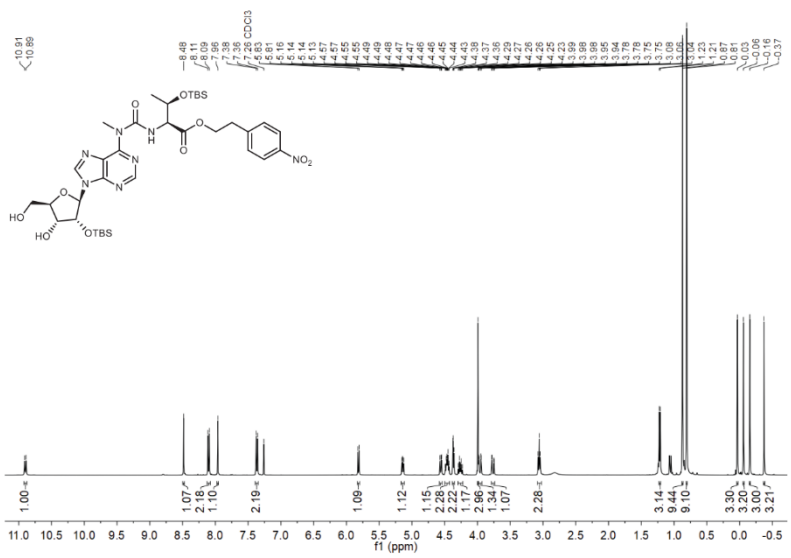


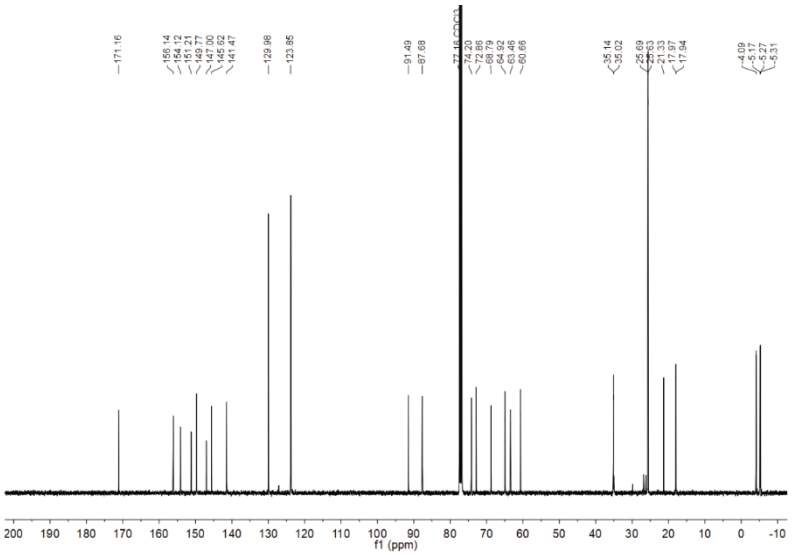
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12d



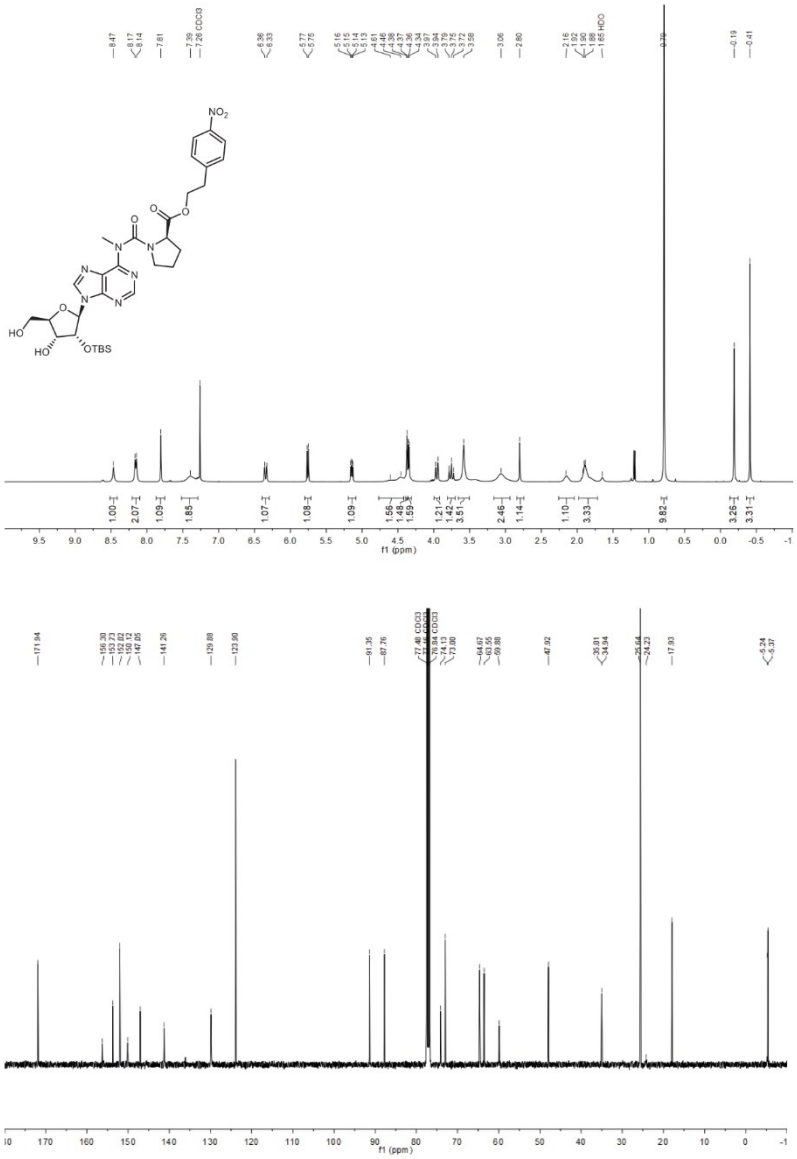


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12e

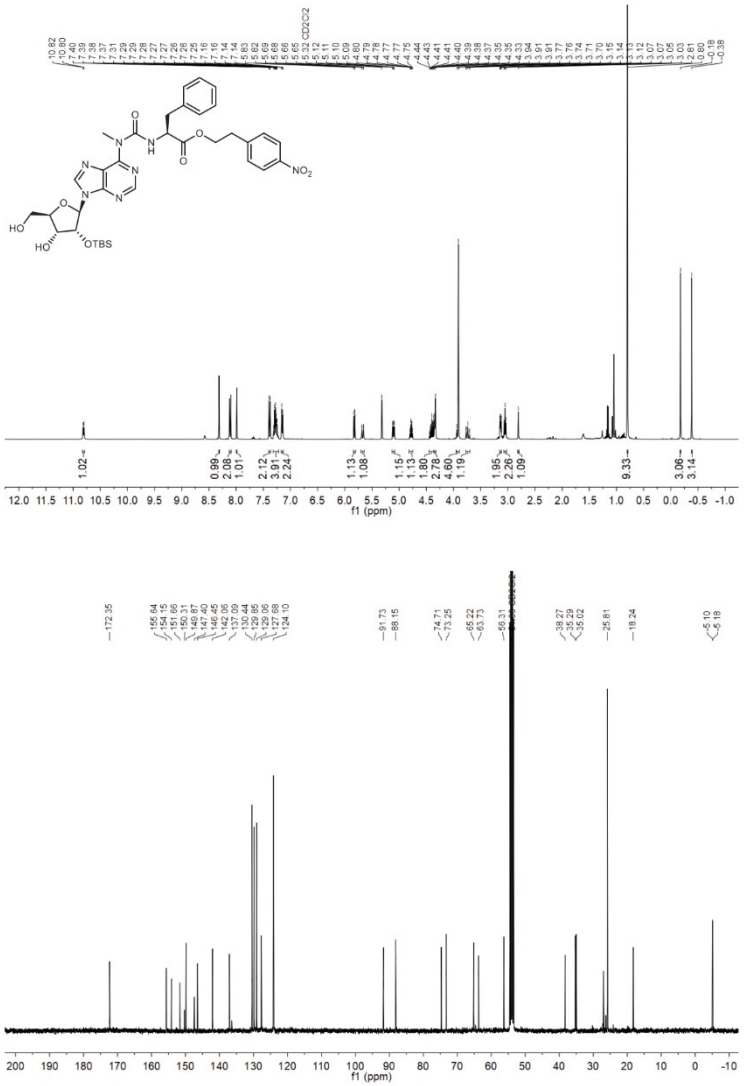




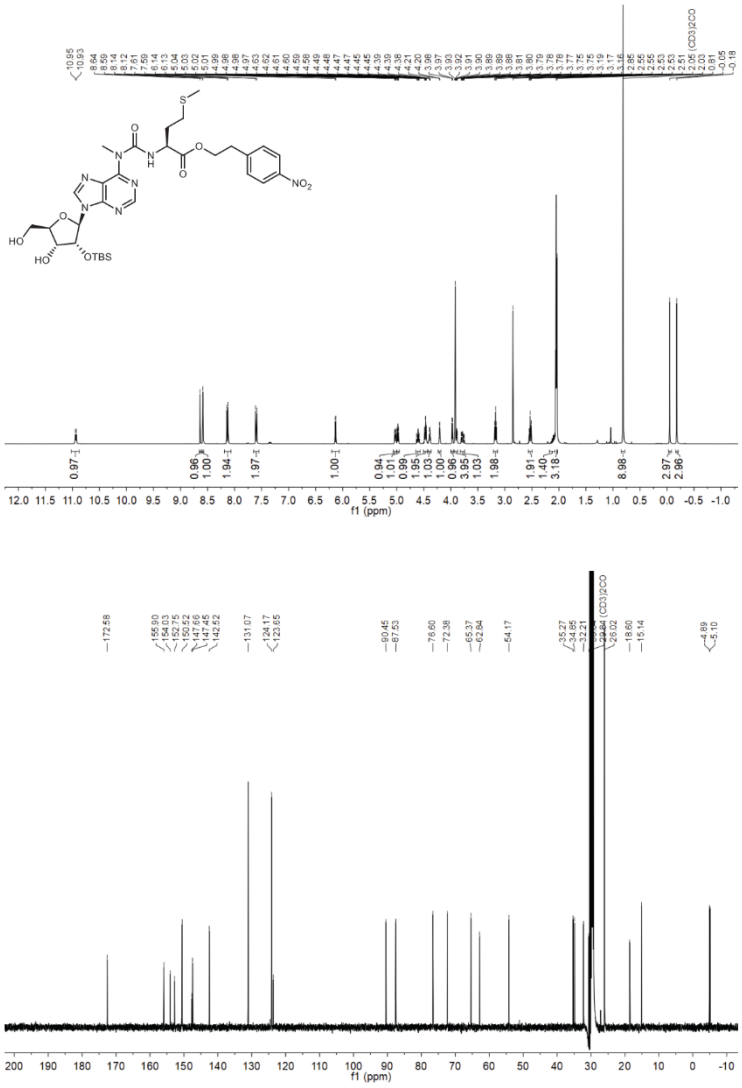
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12f



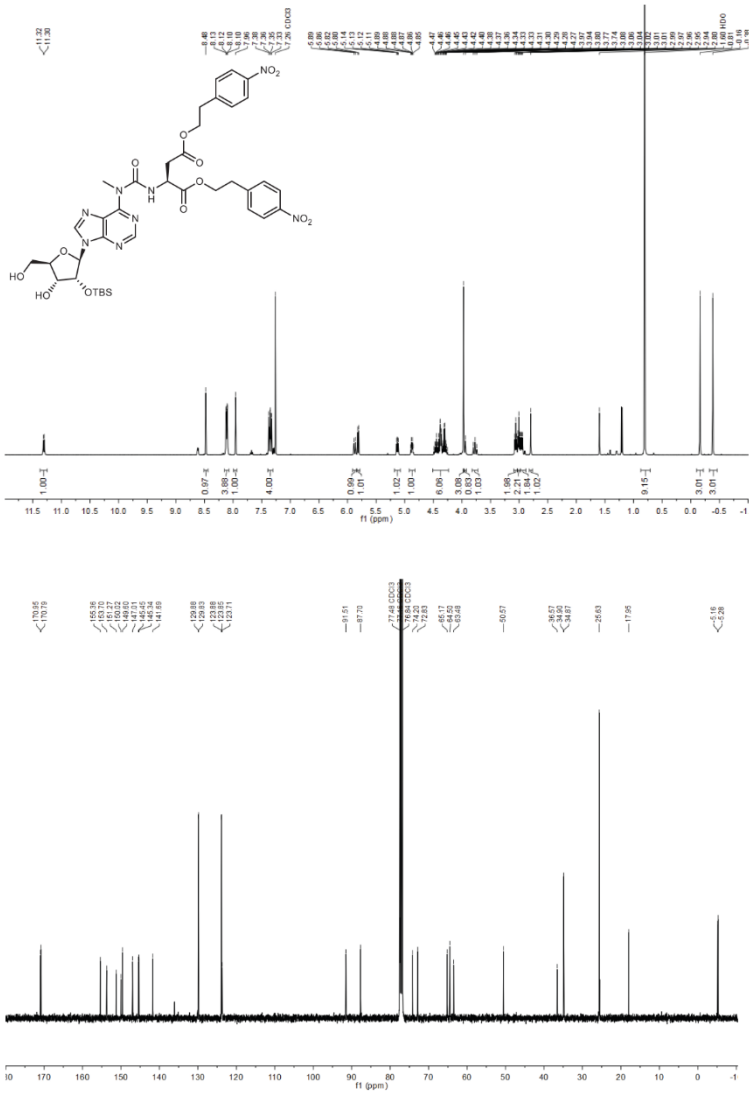
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12g



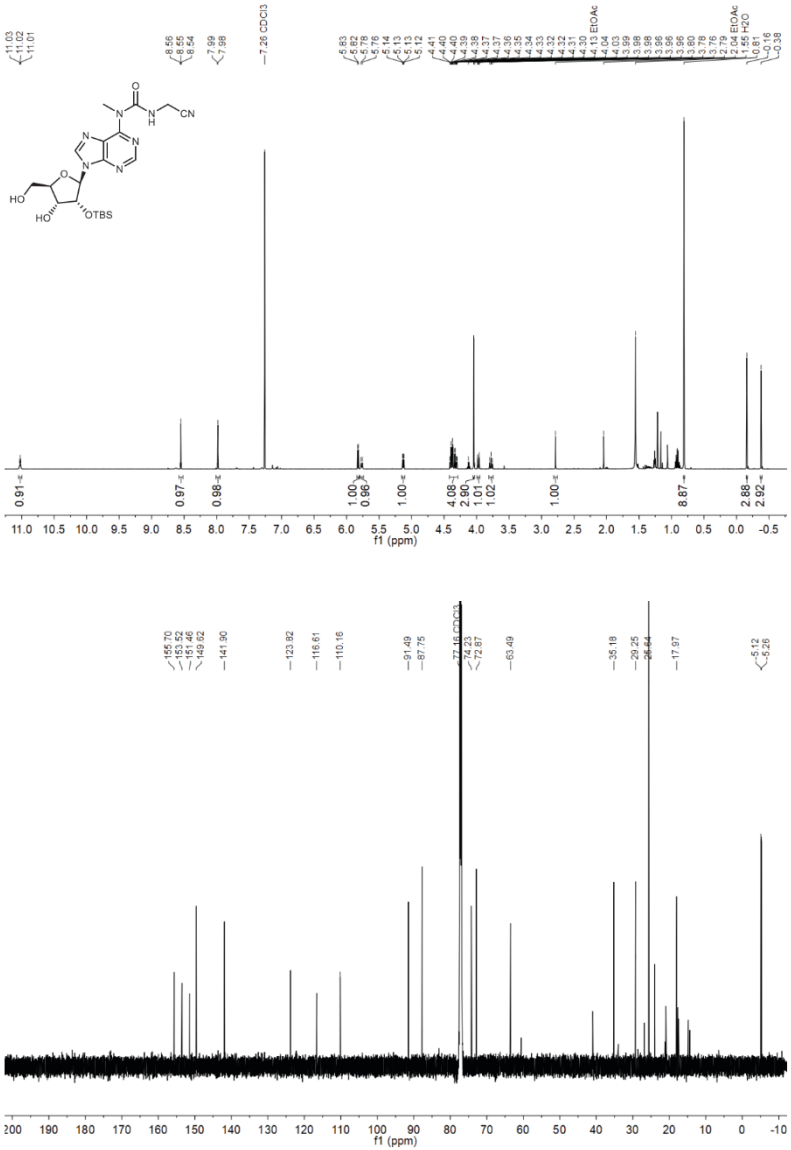
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12h



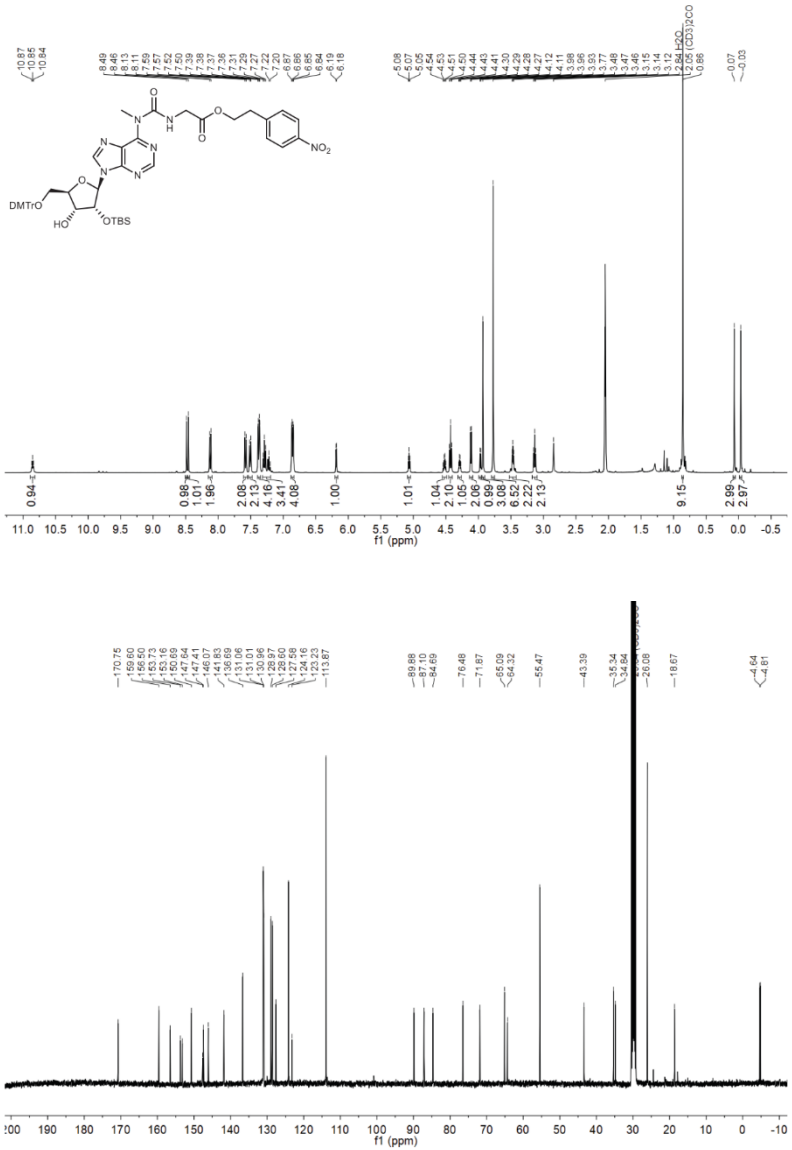
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12i



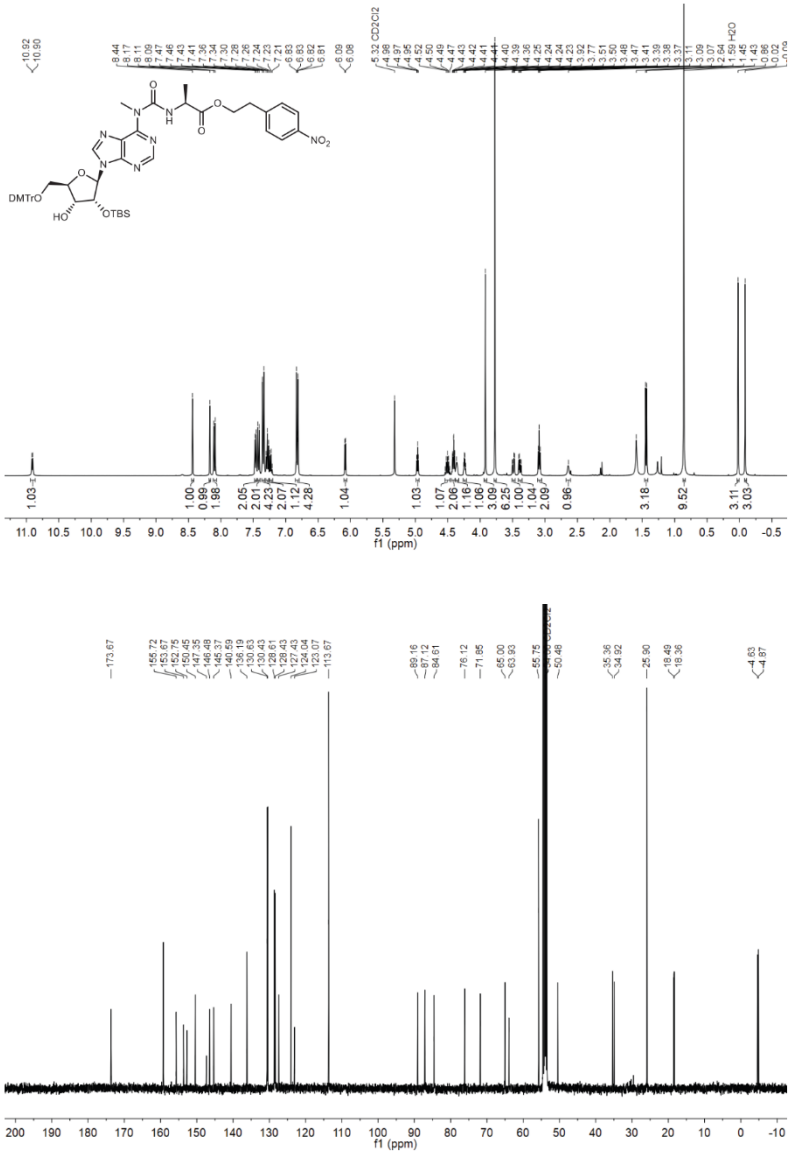
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12j



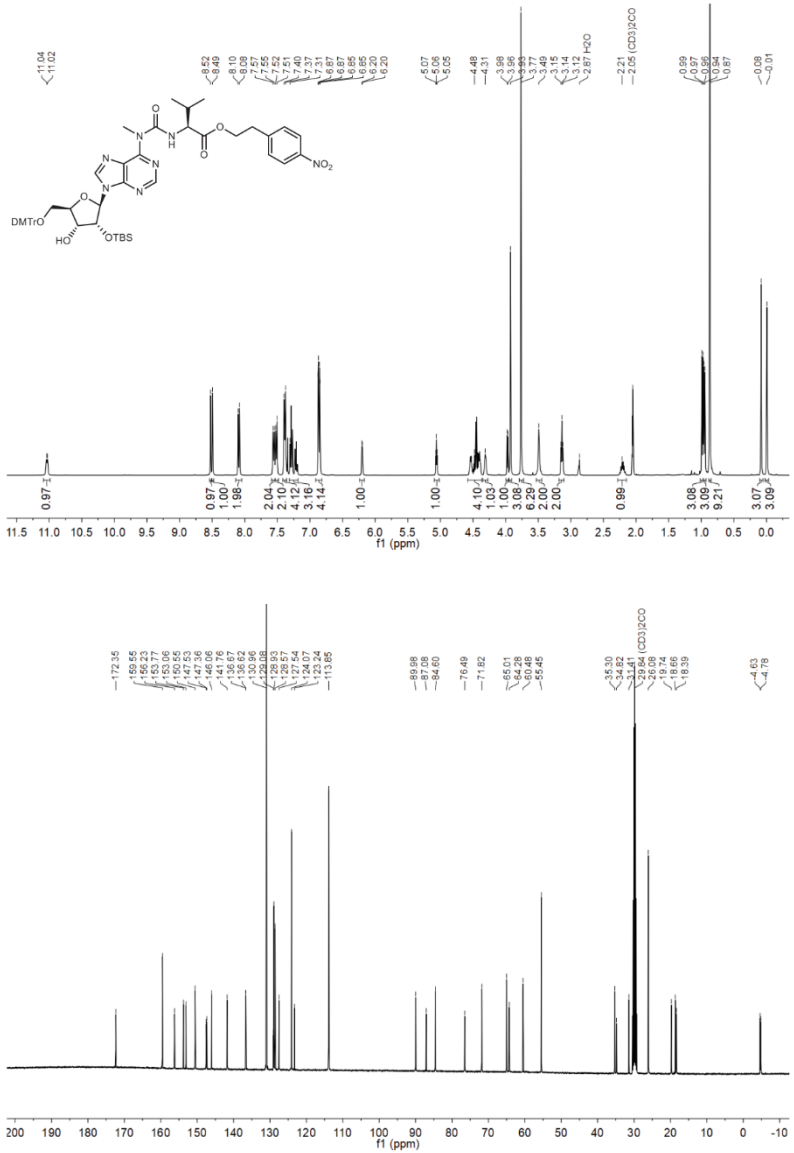
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13a



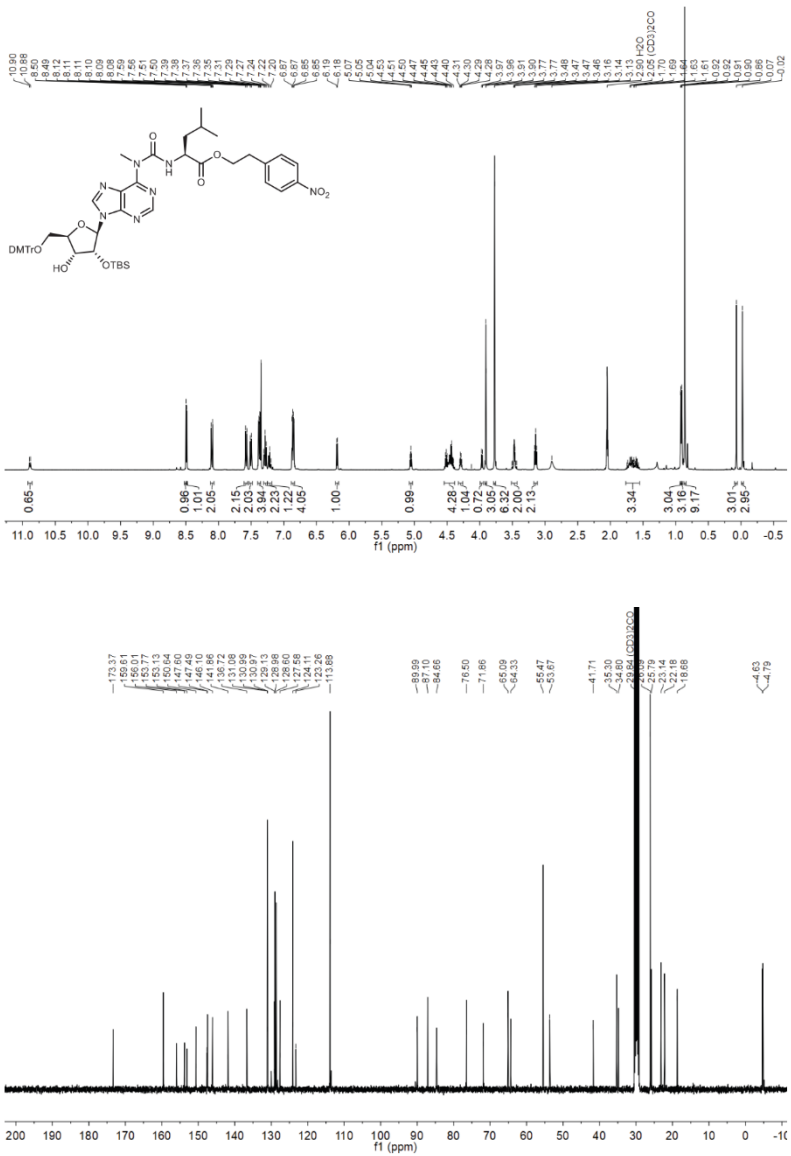
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13b



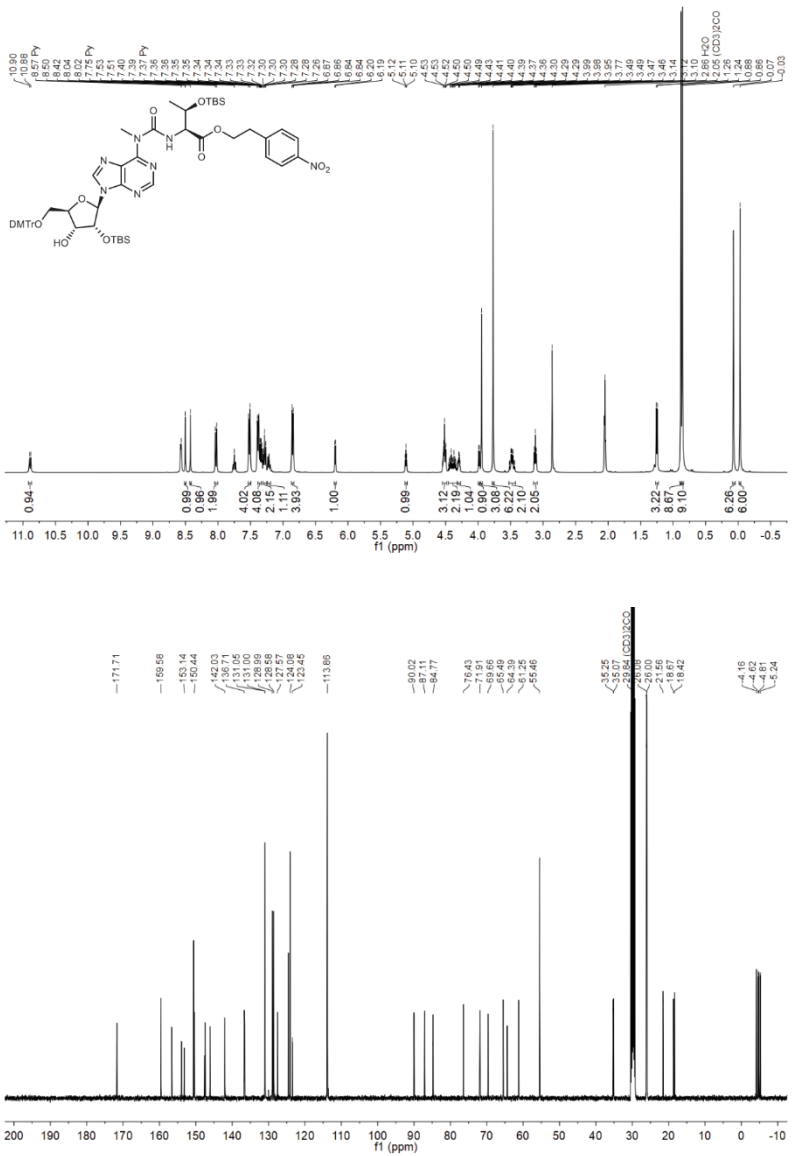
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13c



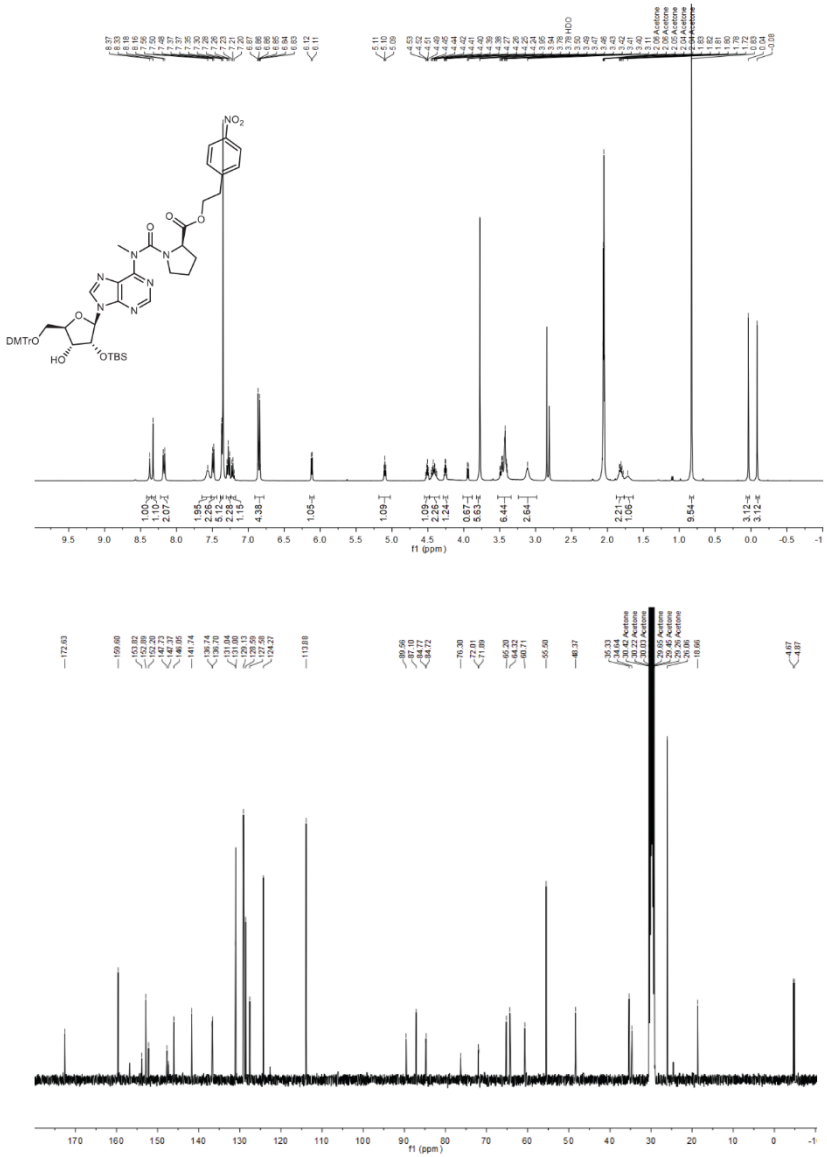
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13d



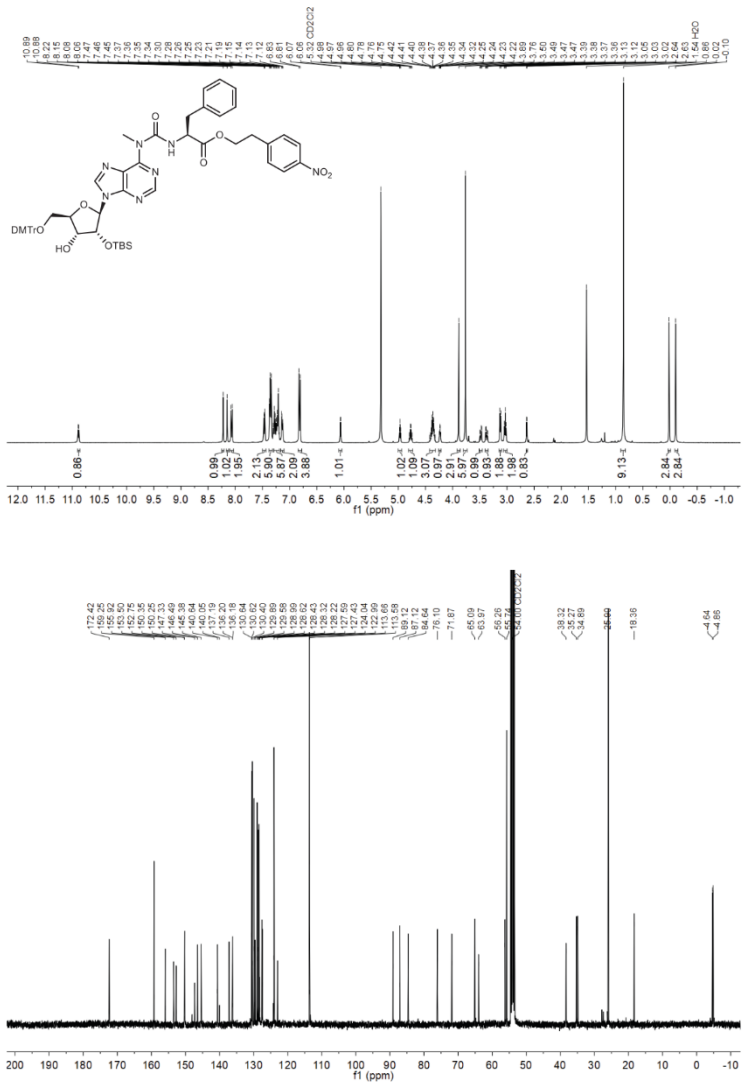
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13e



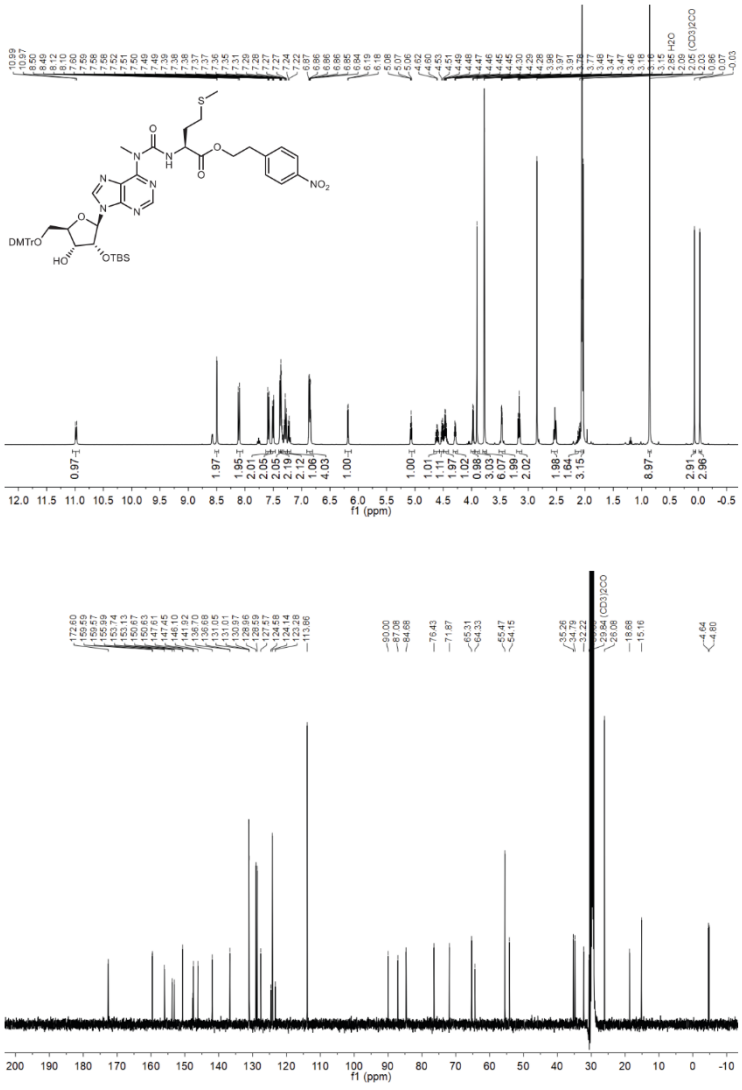
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13f



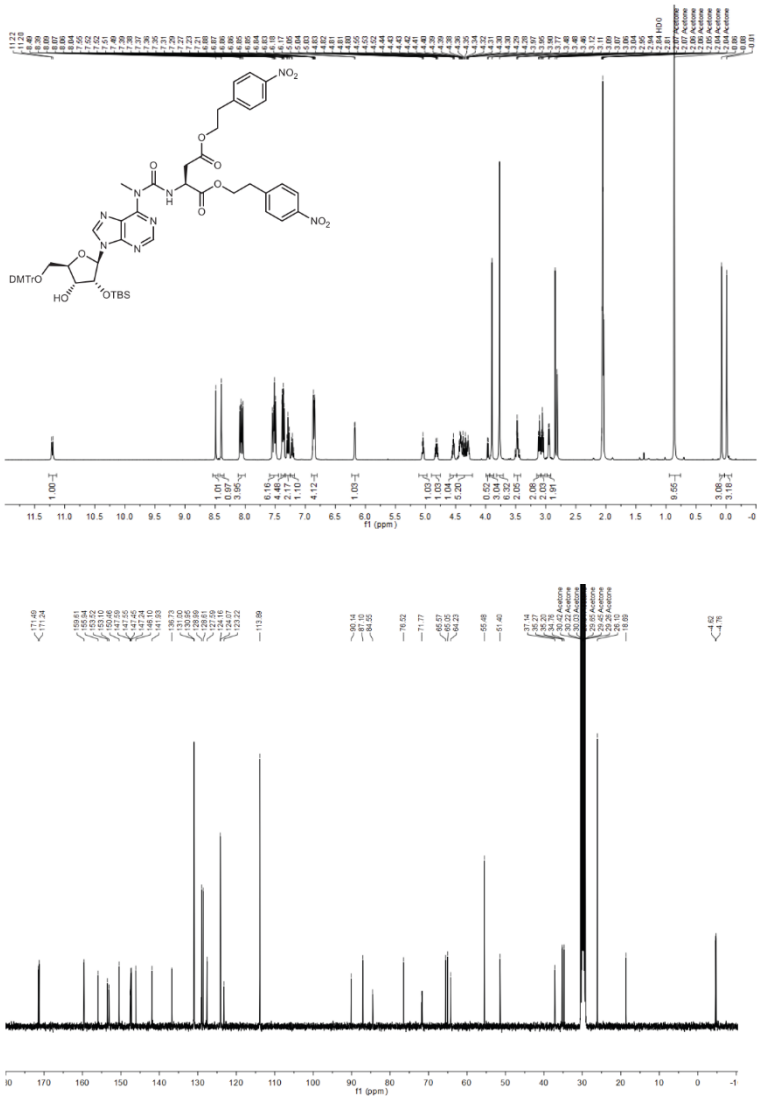
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13g



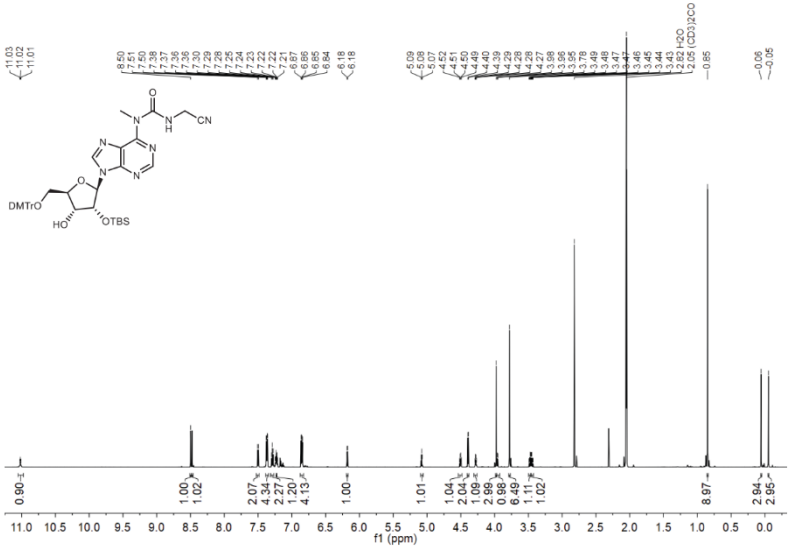
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13h

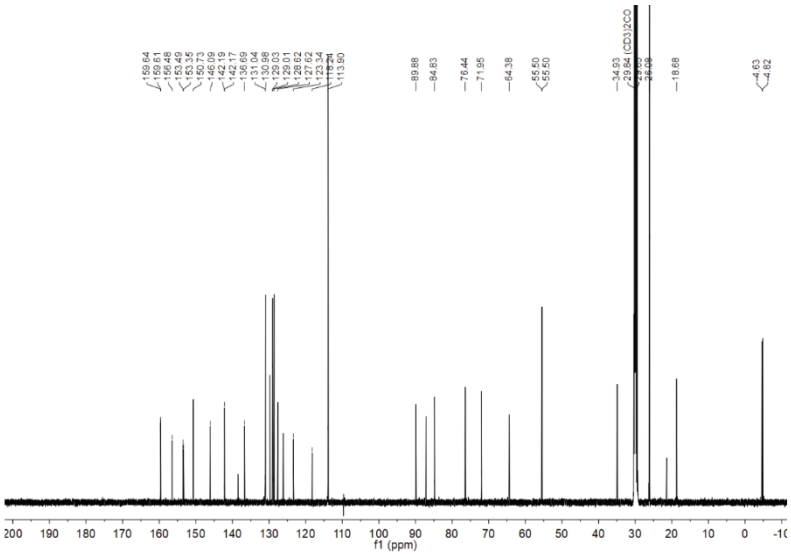


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13i

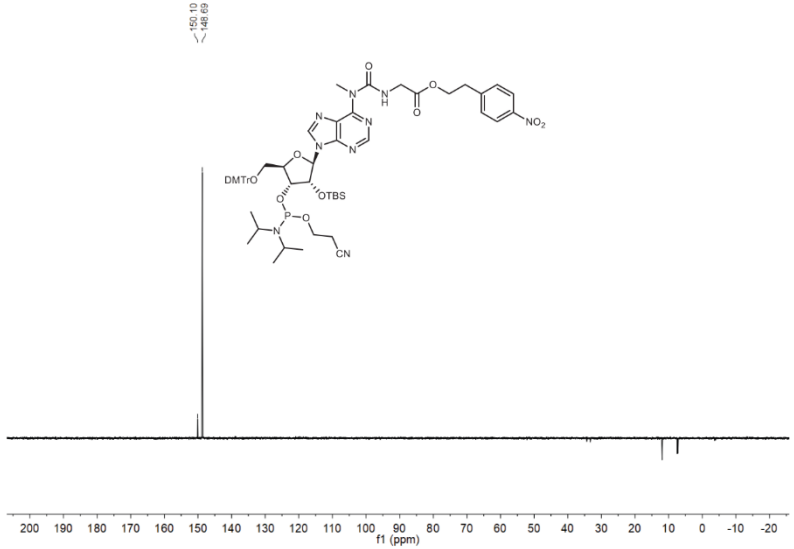


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13j

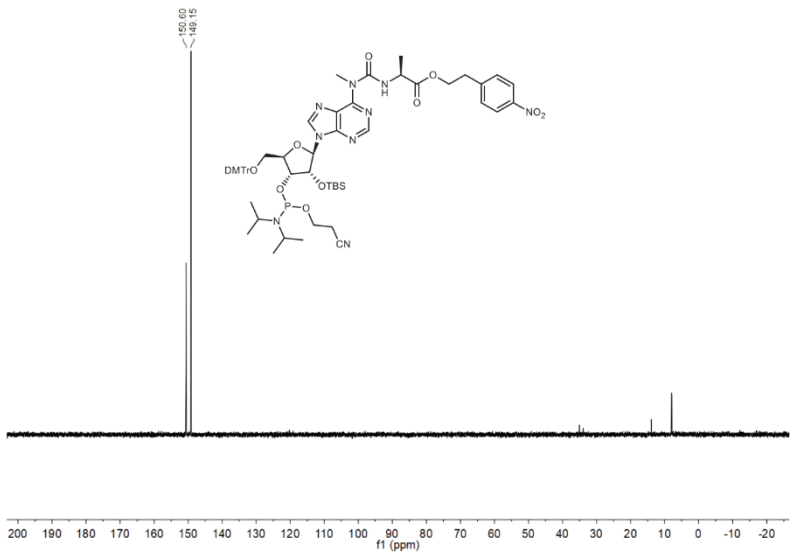




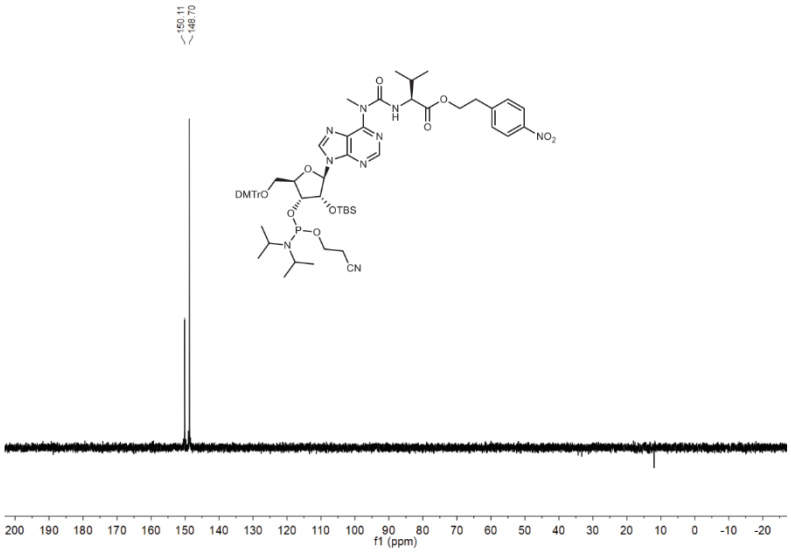
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14a



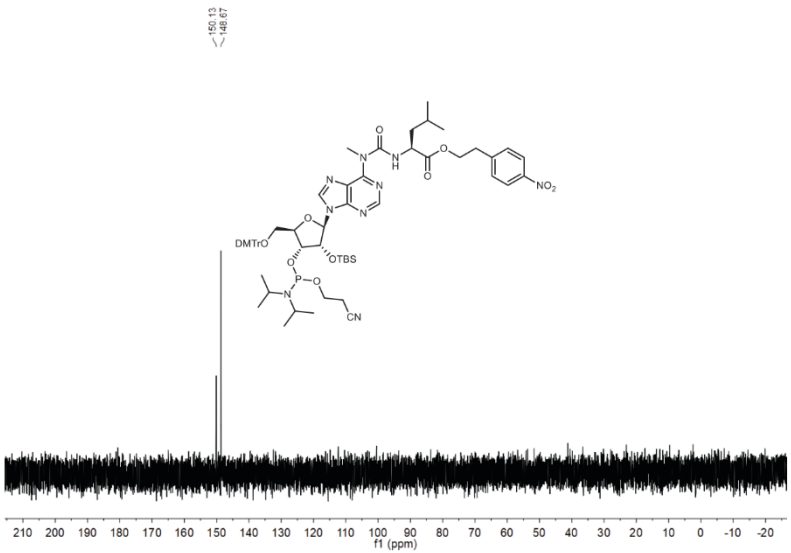
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14b



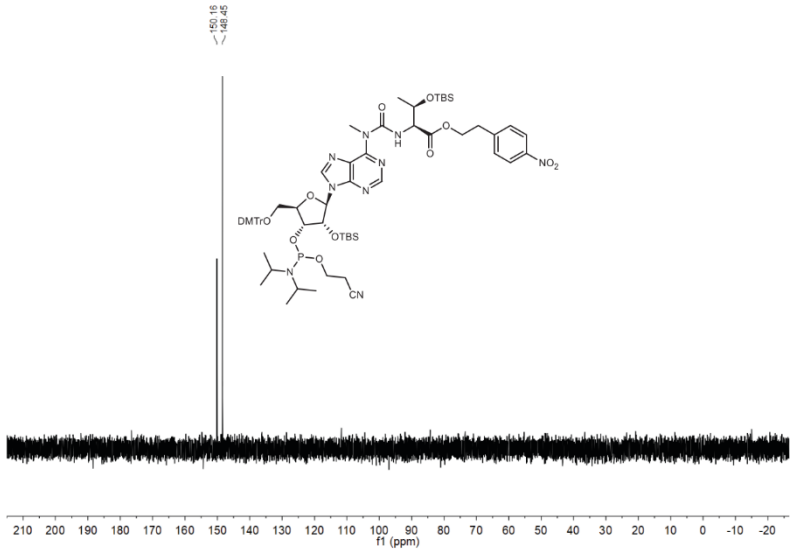
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14c



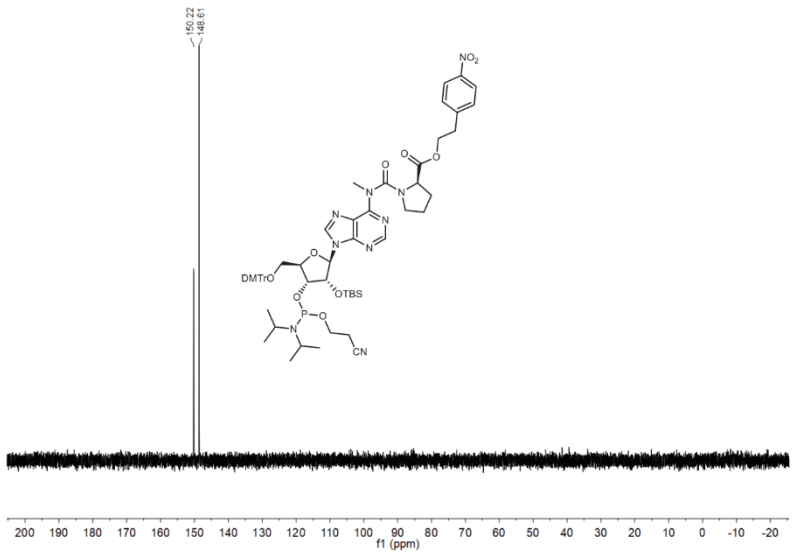
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14d



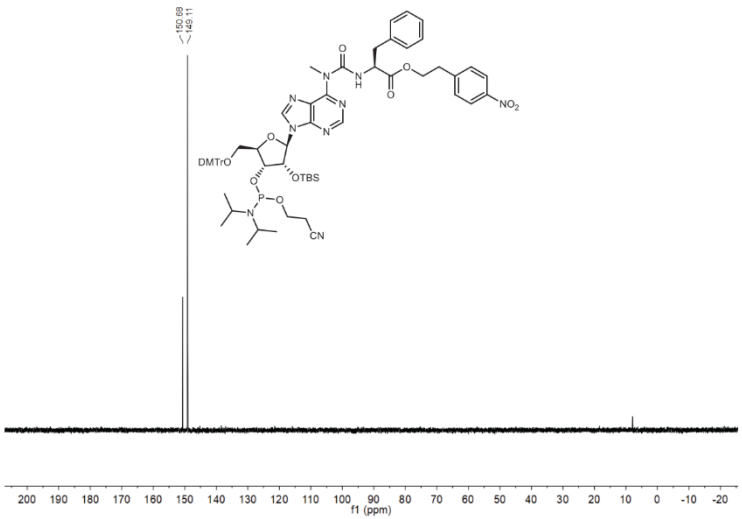
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14e



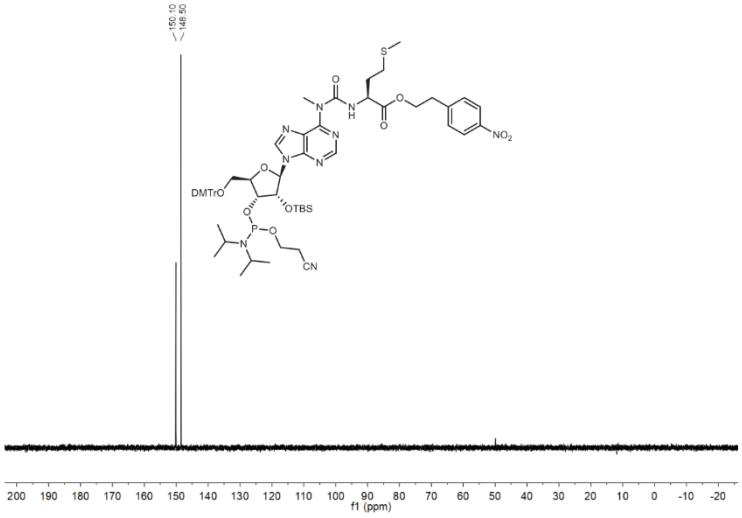
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14f



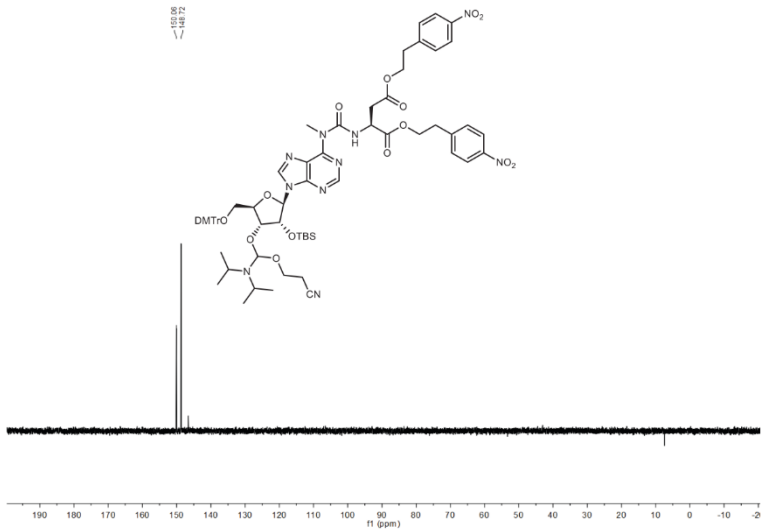
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14g



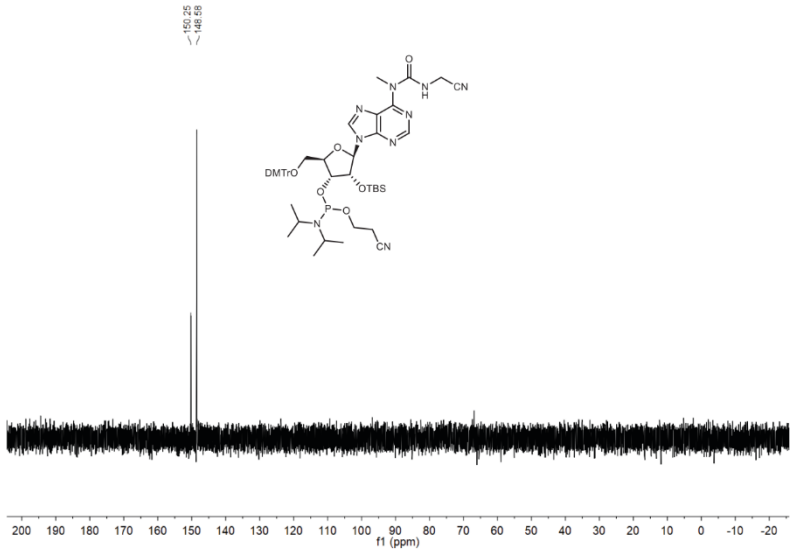
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14h



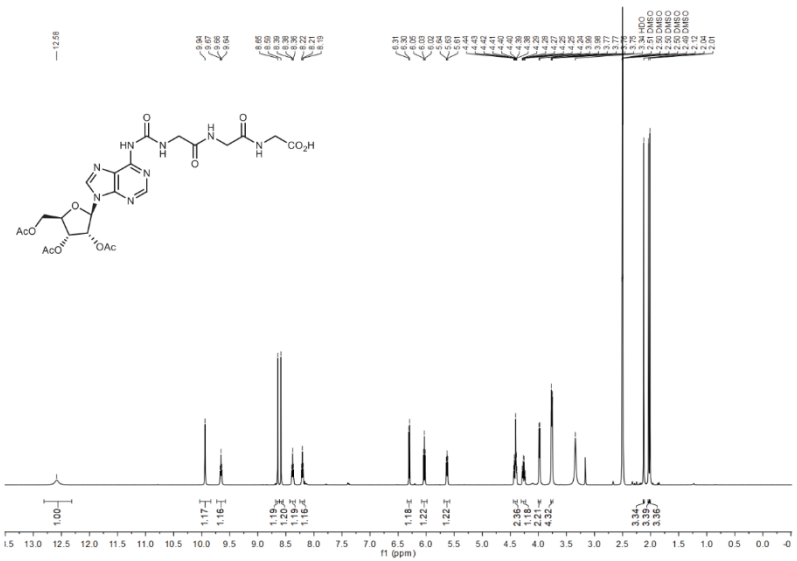
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14i

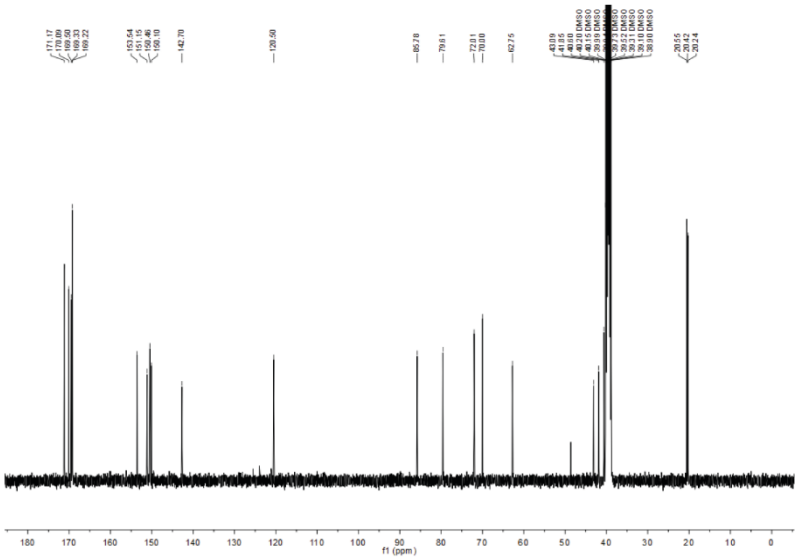


$^{31}P\{^{1}H\}$ NMR spectrum of compound 14j

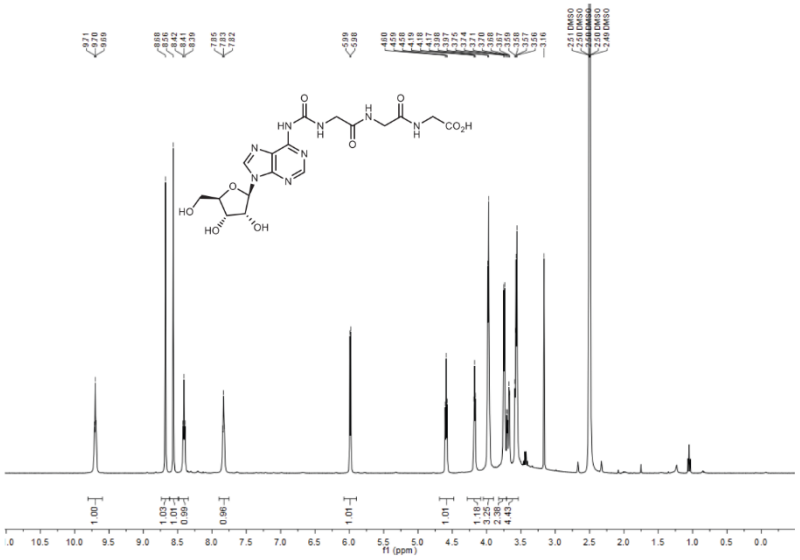


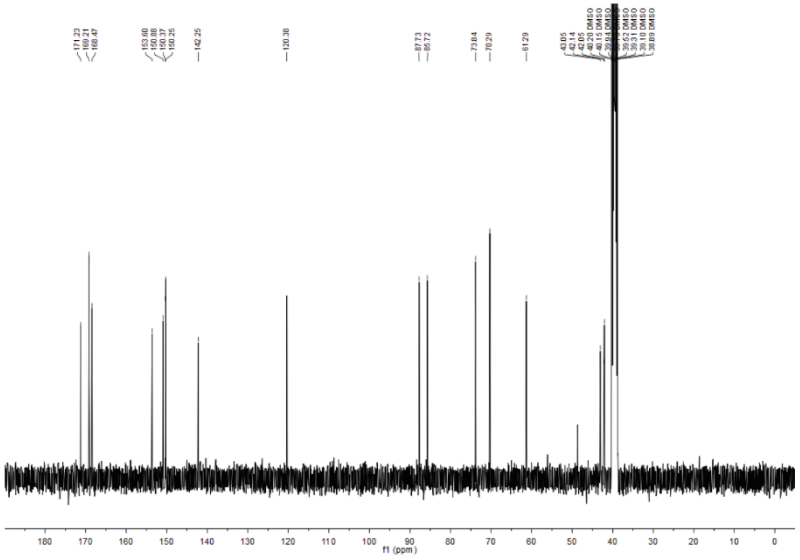
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 16



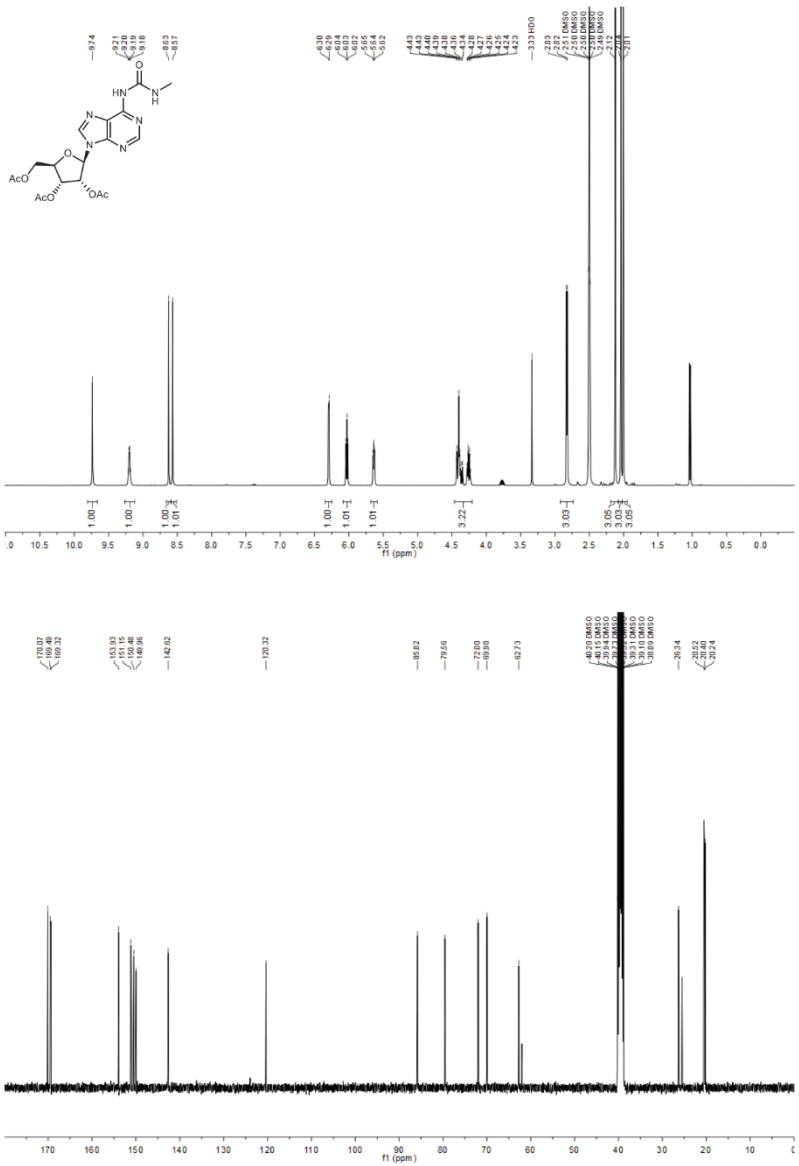


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 17

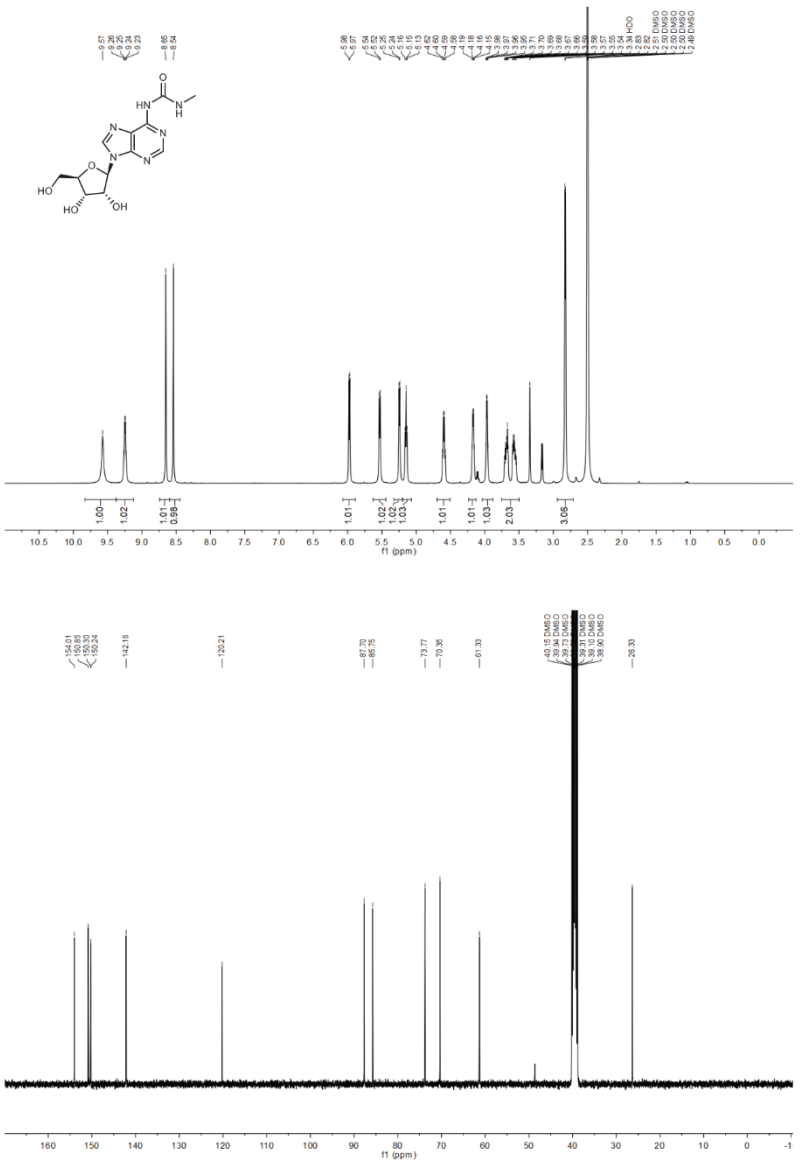




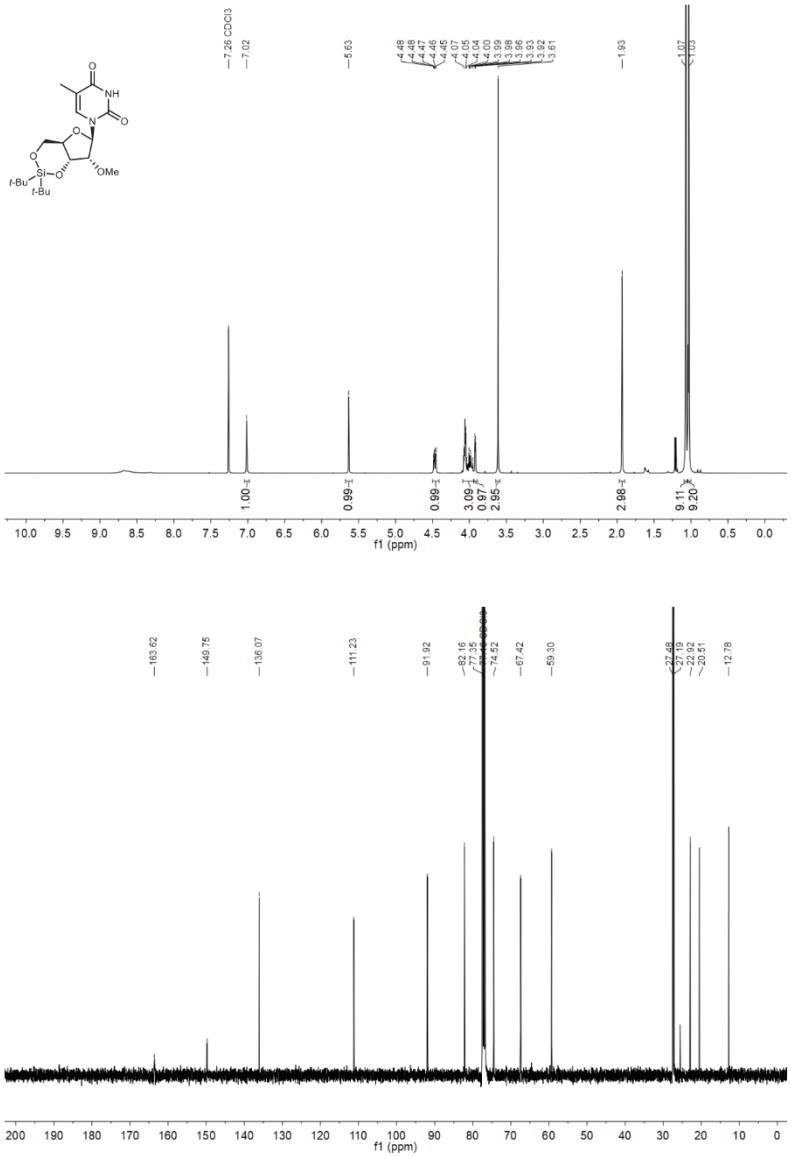
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 18



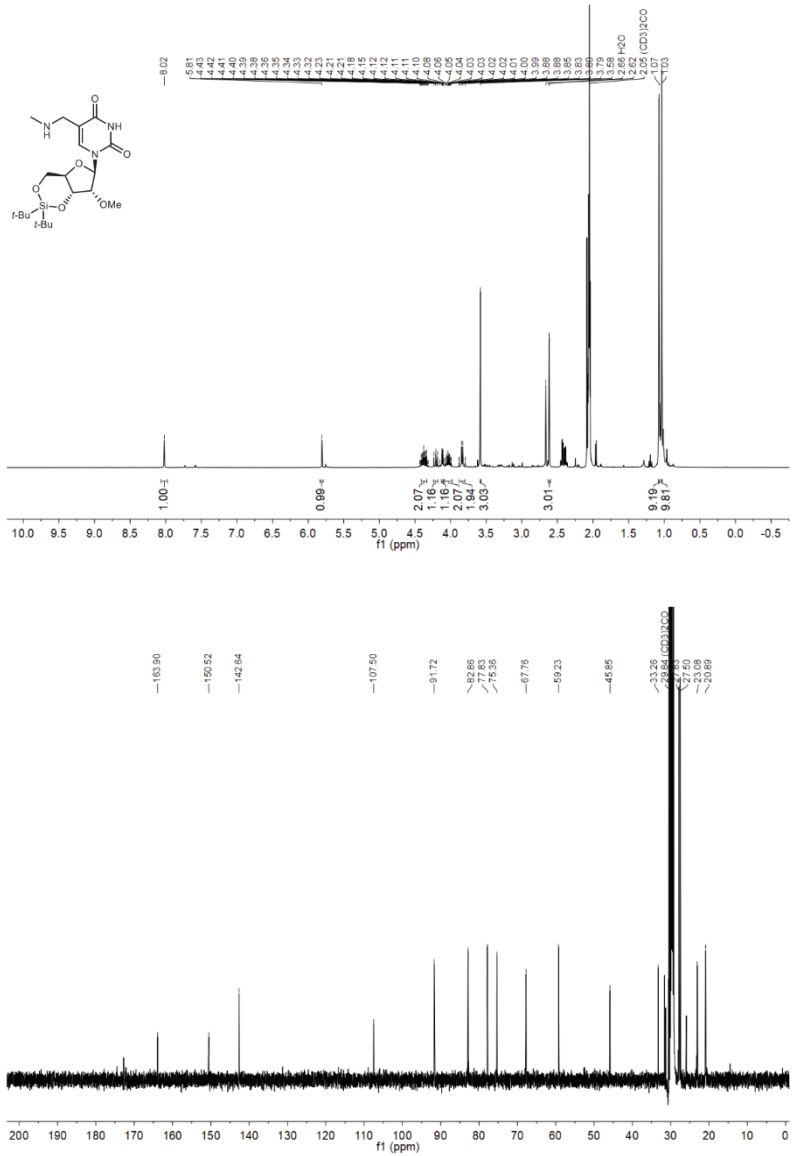
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 19



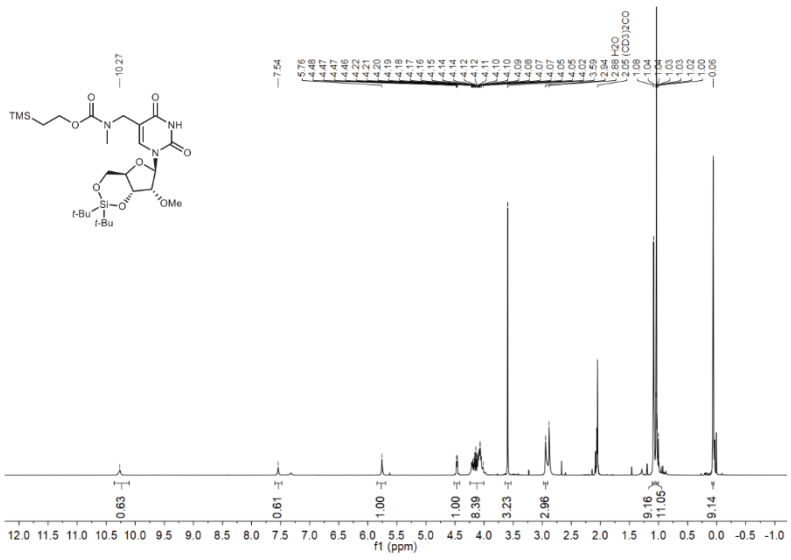
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 21

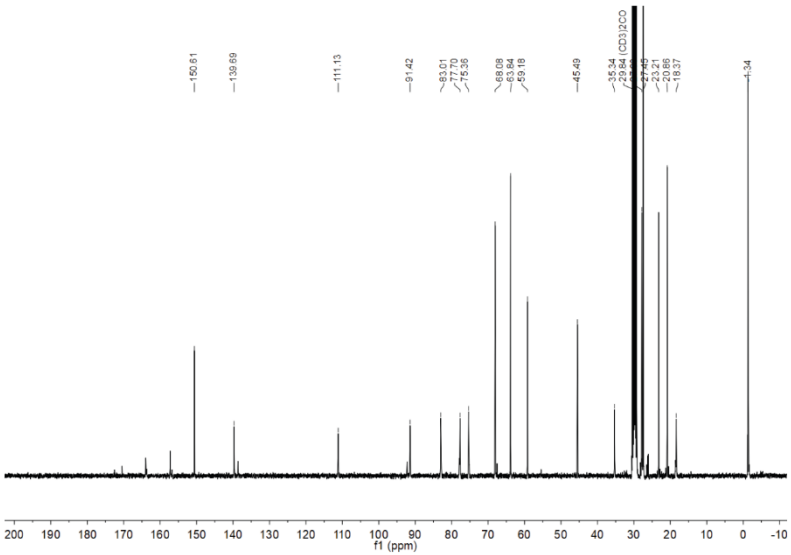


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 22

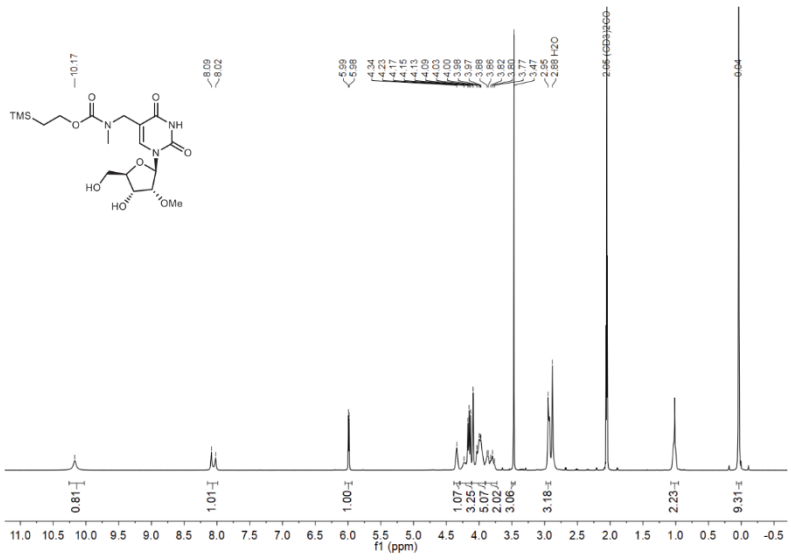


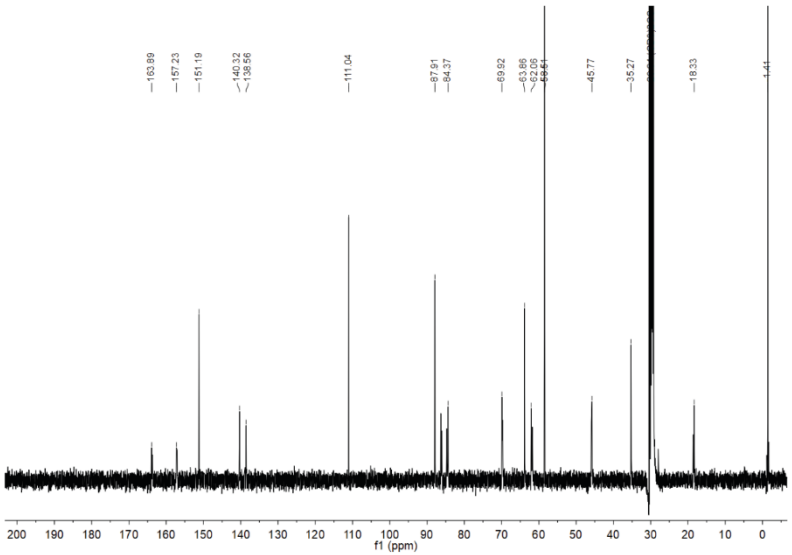
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 23



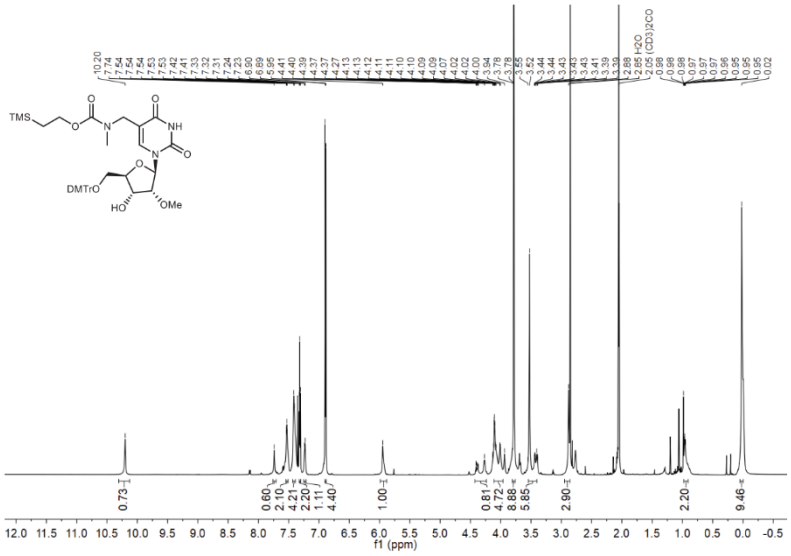


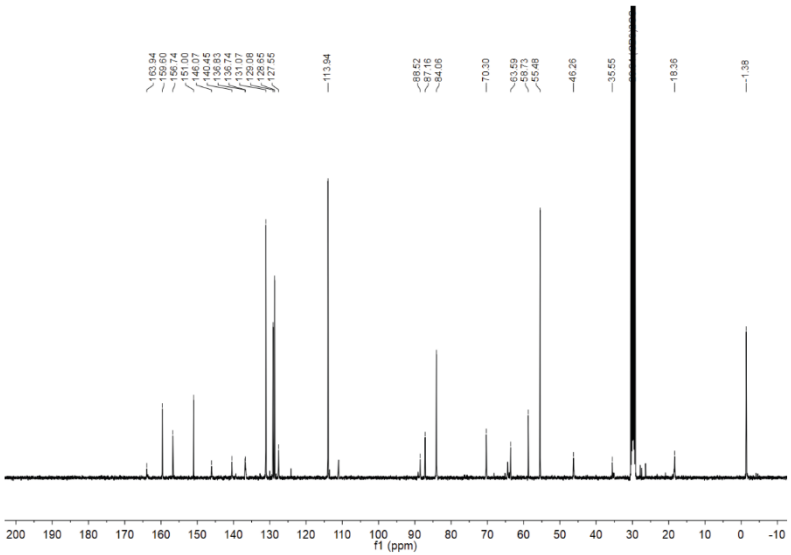
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 24



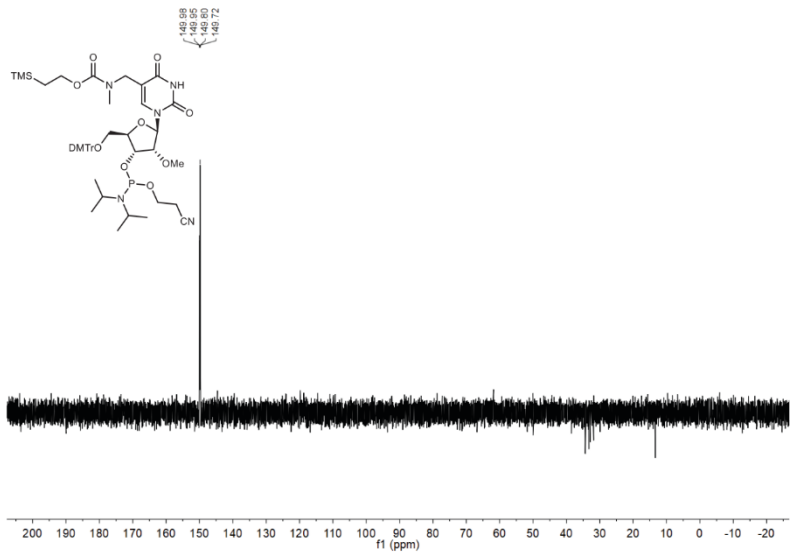


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 25

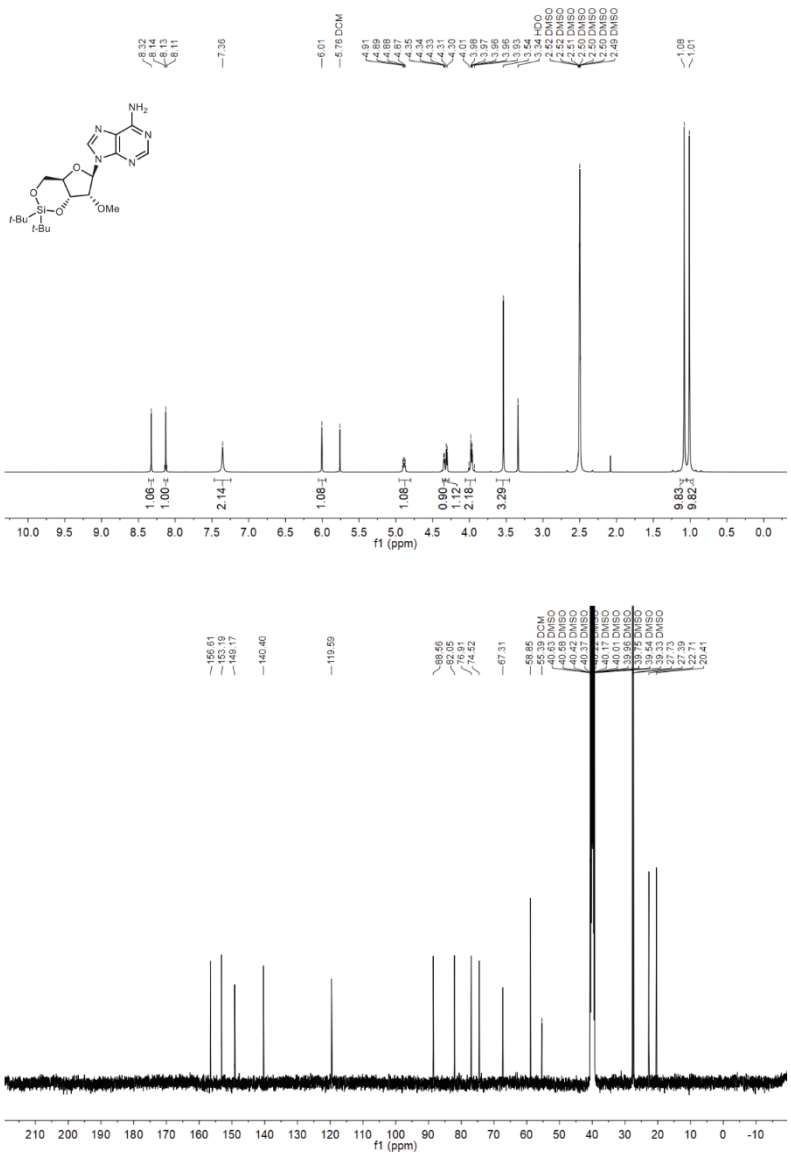




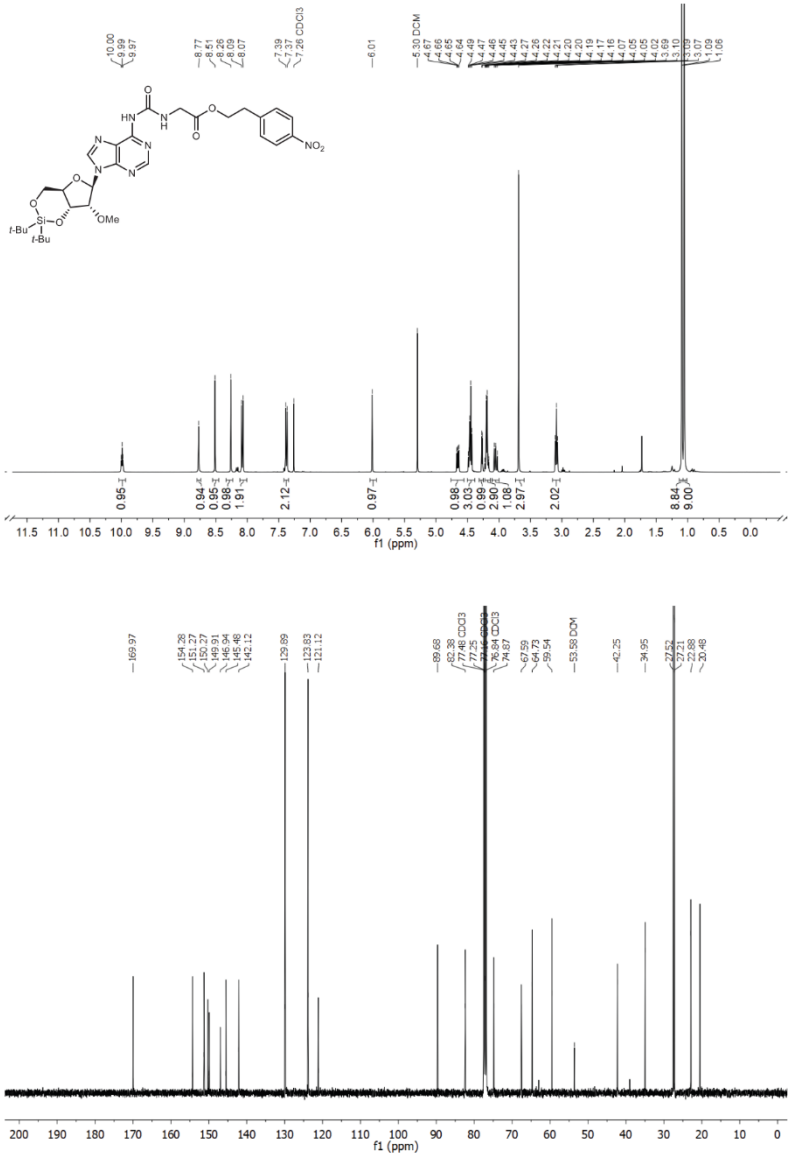
$^{31}P\{^{1}H\}$ NMR spectrum of compound 26



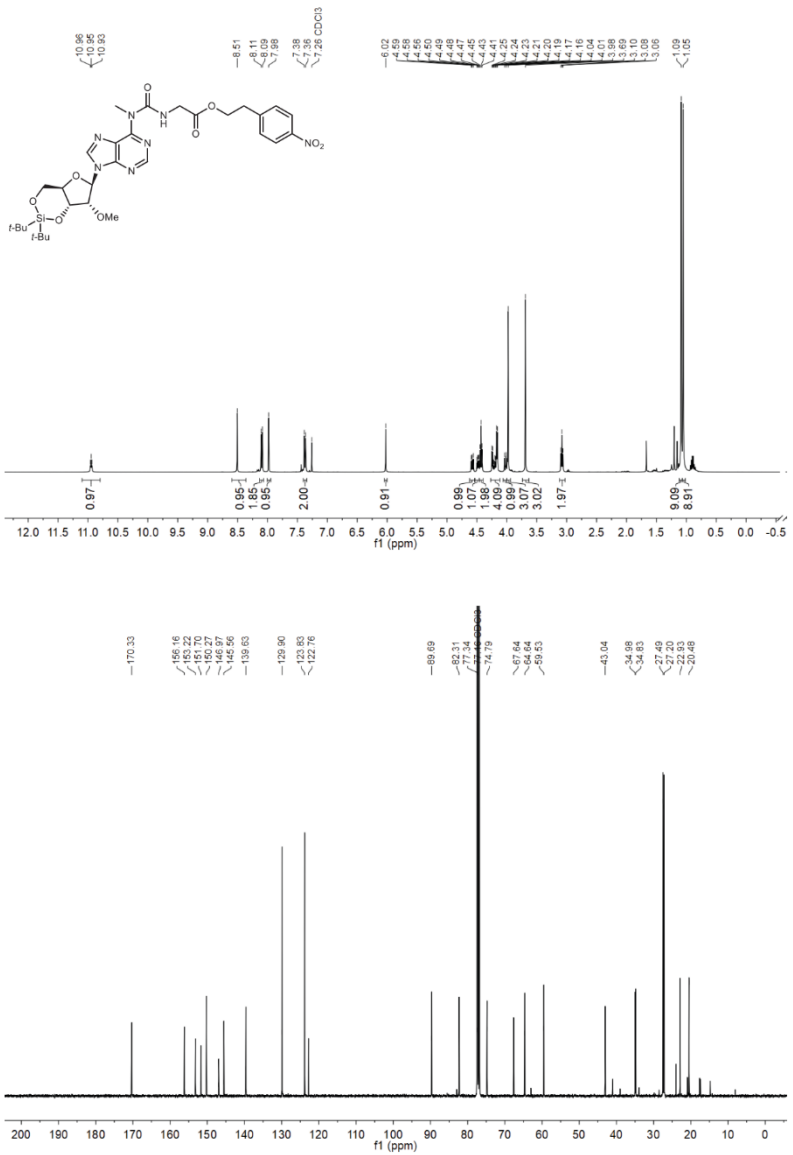
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 28



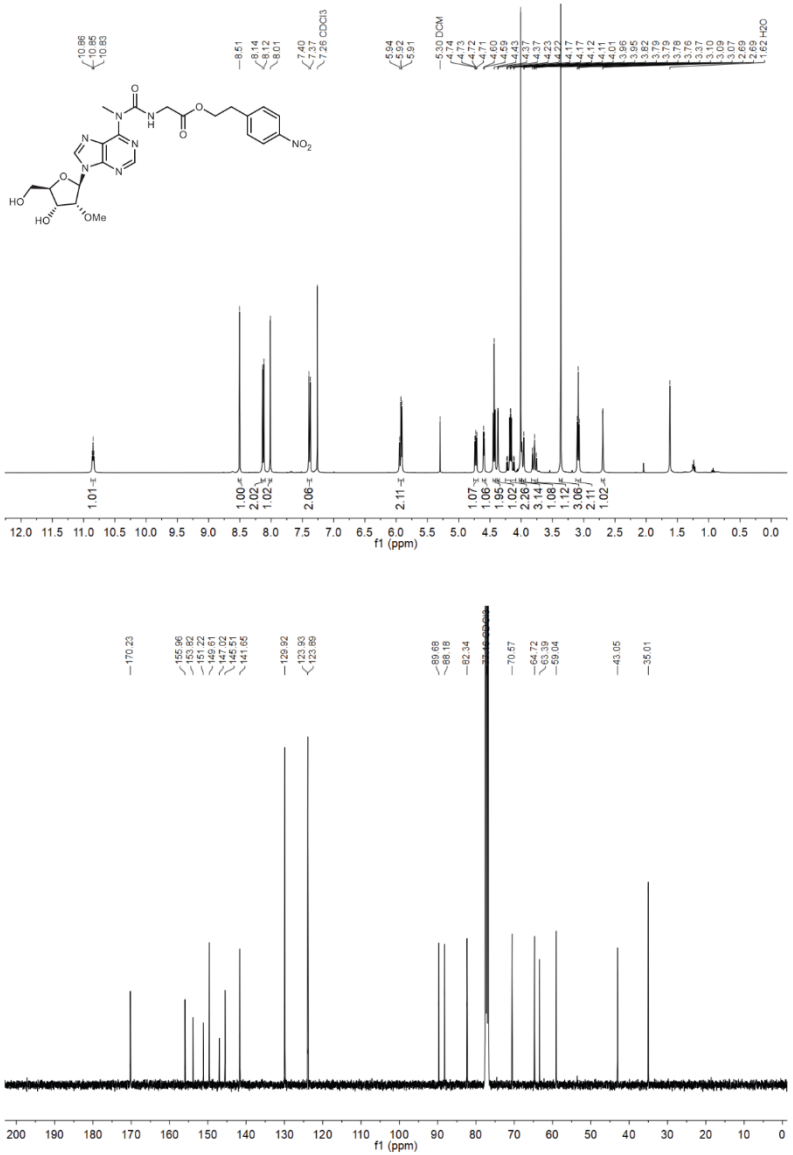
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 29



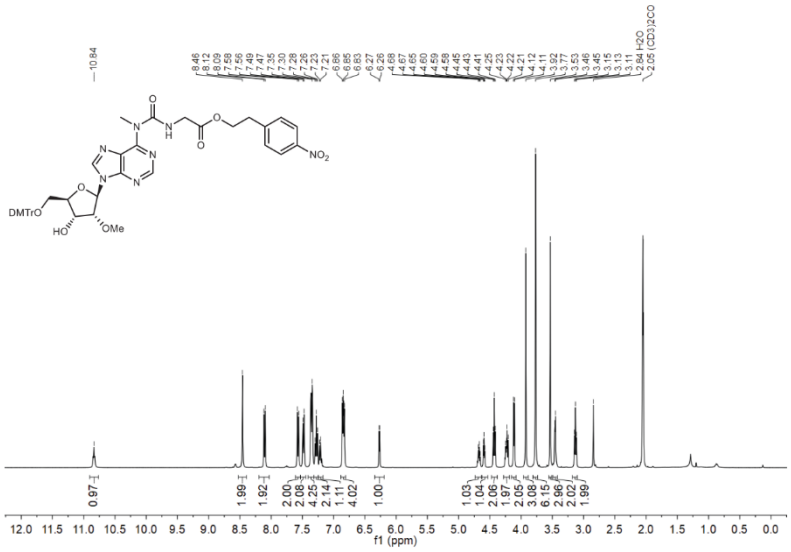
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 30

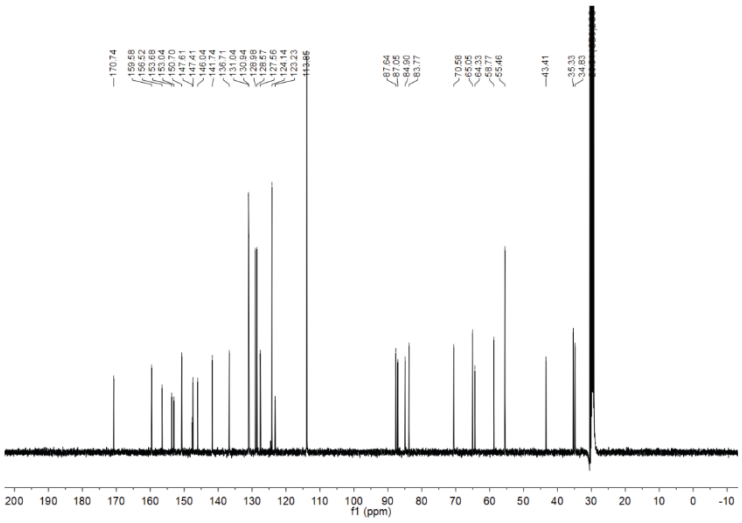


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 31

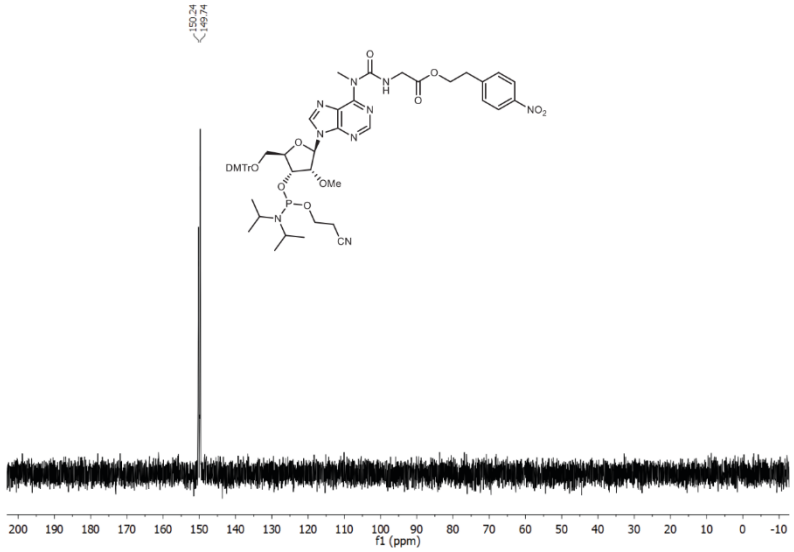


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 32





$^{31}P\{^{1}H\}$ NMR spectrum of compound 33



18. References

1. Fulmer, G. R. *et al.* NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **29**, 2176-2179 (2010).

2. Tanpure, A. A. & Balasubramanian, S. Synthesis and Multiple Incorporations of 2′-*O*-Methyl-5-hydroxymethylcytidine, 5-Hydroxymethylcytidine and 5-Formylcytidine Monomers into RNA Oligonucleotides. *ChemBioChem* **18**, 2236-2241 (2017).

3. Shute, R. E. & Rich, D. H. Synthesis and Evaluation of Novel Activated Mixed Carbonate Reagents for the Introduction of the 2-(Trimethylsilyl)ethoxycarbonyl(Teoc)-Protecting Group. *Synthesis* **1987**, 346-349 (1987).

4. Nainytė, M. *et al.* Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. *Chem. Eur. J.* **26**, 14856–14860 (2020).

5. Serebryany, V. & Beigelman, L. An efficient preparation of protected ribonucleosides for phosphoramidite RNA synthesis. *Tetrahedron Lett.* **43**, 1983-1985 (2002).

6. Sundaram, M., Crain, P. F. & Davis, D. R. Synthesis and Characterization of the Native Anticodon Domain of *E. coli* $tRNA^{Lys}$: Simultaneous Incorporation of Modified Nucleosides mnm^5s^2U, t^6A, and Pseudouridine Using Phosphoramidite Chemistry. *J. Org. Chem.* **65**, 5609-5614 (2000).

7. Matuszewski, M. & Sochacka, E. Stability studies on the newly discovered cyclic form of tRNA N^6-threonylcarbamoyladenosine (ct^6A). *Bioorg. Med. Chem. Lett.* **24**, 2703-2706 (2014).

8. Schneider, C. *et al.* Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth—Generation by Prebiotic Methylations and Carbamoylations. *Angew. Chem. Int. Ed.* **57**, 5943-5946 (2018).

9. Himmelsbach, F., Schulz, B. S., Trichtinger, T., Charubala, R. & Pfleiderer, W. The *p*-Nitrophenylethyl (NPE) Group: A Versatile New Blocking Group for Phosphate and Aglycone Protection in Nucleosides and Nucleotides. *Tetrahedron* **40**, 59-72 (1984).

10. Ferreira, F. & Morvan, F. Silyl Protecting Groups for Oligonucleotide Synthesis Removed by a $ZnBr_2$ Treatment. *Nucleosides, Nucleotides, and Nucleic Acids* **24**, 1009-1013 (2005).

11. Usanov, D. L., Chan, A. I., Maianti, J. P. & Liu, D. R. Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules. *Nat. Chem.* **10**, 704-714 (2018).

12. Hoops, S. *et al.* COPASI—a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067-3074 (2006).

13. Jash, B., Tremmel, P., Jovanovic, D. & Richert, C. Single nucleotide translation without ribosomes. *Nat. Chem.* **13**, 751-757 (2021).

14. Mochrie, S. G. J. The Boltzmann factor, DNA melting, and Brownian ratchets: Topics in an introductory physics sequence for biology and premedical students. *Am. J. Phys* **79**, 1121-1126 (2011).

15. Senior, M. M., Jones, R. A. & Breslauer, K. J. Influence of loop residues on the relative stabilities of DNA hairpin structures. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6242-6246 (1988).

16. Xodo, L. E., Manzini, G., Quadrifoglio, F., Marel, G. v. d. & van Boom, J. H. Hairpin structures in synthetic oligodeoxynucleotides: sequence effects on the duplex-to-hairpin transition. *Biochimie* **71**, 793-803 (1989).

Anhang II

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Isoxazole Nucleosides as Building Blocks for a Plausible Proto-RNA

*F. Xu, A. Crisp, T. Schinkel, R. C. A. Dubini, S. Hübner, S. Becker, F. Schelter, P. Rovó, T. Carell**

SUPPORTING INFORMATION

Table of Contents

SUPPORTING INFORMATION

General Experimental Methods

Chemicals were purchased from Sigma-Aldrich, TCI, Fluka, ABCR, Carbosynth or Acros organics and used without further purification. The solvents were of reagent grade or purified by distillation. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F_{254} TLC plates from Merck KGaA. Flash column chromatography was performed on Silicagel 60 (40-63 µm) silica gel from Macherey-Nagel. Reactions were conducted under a positive pressure of dry nitrogen in oven-dried glassware, and at ambient room temperature, unless otherwise specified. NMR spectra were recorded on Bruker AVIIIHD 400 (400 MHz) or Bruker Avance III (800 MHz) spectrometers. ^{1}H NMR shifts were calibrated to the residual solvent resonances: DMSO-d_6 (2.50 ppm), CD_3OD (4.87 ppm), Acetone-d_6 (2.05 ppm), $CDCl_3$ (7.26 ppm). ^{13}C NMR shifts were calibrated to the residual solvent: DMSO-d_6 (39.52 ppm), CD_3OD (49.00 ppm), $CDCl_3$ (77.16 ppm), Acetone-d_6 (29.84 ppm). All NMR spectra were analyzed using the program MestRE NOVA 10.0.1 from Mestrelab Research S. L. Normal resolved mass spectra were measured on a LTQ FT-ICR by Thermo Finnigan GmbH. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of the Ludwig-Maximilians-Universität München on the following spectrometers (ionization mode in brackets): MAT 95 (EI) and MAT 90 (ESI) from Thermo Finnigan GmbH, unless otherwise specified. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system.

Synthesis and characterization of the Phosphoramidite Building Blocks

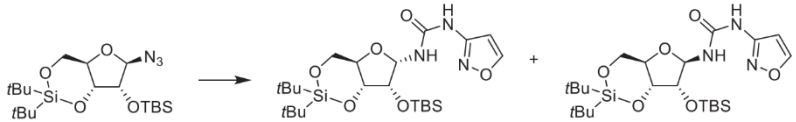
2-*O*-*tert*-butyldimethylsilyl-3,5-*O*-(di-*tert*-butylsilandiyl)-1-β-D-ribofuranosy-azide (11)

10 **11**

10[1] (10.9 g, 34.5 mmol, 1.00 eq.) was dissolved in DMF (36 mL) imidazole (11.8 g, 173 mmol, 5.00 eq.) and TBS-Cl (6.25 g, 41.5 mmol, 1.20 eq.) were added and the mixture was stirred at rt for 16 h. The solvent was evaporated, the residue was taken up in EtOAc (50 mL) and sat. $NaHCO_3$-solution (50 mL) and extracted with EtOAc (2×50 mL). The combined organic layers were dried over Na_2SO_4 and concentrated in vacuo. The crude product was purified by flash column chromatography (silica, iHex:EtOAc = 100:1) to yield **11** as colorless crystals (11.6 g, 26.9 mmol, 78%).

R_f = 0.43 (iHex:EtOAc = 100:1);**^{1}H NMR** (400 MHz, $CDCl_3$): δ (ppm) = 5.07 (s, 1H, 1-H), 4.42 (dd, *J* = 9.1, 5.1 Hz, 1H, 5-H), 4.12 (ddd, *J* = 10.9, 9.6, 5.1 Hz, 1H, 4-H), 4.01 (d, *J* = 4.1 Hz, 1H, 2-H), 3.95 – 3.88 (m, 2H, 3-H, 5-H), 1.05 (s, 9H, Si-C-CH_3), 1.00 (s, 9H, Si-C-CH_3), 0.91 (s, 9H, Si-C-CH_3), 0.13 (s, 3H, Si-CH_3), 0.12 (s, 3H, Si-CH_3);**^{13}C NMR** (101 MHz, $CDCl_3$): δ (ppm) = 96.19 (C1), 76.33 (C4), 75.98 (C2), 74.65 (C3), 68.42 (C5), 27.58 (Si-C-CH_3), 27.15 (Si-C-CH_3), 25.98 (Si-C-CH_3), 22.83 (Si-C), 20.47 (Si-C), 18.50 (Si-C), -4.25 (Si-CH_3), -5.06 (Si-CH_3); **IR** (cm^{-1}): ϑ = 4029 (w), 3843 (w), 3775 (w), 2933 (s), 2884 (m), 2859 (s), 2386 (w), 2205 (w), 2109 (s), 1472 (m), 1387 (m), 1362 (m), 1297 (w), 1253 (s), 1167 (m), 1142 (s), 1049 (s), 998 (s), 925 (m), 880 (m), 826 (s), 776 (s), 753 (w); **HRMS** (EI): calcd. for $C_{18}H_{36}N_3O_4Si_2^+$ [M]$^+$: 414.2239; found: 414.2252.

N-isoxazol-3-yl-*N*'-(1'-(2'-*O*-*tert*-butyldimethylsilyl-3',5'-*O*-(di-*tert*-butylsilandiyl)-β-D-ribofuranosyl))-urea (12)



11 **α-12** **β-12**

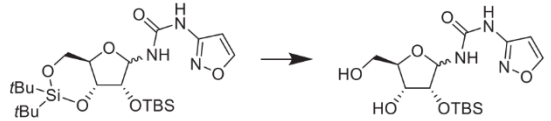
11 (1.01 g, 2.36 mmol, 1.00 eq.) and 3-aminoisoxzole (209 µL, 2.83 mmol, 1.20 eq.) were dissolved in dry toluene (23 mL). Palladium on carbon (10%, 126 mg, 118 µmol, 0.05 eq.) and XPhos (113 mg, 236 µmol, 0.1 eq.) were added to the solution. The reaction mixture was stirred at 60°C under CO-atmosphere (5 bar) for 18 h. The catalyst was filtered off with a pad of $MgSO_4$ and washed with toluene. The filtrate was concentrated under reduced pressure and the crude product was purified by flash column chromatography (silica, *i*Hex:EtOAc = 4:1) to yield **α-12** (255 mg, 496 µmol, 21%) and **β-12** (775 mg, 1.51 mmol, 64%) as colorless solids.

SUPPORTING INFORMATION

α-12:
R_f = 0.19 (*i*Hex:EtOAc = 4:1); **^{1}H NMR** (600 MHz, $CDCl_3$): δ (ppm) = 8.42 (s, 1H, NH), 8.20 (d, *J* = 1.8 Hz, 1H, 5-H), 6.43 (s, 1H, 4-H), 5.90 (dd, *J* = 9.2, 4.1 Hz, 1H, 1'-H), 4.39 – 4.34 (m, 2H, 2'-H, 5'-H), 4.04 (td, , *J* = 10.0, 5.2 Hz, 1H, 4'-H), 3.94 (dd, *J* = 9.7, 4.5 Hz, 1H, 3'-H), 3.86 – 3.82 (m, 1H, 5'-H), 1.06 (s, 9H, Si-C-CH_3), 1.02 (s, 9H, Si-C-CH_3), 0.96 (s, 9H, Si-C-CH_3), 0.18 (s, 3H, Si-CH_3), 0.17 (s, 3H, Si-CH_3); **^{13}C NMR** (151 MHz, $CDCl_3$): δ (ppm) = 158.3 (C5), 157.7 (C3), 153.8 (C=O), 98.2 (C4), 82.2 (C1'), 77.7 (C3'), 72.5 (C4'), 71.0 (C4'), 68.3 (C5'), 27.5 (C), 27.1 (Si-C-CH_3), 26.2 (Si-C-CH_3), 22.7 (Si-C-CH_3), 22.7 (Si-C), 20.3 (Si-C), 18.6 (Si-C), -4.3 (Si-CH_3), -4.9 (Si-CH_3); **IR** (cm^{-1}): ϑ = 3342 (w), 2931 (m), 2859 (m), 1707 (s), 1590 (s), 1504 (s), 1468 (m), 1389 (w), 1260 (w), 1209 (w), 1174 (w), 1114 (m), 1053 (s), 964 (m), 905 (w). 827 (s), 773 (m), 705 (w); **HRMS** (ESI): calcd. for $C_{23}H_{44}N_3O_6Si_2^+$ [M+H]$^+$: 514.27632, found: 514.27643.

β-12:
R_f = 0.21 (*i*Hex:EtOAc = 4:1); **^{1}H NMR** (600 MHz, $CDCl_3$): δ (ppm) = 8.85 (s, 1H, NH), 8.20 (d, ^3J = 1.8 Hz, 1H, 5-H), 6.27 (s, 1H, 4-H), 5.47 (d, ^3J = 8.5 Hz, 1H, 1'-H), 4.40 (dd, ^2J = 9.2, ^3J = 5.0 Hz, 1H, 5'-H), 4.21 (d, ^3J = 4.7 Hz, 1H, 2'-H), 4.04 (ddd, ^3J = 10.4 Hz, ^3J = 9.5 Hz, ^3J = 5.1 Hz, 1H, 4'-H), 3.93 - 3.87 (m, 2H, 3'-H, 5'-H), 1.07 (s, 9H, Si-C-CH_3), 1.02 (s, 9H, Si-C-CH_3), 0.93 (s, 9H, Si-C-CH_3), 0.15 (s, 3H, Si-CH_3), 0.13 (s, 3H, Si-CH_3); **^{13}C NMR** (151 MHz, $CDCl_3$): δ (ppm) = 158.4 (C5), 158.1 (C=O), 153.7 (C3), 98.2 (C4), 89.4 (C1'), 77.1 (C3'), 76.2 (C2'), 73.7 (C4'), 68.6 (C5'), 27.6 (Si-C-CH_3), 27.2 (Si-C-CH_3), 26.1 (Si-C-CH_3), 22.9 (Si-C), 20.50 (Si-C), 18.52 (Si-C), -4.2 (Si-CH_3), -4.9 (Si-CH_3); **IR** (cm^{-1}): ϑ = 3340 (w), 2931 (m), 2858 (m), 1706 (s), 1593 (m), 1503 (m), 1468 (m), 1388 (w), 1260 (w), 1113 (m), 1054 (s), 963 (s), 827 (s), 774 (s), 705 (m); **HRMS** (ESI): calcd. for $C_{23}H_{44}N_3O_6Si_2^+$ [M+H]$^+$: 514.27632, found: 514.27662.

N-isoxazol-3-yl-*N*'-(1'-(2'-*O*-*tert*-butyldimethylsilyl)-β-D-ribofuranosyl)-urea (13)



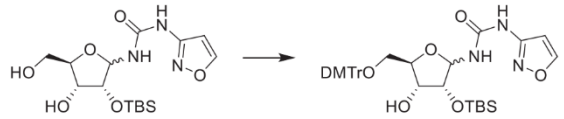
α-12 **α-13**
β-12 **β-13**

α-12 or **β-12** (1.00 eq.) was dissolved in DCM and treated with Pyridine (0.01 M) and HF (70% in Pyridine, 5.00 eq.) at 0°C. The resulting mixture was stirred for 4 h and afterwards quenched by the addition of methoxytrimethylsilane. The solution was diluted with DCM and washed with sat. $NaHCO_3$. The aqueous phase was extracted with DCM (3×). The combined organic layers were washed with Brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by flash column chromatography (silica, DCM:MeOH = 100:5) to yield **13** as a colorless foam.

β-13
yield: 91%; R_f = 0.43 (DCM:MeOH = 100:5); **^{1}H NMR** (600 MHz, $CDCl_3$): δ (ppm) = 9.29 (s, 1H, NH), 8.21 (d, *J* = 1.8 Hz, 1H, 5-H), 6.50 (s, 1H, 4-H), 5.24 (t, *J* = 6.6 Hz, 1H, 1'-H), 4.54 (t, *J* = 5.3 Hz, 1H, 2'-H), 4.20 (dt, *J* = 5.4, 3.4 Hz, 1H, 3'-H), 4.08 (t, ^3J = 2.6 Hz, 1H, 4'-H), 3.86 (dd, *J* = 9.6, 2.3 Hz, 1H, 5'-H), 3.70 (m , 2H, 5'-H, 5'-OH), 2.81 (d, *J* = 3.7 Hz, 1H, 3'-OH), 0.91 (s, 9H, Si-C-CH_3), 0.13 (s, 3H, Si-CH_3), 0.12 (s, 3H, Si-CH_3); **^{13}C NMR** (151 MHz, $CDCl_3$): δ (ppm) = 158.6 (C5), 158.2 (C=O), 154.7 (C3), 98.3 (C4), 87.4 (C1'), 84.8 (C4'), 74.6 (C2'), 71.9 (C3'), 62.9 (C5'), 25.9 (Si-C-CH_3), 18.18 (Si-C), -4.49 (Si-CH_3), -4.91 (Si-CH_3); **IR** (cm^{-1}): ϑ = 3304 (m), 2928 (m), 2856 (m), 1676 (s), 1594 (s), 1531 (s), 1471 (m), 1373 (w), 1254 (w), 1126 (w), 996 (m), 903 (w), 837 (s), 780 (s), 667 (w); **HRMS** (ESI): calcd. for $C_{15}H_{27}N_3O_6Si^+$ [M+H]$^+$: 374.17419, found: 374.17397.

α-13:
yield: 90%; R_f = 0.42 (DCM:MeOH = 100:5); **^{1}H NMR** (600 MHz, $CDCl_3$): δ (ppm) = 8.46 (s, 1H, NH), 8.20 (d, *J* = 1.8 Hz, 1H, 5-H), 6.42 (brs, 1H, 4-H), 5.80 (dd, *J* = 8.4 Hz, 5.6 Hz, 1H, 1'-H), 4.33 (t, *J* = 5.6 Hz, 1H, 2'-H), 4.13 – 4.11 (m, 1H, 3'-H), 4.09 (q, *J* = 3.3 Hz, 1H, 4'-H), 3.82 (d, *J* = 12.3 Hz, 1H, 5'-H), 3.70 – 3.64 (m, 1H, 5'-H), 2.77 (d, *J* = 9.2 Hz, 1H, 3'-OH), 2.40 (d, *J* = 7.9 Hz, 1H, 5'-OH), 0.93 (s, 9H, Si-C-CH_3), 0.16 (s, 3H, Si-CH_3), 0.12 (s, 3H, Si-CH_3); **^{13}C NMR** (151 MHz, $CDCl_3$): δ (ppm) = 158.5 (C5), 158.0 (C3), 154.4 (C=O), 98.2 (C4), 83.5 (C4'), 81.9 (C1'), 72.0 (C3'), 71.6 (C2'), 62.6 (C5'), 25.9 (Si-C-CH_3), 18.3 (Si-C), -4.60 (Si-CH_3), -4.91 (Si-CH_3); **IR** (cm^{-1}): ϑ = 3295 (m), 2926 (m), 1678 (s), 1594 (s), 1532 (s), 1472 (m), 1368 (w), 1248 (w), 1117 (w), 995 (m), 839 (s), 782 (s); **HRMS** (ESI): calcd. for $C_{15}H_{27}N_3O_6Si^+$ [M+H]$^+$: 374.17419, found: 374.17463.

_N_-isoxazol-3-yl-_N'_-(1'-(5'-_O_-(4,4'-dimethoxytrityl)-2'-_O-tert_-butyldimethylsilyl)-β-D-ribofuranosyl)-urea (14)



α-13
β-13

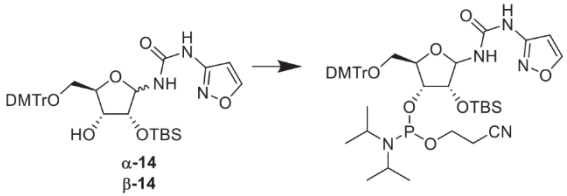
α-14
β-14

α-13 or **β-13** (1.00 eq.) was dissolved in pyridine and pre-dried DMT-Cl (1.40 eq.) was added. The resulting mixture was stirred at room temperature for 18 h. The Reaction was quenched by the addition of a few drops of methanol. Afterwards the solvent was evaporated *in vacuo* and the crude product was purified by column chromatography (silica, *i*Hex:EE = 2:1 + 0.1% pyridine) to yield **14** as a yellow foam.

β-14
yield: 58%; R_f = 0.31 (*i*Hex:EtOAc = 2:1);**^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 8.87 (s, 1H, NH), 8.54 (d, *J* = 1.5 Hz, 1H, 5-H), 7.51 – 7.49 (m, 2H, Ar-H), 7.38 – 7.35 (m, 4H, Ar-H), 7.31 – 7.28 (m, 2H, Ar-H), 7.25 – 7.21 (m, 2H, Ar-H), 7.18 – 7.16 (m, 2H, Ar-H, NH), 7.14-7.12 (m, 1H, Ar-H), 6.75 (d, *J* = 1.5 Hz, 1H, 4-H), 5.57 (dd, *J* = 9.5 Hz, *J* = 5.5 Hz, 1H, 1'-H), 4.28 (t, *J* = 5.3 Hz, 1H, 2'-H), 4.06 (td, *J* = 4.9 Hz, 3.7 Hz, 1H, 3'-H), 3.99 (q, *J* = 3.8 Hz, 1H, 4'-H), 3.77 (s, 6H, O-CH$_3$), 3.56 (d, *J* = 4.9 Hz, 1H, 3'-OH), 3.31 (dd, *J* = 10.2 Hz, 3.2 Hz, 1H, 5'-H), 3.10 (dd, *J* = 10.2; 4.2 Hz, 1H, 5'-H), 0.93 (s, 9H, Si-C-CH$_3$), 0.18 (s, 3H, Si-CH$_3$) 0.17 (s, 3H, Si-CH$_3$); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 159.93 (C3) 159.57 (C5), 159.49 (ArC-OCH$_3$), 154.04 (C=O), 146.21 (ArC), 136.97 (ArC), 136.78 (ArC), 131.02 (ArC), 139.76 (ArC), 129.07 (ArC), 128.59 (ArC), 127.50 (ArC), 126.12 (ArC), 113.87 (ArC), 99.08 (C4), 86.84 (C-Ar$_3$), 86.03 (C1'), 83.13 (C4'), 77.31 (C2'), 72.65 (C3'), 65.34 (C5'), 55.47 (O-CH$_3$), 26.21 (Si-C-CH$_3$), 18.89 (Si-C), -4.45 (Si-CH$_3$), -4.52 (Si-CH$_3$); **IR** (cm^{-1}): ṽ = 3306 (w), 2952 (m) 2929 (m), 2361 (w), 1708 (m), 1606 (m), 1507 (s), 1483 (m), 1363 (m), 1300 (m), 1249 (s), 1175 (s), 1030 (s), 992 (m), 909 (m), 833 (s), 780 (s), 701 (m); **HRMS** (ESI): calcd. for $C_{36}H_{45}N_3O_8Si^+$ [M+H]$^+$: 674.29032, found: 674.29055.

α-14:
yield: 78%; R_f = 0.30 (*i*Hex:EtOAc = 2:1)**;** **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.10 (s, 1H, NH), 8.51 (d, *J* = 1.7 Hz, 1H, 5-H), 7.53 – 7.45 (m, 2H, Ar-H), 7.37 – 7.33 (m, 4H, Ar-H), 7.33 – 7.28 (m, 2H, Ar-H), 7.25 – 7.20 (m, 1H, Ar-H), 6.91 – 6.86 (m, 4H, Ar-H), 6.79 (s, 1H, 4-H), 5.89 (dd, *J* = 9.5, 5.6 Hz, 1H, 1'-H), 4.57 – 4.52 (m, 1H, 2'-H), 4.11 (dt, *J* = 4.9, 3.2 Hz, 1H, 3'-H), 4.08 (q, *J* = 3.6 Hz, 1H, 4'-H), 3.88 (d, ^{3}*J* = 3.5 Hz, 1H, 3'-OH), 3.79 (s, 6H, O-CH$_3$), 3.28 (dd, *J* = 10.2, 3.6 Hz, 1H, 5'-H), 3.06 (dd, *J* = 10.2, 3.7 Hz, 1H, 5'-H), 0.93 (s, 9H, Si-C-CH$_3$), 0.17 (s, 3H, Si-CH$_3$), 0.15 (s, 3H, Si-CH$_3$; **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 159.6 (C3), 159.6 (C5), 159.6 (ArC-OCH$_3$), 154.1 (C=O), 137.0 (ArC), 136.8 (ArC), 131.0 (ArC), 129.0 (ArC), 128.6 (ArC), 127.5 (ArC), 113.9 (ArC), 99.2 (C4), 86.9 (C-Ar$_3$), 83.1 (C4'), 82.3 (C1'), 73.6 (C3'), 72.8 (C2'), 65.0 (C5'), 55.5 (O-CH$_3$), 55.5 (O-CH$_3$), 26.3 (Si-C-CH$_3$), 18.8 (Si-C), -4.6 (Si-CH$_3$), -4.7 (Si-CH$_3$); **IR** (cm^{-1}): ṽ = 3314 (w), 2953 (m) 2930 (m), 2361 (w), 1708 (m), 1606 (m), 1506 (s), 1484 (m), 1363 (m), 1300 (m), 1250 (s), 1176 (s), 1032 (s), 994 (m), 910 (m), 835 (s), 780 (s), 702 (m); **HRMS** (ESI): calcd. for $C_{36}H_{45}N_3NaO_8Si^+$ [M+Na]$^+$: 698.28736, found: 698.28764.

_N_-isoxazol-3-yl-_N'_-(1'-(5'-_O_-(4,4'-dimethoxytrityl)-3'-_O_-[2-cyanoethoxy(diisopropylamino)phosphino]-2'-_O-tert_-butyldimethylsilyl)-β-D-ribofuranosyl)-urea (8)



α-14
β-14

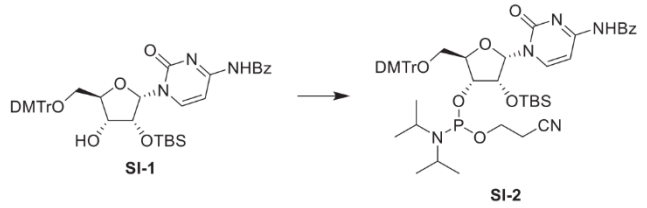
α-8
β-8

α-14 or **β-14** (1.00 eq.) was pre-dried by lyophilization from benzene (3×) and dissolved in DCM at 0 °C under argon atmosphere. DIPEA (4.00 eq.) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (2.50 eq.) were added dropwise. The resulting solution was stirred at room temperature for 3 h and subsequently quenched by the addition of sat. NaHCO$_3$ solution . The aqueous layer was extracted with DCM (3x). The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, and concentrated *in vacuo*. The residue was purified by flash column chromatography (silica, *i*Hex:EE = 2:1 + 0.1% pyridine, HPLC grade solvents) to yield **8** as mixture of diastereoisomers, as a colorless solid.

β-8:
yield: 86%; R_f = 0.40, 0.42 (*i*Hex:EtOAc = 2:1); **^{31}P NMR** (162 MHz, Acetone-d_6): δ (ppm) = 150.07, 147.99; **IR** (cm^{-1}): ṽ = 2927 (w), 1695 (m), 1593 (m), 1508 (s), 1463 (m), 1363 (w), 1298 (w), 1249 (s), 1177 (s), 1030 (s), 833 (s), 777 (s), 670 (m); **HRMS** (ESI): calcd. for $C_{45}H_{63}N_5O_9PSi^+$ [M+H]$^+$: 876.41272, found: 876.41281.

α-8:
yield: 84%; R_f = 0.40, 0.41 (*i*Hex:EtOAc = 2:1); **^{31}P NMR** (162 MHz, Acetone-d_6): δ (ppm) = 150.20, 148.93; **IR** (cm^{-1}): ṽ = 2926 (m), 1694 (m), 1590 (m), 1508 (s), 1464 (m), 1364 (w), 1230 (w), 1249 (s), 1176 (m), 1031 (s), 978 (m), 833 (s), 778 (s), 701 (m); **HRMS** (ESI): calcd. for $C_{45}H_{63}N_5O_9PSi^+$ [M+H]$^+$: 876.41272, found: 876.41341.

5'-_O_-(4,4'-dimethoxytrityl)-3'-_O_-[2-cyanoethoxy(diisopropylamino)phosphino]-2'-_O-tert_-butyldimethylsilyl)-4-_N_-benzoyl-α-cytidine (SI-2)



SI-1

SI-2

5'-O-(4,4'-dimethoxytrityl-2'-O-tert-butyldimethylsilyl-4-*N*-benzoyl-α-cytidine (**SI-1**) was synthesized in an eight-step procedure according to literature.[2-3] **SI-1** (170 mg, 223 µmol, 1.00 eq.) was pre-dried by lyophilization from benzene and dissolved in DCM (3 mL) at 0 °C under argon atmosphere. DIPEA (155 µL, 890 µmol, 4.00 eq.) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (124 µL, 556 µmol, 2.50 eq.) were added dropwise. The resulting solution was stirred at room temperature for 5 h and subsequently quenched by the addition of sat. NaHCO$_3$ solution (5 mL). The aqueous layer was extracted with DCM (3 x 10 mL). The combined organic layers were washed with brine (10 mL), dried over Na$_2$SO$_4$, and concentrated *in vacuo*. The residue was purified by flash column chromatography (silica, *i*Hex:EE = 2:1 → 1:2 + 0.1% pyridine, HPLC grade solvents) to yield **SI-2** (180 mg, 186 µmol, 84%) as a mixture of diastereoisomers, as a pale-yellow foam.

R_f = 0.50, 0.52 (*i*Hex:EtOAc = 1:2); **^{31}P NMR** (162 MHz, Acetone-d_6): δ (ppm) = 150.16, 150.02; **IR** (cm^{-1}): ṽ = 2929 (w), 1667 (m), 1622 (m), 1554 (m), 1508 (m), 1483 (s), 1395 (m), 1299 (m), 1249 (s), 1177 (s), 1157 (m), 1069 (w), 1031 (s), 978 (m), 927 (w), 829 (s), 779 (s), 704 (s); **HRMS** (ESI): calcd. for $C_{52}H_{66}N_5O_9PSi^+$ [M+H]$^+$: 964.44402, found: 964.44278.

Synthesis and purification of oligonucleotides

Phosphoramidites of canonical ribonucleosides (Bz-A-CE, Dmf-G-CE, Ac-C-CE and U-CE) were purchased from LinkTech and Sigma-Aldrich. Oligonucleotides were synthesized on a 1 µmol scale using RNA SynBase™ CPG 1000/110 as solid supports using an RNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with a standard phosphoramidite chemistry. Oligonucleotides were synthesized in DMT-OFF mode using DCA as a deblocking agent in CH_2Cl_2, Activator 42® as activator in MeCN, Ac_2O as capping reagent in pyridine/THF and I_2 as oxidizer in pyridine/H_2O. The cleavage and deprotection of the CPG bound oligonucleotides were performed with a 1:1 aqueous solution mixture (0.6 mL) of 30% NH_4OH and 40% $MeNH_2$. The suspension was heated at 65ºC for 5 min for SynBase™ CPG 1000/110. Subsequently, the supernatant was collected, and the beads were washed with water (2×0.5 mL). The combined aqueous solutions were concentrated under reduced pressure using a SpeedVac concentrator. After that, the crude was dissolved in DMSO (100 µL) and triethylamine trihydrofluoride (125 µL) was added. The solution was heated at 65 ºC for 1.5 h. Finally, the Oligonucleotides were precipitated by adding 3 M NaOAc in water (25 µL) and *n*-BuOH (1 mL). The mixture was kept at -80 ºC for 2 h and centrifuged at 4ºC for 1 h. The supernatant was removed, and the white precipitate was lyophilized. The oligonucleotides were further purified by semi-preparative reverse-phase HPLC using a 1260 Infinity II Manual Preparative LC System from Agilent (G7114A detector) equipped with the column VP 250/10 Nucleodur 100-5 C18ec from Macherey Nagel. A flow rate of 5 mL/min with varying gradients between 0-15% and 0-40% of buffer B in 45 min was applied for the purifications. The following buffer system was used: buffer A: 100 mM NEt_3/HOAc (pH 7.0) in H_2O and buffer B: 100 mM NEt_3/HOAc in 80% (v/v) acetonitrile. The purified oligonucleotides were analyzed by analytical RP-HPLC on a 1260 Infinity II LC System from Agilent (G7165A detector) equipped with the column an EC 250/4 Nucleodur 100-3 C18ec from Macherey Nagel using a flow of 1 mL/min, a gradient of 0-15% or 0-20% of buffer B in 45 min was applied. Finally, the purified oligonucleotides were desalted using a C18 RP-cartridge from Waters. The absorbance of the synthesized oligonucleotides in H_2O solution were measured using an IMPLEN NanoPhotometer® N60/N50 at 260 nm. The extinction coefficients of the oligonucleotides were calculated using the OligoAnalyzer Version 3.0 from Integrated DNA Technologies. For strands containing mainly Isoxazole, extinctions coefficients were calculated based on the base composition method at 223 nm using estimated extinctions coefficients (e.g., 10726 $M^{-1}cm^{-1}$ for IO3 and 7616 $M^{-1}cm^{-1}$ for cytidine). The structural integrity of the synthesized oligonucleotides was analyzed by MALDI-TOF mass measurement. For this purpose, the synthesized oligonucleotides (2-3 µL) were desalted on a 0.025 µm VSWP filter (Millipore), co-crystallized in a 3-hydroxypicolinic acid matrix (HPA, 1 µL) and measured on a Bruker Autoflex II. UV spectra, melting profiles and the concentrations of purified oligonucleotides were measured on a JASCO V-650 spectrometer.

WILEY-VCH

SUPPORTING INFORMATION

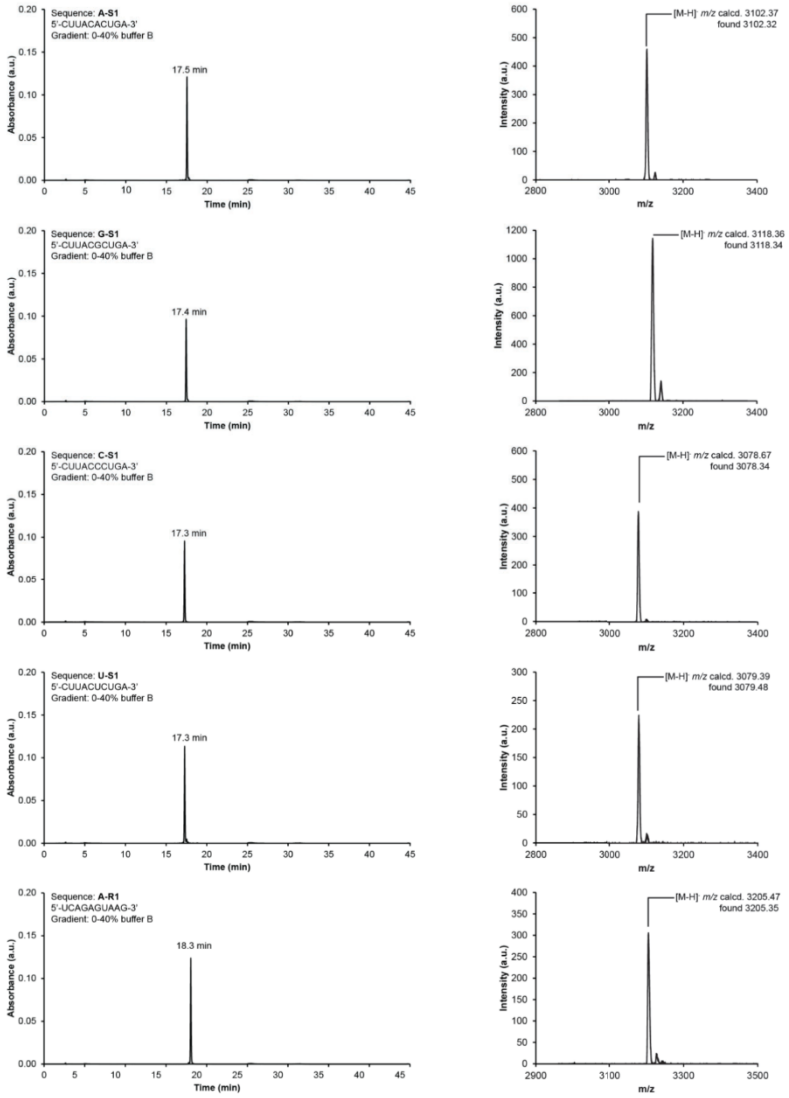


Figure S1. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the purified Oligonucleotides

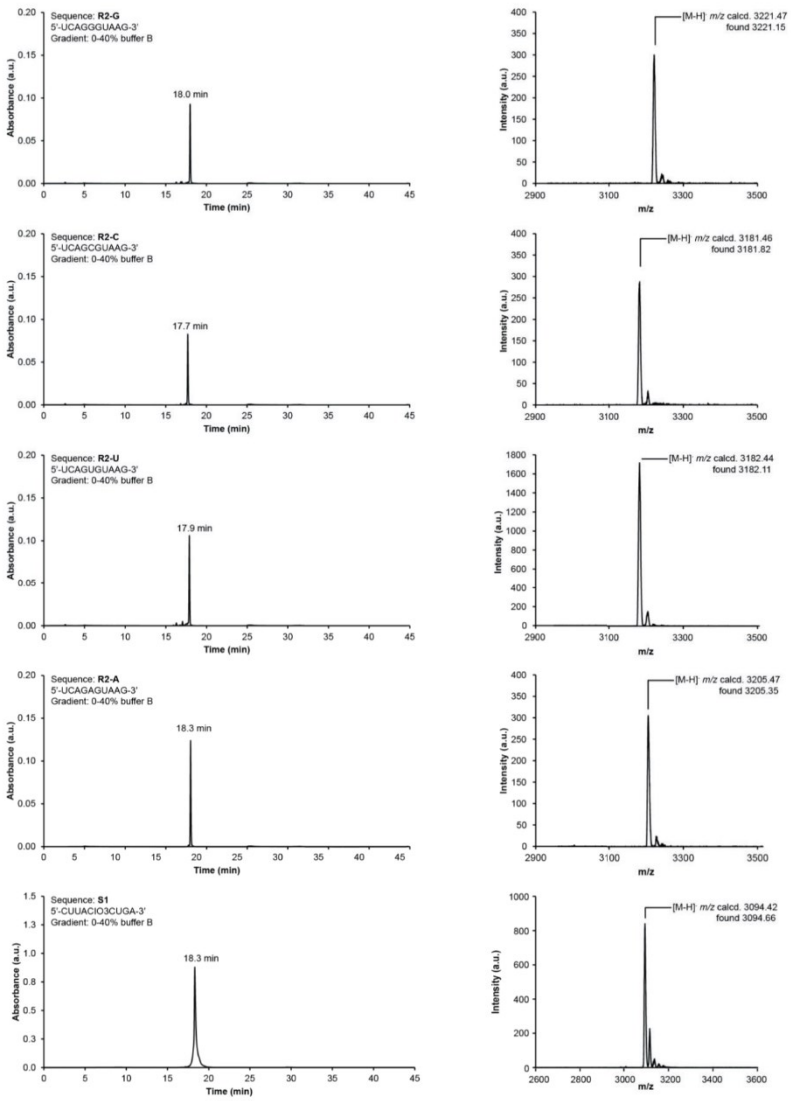


Figure S2. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the purified Oligonucleotides (continued)

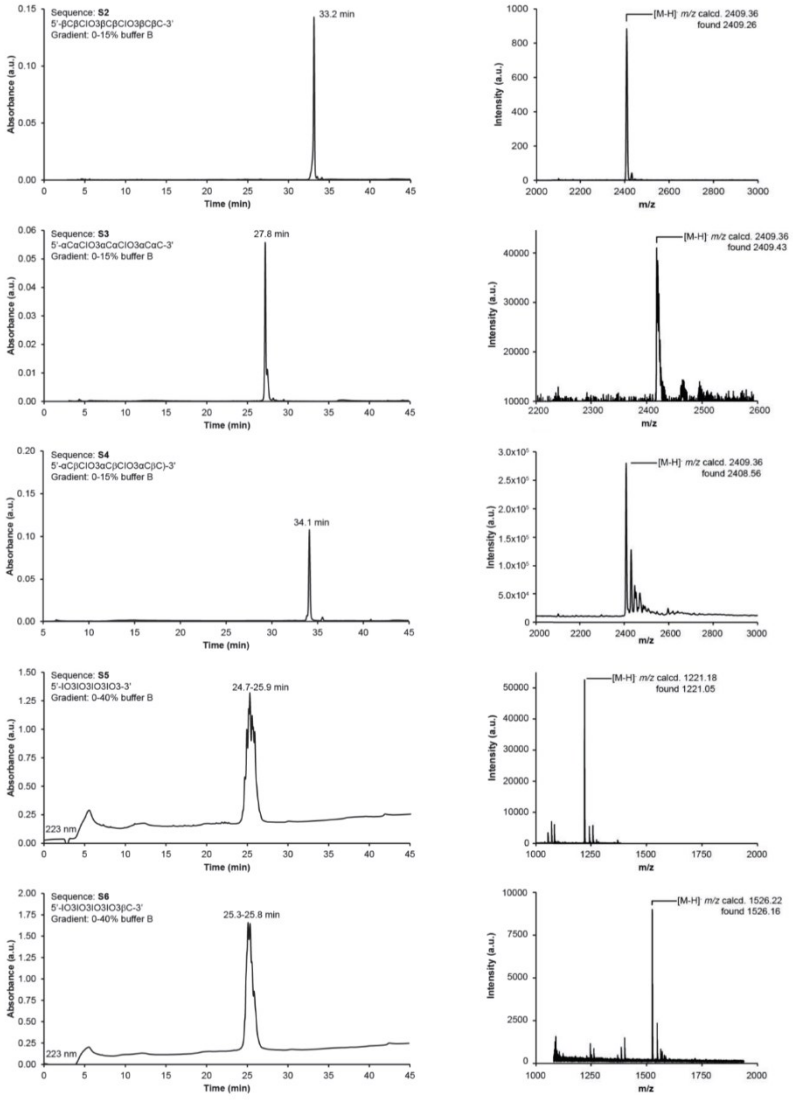


Figure S3. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the purified Oligonucleotides (continued)

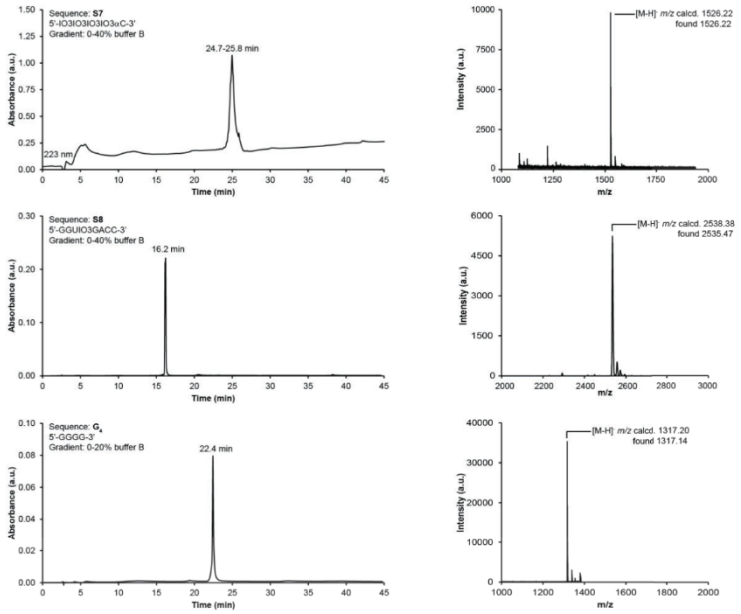


Figure S4. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the purified Oligonucleotides (continued)

UV Melting Curve Measurements

The UV melting curves were measured on JASCO V-650 spectrometer using 10 mm QS cuvettes, purchased from Hellma Analytics. A solution (80 µL) of equimolar amounts of oligonucleotides **S1** and **R1** (4 µM each) in the buffer solution containing 10 mM sodium phosphate buffer (pH 7.0) and 150 mM NaCl was heated at 50°C for 5 min and gradually cooled to 4°C prior to the measurement. Melting profiles were recorded at temperatures between 5 and 70°C with a ramping and scanning rate of 1°C/min at 260 nm. All samples were measured at least three times.

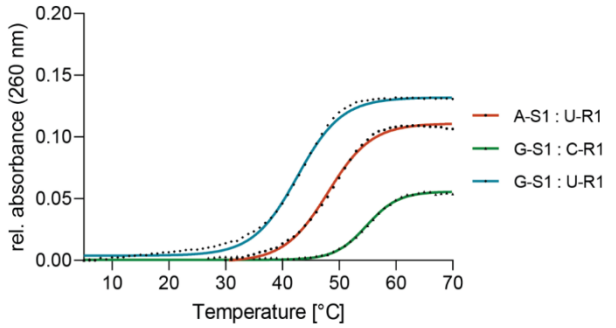


Figure S5. UV Melting curves of dsRNAs: **A-S1:U-R1** (T_m = 48.0°C), **G-S1:C-R1** (T_m = 54.8°C), **G-S1:U-R1** (T_m = 42.6°C). Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function.

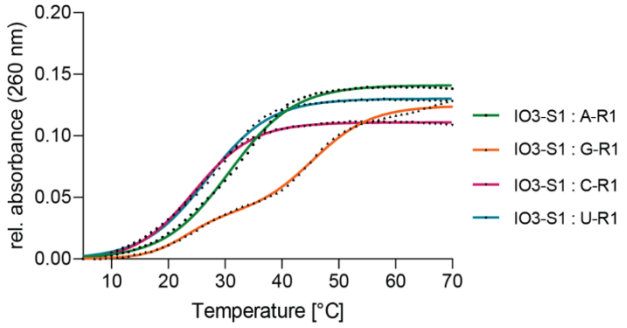


Figure S6. UV Melting curves of dsRNAs: **IO3-S1:A-R1** (T_m = 30.7°C), **IO3-S1:G-R1** (T_m = 44.6°C), **IO3-S1:C-R1** (T_m = 24.5°C),), **IO3-S1:C-R1** (T_m = 26.8°C). Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function.

In-strand cytidine formation reactions

Stock solutions of boric acid (pH 9.7, 100 mM) and Na_2CO_3 (1000 mM) were prepared in water. The oligonucleotide **S1**-**S7** (2-10 nmol) was with mixed with buffer, Na_2CO_3, and water. The Fe^{2+} source (FeS or FeS_2) and DTT were added to the mixture. The final concentration of the components: 100 uM of Oligo, 50 mM Buffer, 100 mM Na_2CO_3, 100 mM Fe^{2+}, 300 mM DTT. The mixture was heated at 90 °C for 2 h in a *TAdvanced Thermocycler* by *Biometra*. After cooling to room temperature, the solids were removed by centrifugation and washed with water (2×0.2 mL), residual solids were removed using a syringe filter (0.20 um, PTFE-membrane). The reaction mixture was concentrated by lyophilization and subsequently analyzed by reverse-phase HPLC. The yields of the reactions were calculated by integration of the chromatographic peaks of the products and the use of the calibration curves of the synthetically prepared product.

Digestion and LC-HESI-MS analysis

Reaction buffer 10X and *Enzyme mix* was bought as a Nucleoside Digestion Mix (M0649S) kit (*New England BioLabs Inc.*). The purified oligonucleotide (250-500 ng in 46 uL) was incubated with *Reaction buffer 10X* (5 uL) and *Enzyme mix* (1 uL) at 37°C for 1.5 h. The mixture was subsequently diluted to 90 uL and then analyzed by LC-HESI-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 YMC-Triart C18 column column with a flow of 0.15 ml/min and a constant column temperature of 30 °C The following buffer system was used: buffer A: 2 mM $HCOONH_4$ in H_2O (pH 5.5) and buffer B: 2 mM $HCOONH_4$ in 80% (v/v) acetonitrile (pH 5.5). 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. The synthetic standards for the co-injection experiments were synthesized in our lab (see synthetic procedures or according to reported literature[3]) or purchased.

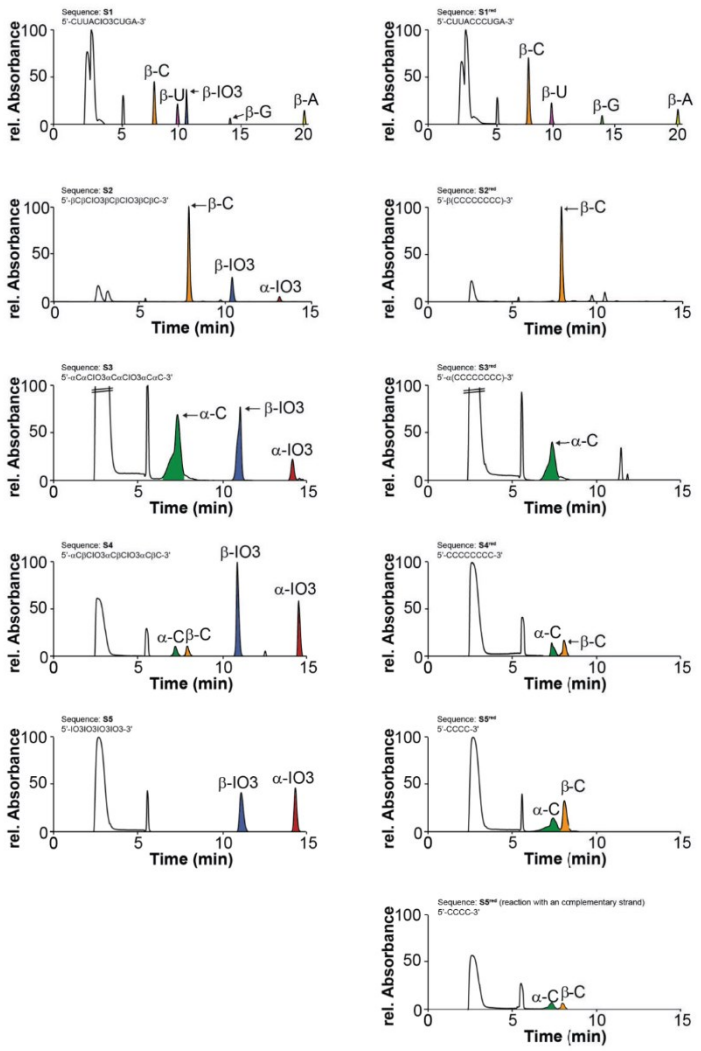


Figure S7. HPLC-MS chromatograms of strands **S1-S7** (left) and their respective N-O cleavage and cyclization products **S1^{red}-S7^{red}** (right).

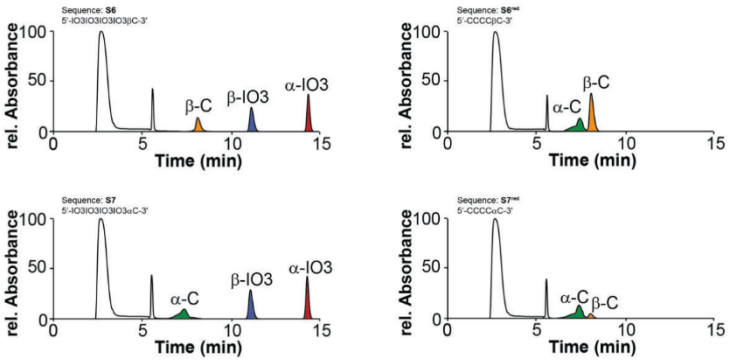


Figure S8. HPLC-MS chromatograms of strands **S1-S7** (left) and their respective N-O cleavage and cyclization products **S1**^{red}**-S7**^{red} (right).(continued)

High-resolution NMR studies of IO-containing RNA oligonucleotides

Canonical dsRNA sample **Can$_8$** (Figure S1) was purchased after purification via high-performance liquid chromatography (HPLC) from Ella Biotech GmbH (Martinsried, Germany). The modified dsRNA sample **S7** (Figure S1) was synthesized as described above. We dissolved both Can$_8$ and S7 samples in an aqueous phosphate buffer. The buffer consisted in Na_2HPO_4/NaH_2PO_4 (15 mM, pH 7.0) and NaCl (25 mM) in H_2O. We added sodium trimethylsilylpropanesulfonate (DSS) (50 µm) as a chemical shift standard for proton NMR spectroscopy, NaN_3 (0.01% m/w) to avoid the growth of bacteria in the sample and D_2O (5%, 9.0 µL) as a lock reference for the spectrometer. Both samples were measured in 3 mm NMR tubes, with a final volume of 0.189 mL and a concentration of 0.53 mM and 1.1 mM, respectively. Experiments were performed on a Bruker Avance III spectrometer operating at a ^{1}H Larmor frequency of 800 MHz equipped with a 5 mm triple channel ^{1}H, ^{13}C and ^{15}N cryoprobe. Spectra were acquired and processed using Topspin 3.5 and 4.0, respectively, and further analyzed by NMRFAM-Sparky.[4] Unless otherwise stated, all experiments were measured at 2 °C to ensure the presence of a single conformer in solution. ^{1}H–^{1}H NOESY spectra were measured with mixing time points at 40, 80, 120, 160, 200, 240, 280, and 320 ms; with 8192 x 1024 complex points, 234 and 29.3 ms acquisition times in the direct and indirect dimension, respectively. Water suppression was achieved using excitation sculpting pulse sequences. The experimental time was about 6 hours for each 2D NOESY spectrum. ^{1}H–^{1}H TOCSY spectra were recorded with a mixing time of 80 ms using MLEV17 mixing scheme, with 4096 x 256 complex points, 51 and 16 ms acquisition times in the direct and indirect dimension, respectively. For the purpose of comparison, the same measurements were performed for the canonical and the modified 8-mer. The used methods as well as the conditions for both samples are listed below. For resonance assignment standard ^{1}H–^{1}H NOESY spectra were measured as well as a TOCSY for the modified oligomer. For more detailed analysis a natural abundance ^{1}H–^{15}N HMQC was recorded using the SOFAST pulse sequence and a ^{1}H–^{13}C HSQC was recorded using a Bruker standard pulse sequence. The basis of the assignment of a NOESY spectra is provided by so-called 'sequential paths', which are formed by NOE cross peaks from adjacent ribose backbone base segments.

Molecular modelling of IO-containing RNA oligonucleotides

For the structure calculation the parameters for the modified base first needed to be incorporated into the topology and the parameter files used by CNSsolve to generate an RNA structure.[5] The bond length, angles as well as dihedral angles were determined using an optimized structure of the modified nucleoside. The calculation for this purpose was performed using a standard Hartree Fock method with a 6-31G base. After the integration of the modified base, the three-dimensional structure was determined by an *in-silico* annealing procedure. The restriction data used as input for the calculation has been directly derived from the ^{1}H–^{1}H NOESY spectra. Due to the characteristic of nucleic acids having a limited amount of protons, the data obtained from the ^{1}H–^{1}H NOESY alone is not sufficient to provide the software with enough input for a *de novo* structure calculation. For this reason, the sugar-phosphate backbone the dihedral angles were determined according to the literature values for an A-form RNA, which was assumed due to the consistency of the fingerprint region with said conformation. For all canonical bases (meaning all the bases except for IO3) the dihedral angles, bond length and bond angles embedded in the CNSsolve program were used. Besides the ^{1}H–^{1}H distances extracted from NOESY experiments, the distances between carbon, oxygen and nitrogen atoms were identified, gauging a 3D model of the canonical structure using Chimera 1.14. The distances were added to the restraint distance file supplementing the NOESY data. Approximate ^{1}H–^{1}H distances were defined by the NOESY spectra recorded at different (40–320 ms) mixing times. The distances could be divided into two categories based on cross-peaks intensities. Protons close in space are therefore visible for short mixing times, conversely protons further away only appear at higher mixing times. The extracted peak heights were converted to distances taking the cross-peak between C7H5 and C7H6 as a reference due to the fixed and known distance of 2.421 Å between the two protons. The values were generated according to the relationship that the intensity is reciprocal to the sixth of the distance, $J \sim 1/r^6$.[6] Overlapping signals were excluded from the analysis and did not contribute to the molecular modelling procedure.

Schematic sample representation

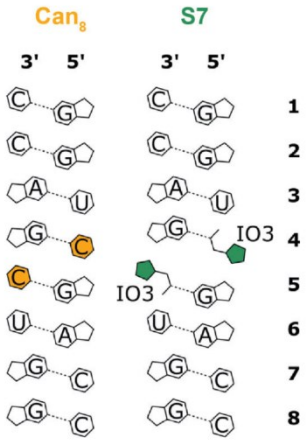


Figure S9. Scheme of the canonical (Can_8, left) and modified (S7, right) RNA samples

Homo- and heteronuclear correlation spectra

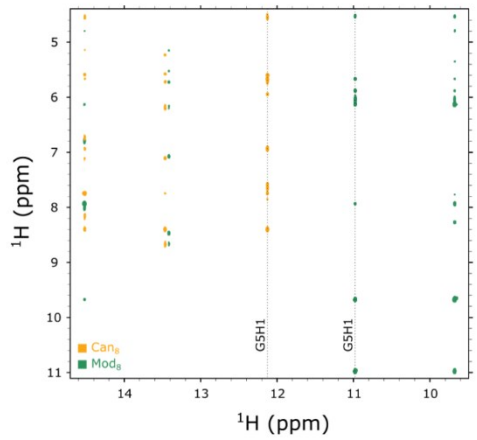


Figure S10. Overlap of the NOESY spectra measured with τ=320 ms for Can_8 and S7 showing the imino region. The chemical shift difference of G5H1, reporting on the stability of the IO3-G base pair, is highlighted by the dotted lines.

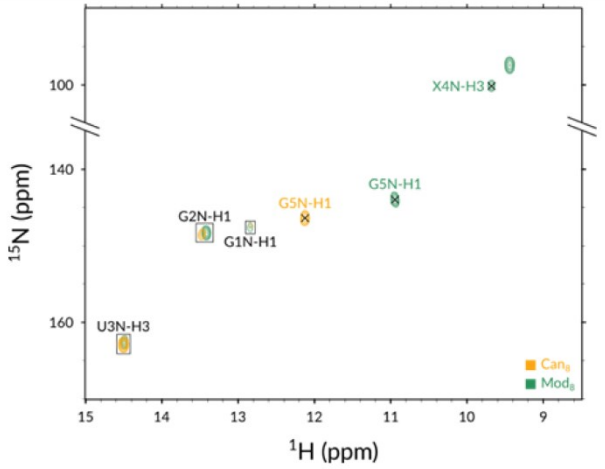


Figure S11. Comparison of the ^{1}H-^{15}N HMQC spectra for Can_8 (yellow) and S7 (green) highlighting the nuclei involved in base pairing

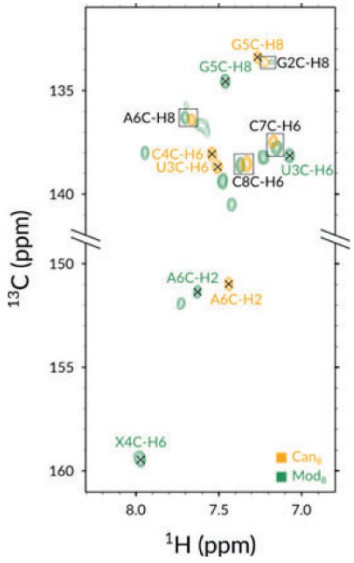


Figure S12. Comparison of the aromatic region of ^{1}H-^{13}C HSQC spectra for Can_8 (yellow) and S7 (green).

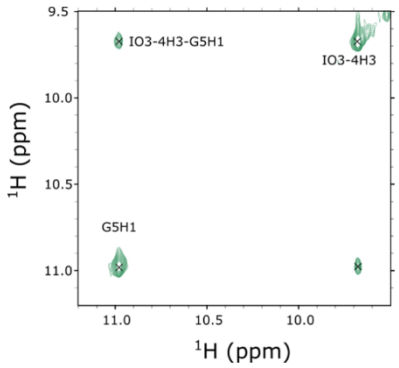


Figure S13. Excerpt from the ^{1}H-^{1}H NOESY spectrum (t_{mix} = 40 ms) of S7 showing the interstrand cross-peak between G5H1 and IO3-4H3.

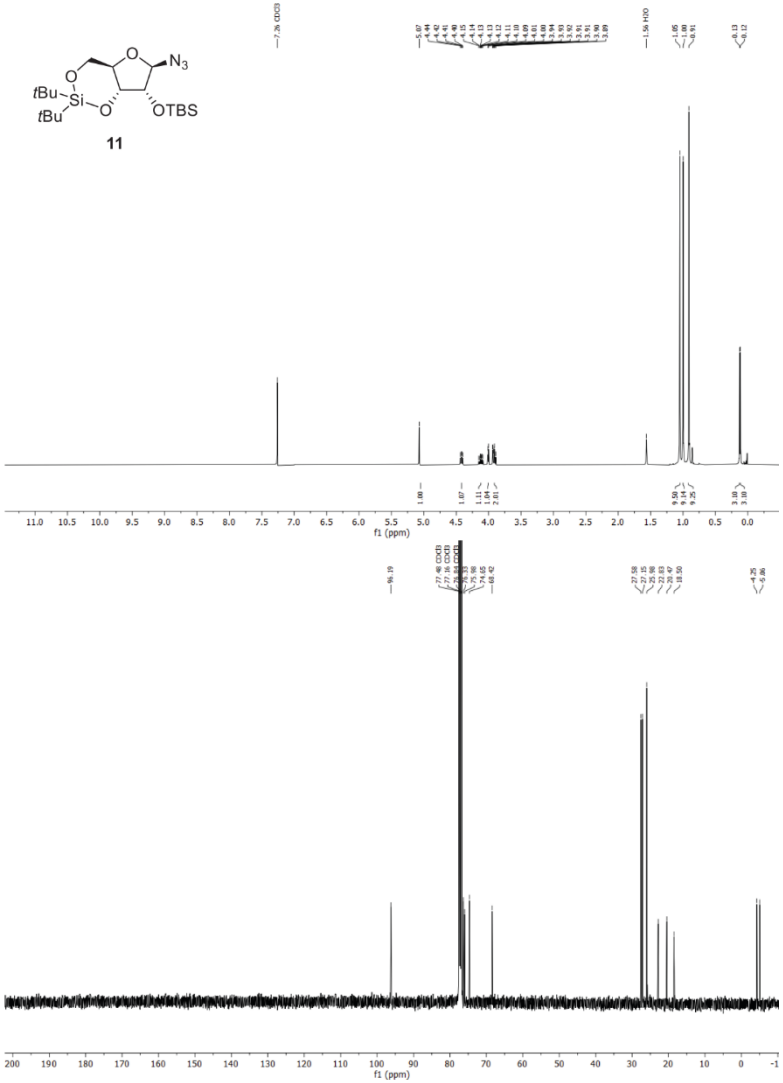
Chemical shift assignment tables

Table S1. Chemical shifts of the signals for Can$_8$ and S7

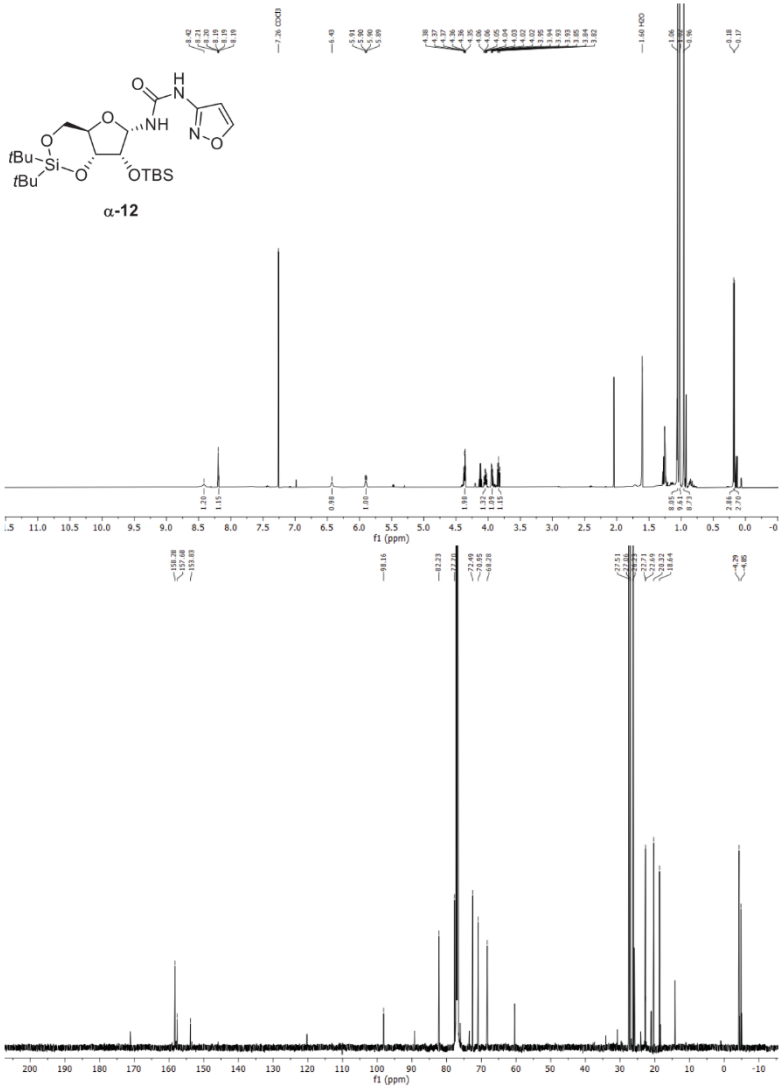
Nucleobase	Nucleus	Can$_8$ chemical shift	S7 Chemical shift
G2	H1	13.464	13.414
	N1	148.5	148.3
	H1'	5.927	5.912
	H2'	4.533	4.567
	H3'	4.615	4.476
	H5'	4.144	4.151
	H8	7.515	7.551
	C8	133.70	133.61
U3	H1'	5.580	5.529
	H2'	4.549	4.837
	H3	14.517	14.520
	N3	162.9	162.6
	H5	5.140	4.799
	H5'	4.112	4.115
	H6	7.819	7.385
	C6	139.07	138.44
X4	H1'	5.594	5.348
	H2'	4.565	4.539
	H3'	4.459	4.881
	H5'	4.126	3.994
G5	H1	12.126	10.979
	N1	146.3	144.4
	H1'	5.602	5.674
	H3'	4.498	4.525
	H5'	4.132	4.149
	H8	7.577	7.772
	C8	133.78	134.94
A6	H1'	5.946	5.879
	H2	7.746	7.934
	C2	151.28	151.65
	H2'	4.568	4.531
	H3'	4.682	4.679
	H5'	4.135	4.139
	H8	7.982	8.025
	C8	136.73	136.68
C7	H1'	5.381	5.439
	H2'	4.151	4.210
	H3'	4.360	4.355
	H41	8.400	8.469
	H42	7.109	7.080
	H5	5.231	5.157
	H6	7.481	7.469
	C6	137.84	138.15
C8	H1'	5.722	5.730
	H2'	4.025	4.178
	H41	8.412	8.365
	H42	6.990	7.006
	H5	5.476	5.475
	H5'	3.920	4.038
	H6	7.638	7.674
	C6	138.89	138.97

NMR spectra of synthesized compounds

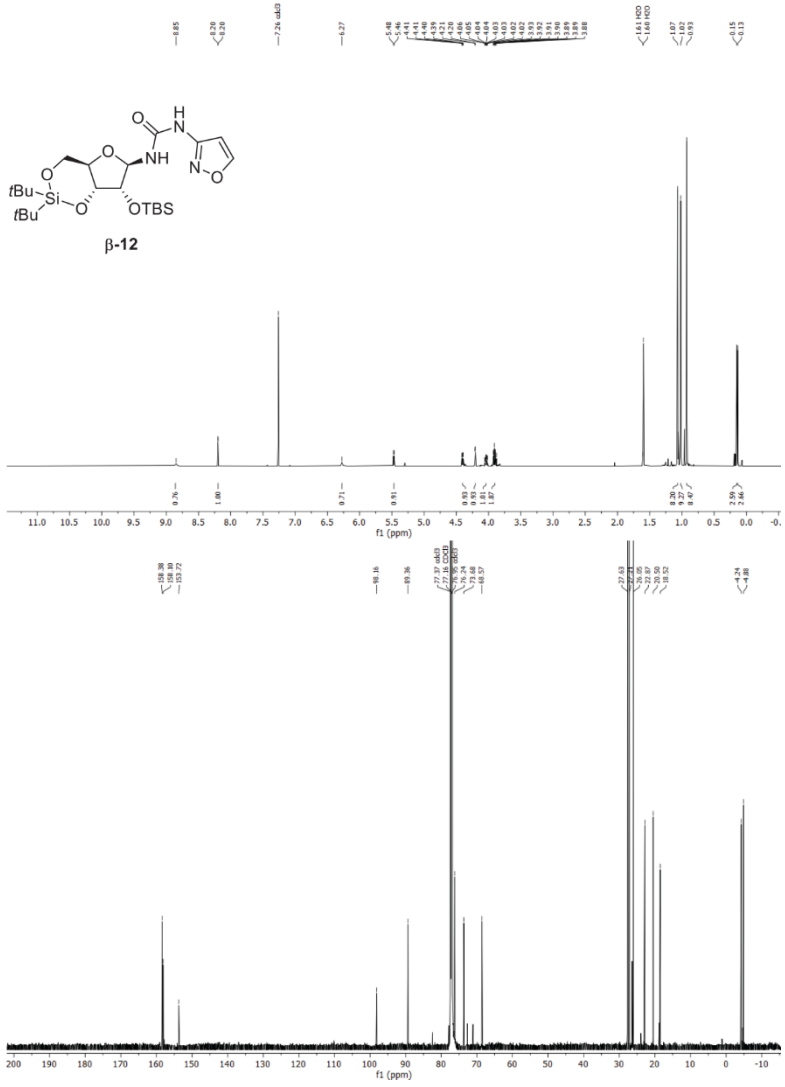
^{1}H and ^{13}C{^{1}H} NMR spectra of compound 11



^{1}H and ^{13}C{^{1}H} NMR spectra of compound α-12

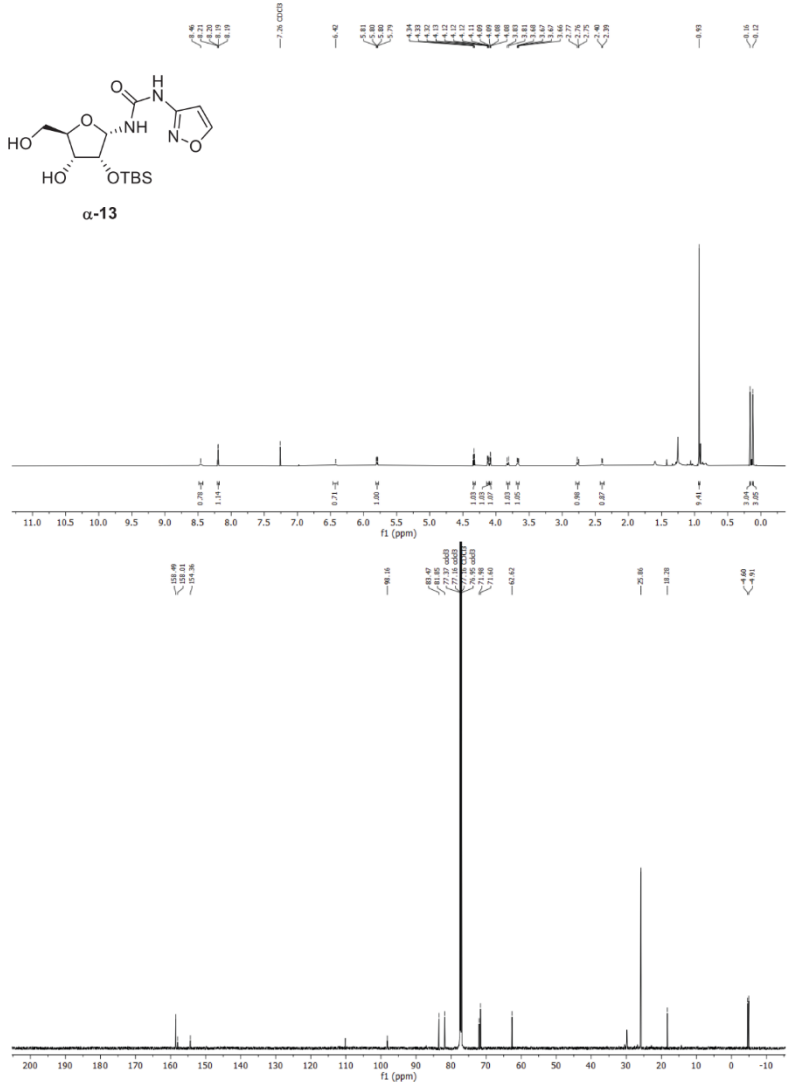


^{1}H and ^{13}C{^{1}H} NMR spectra of compound β-12



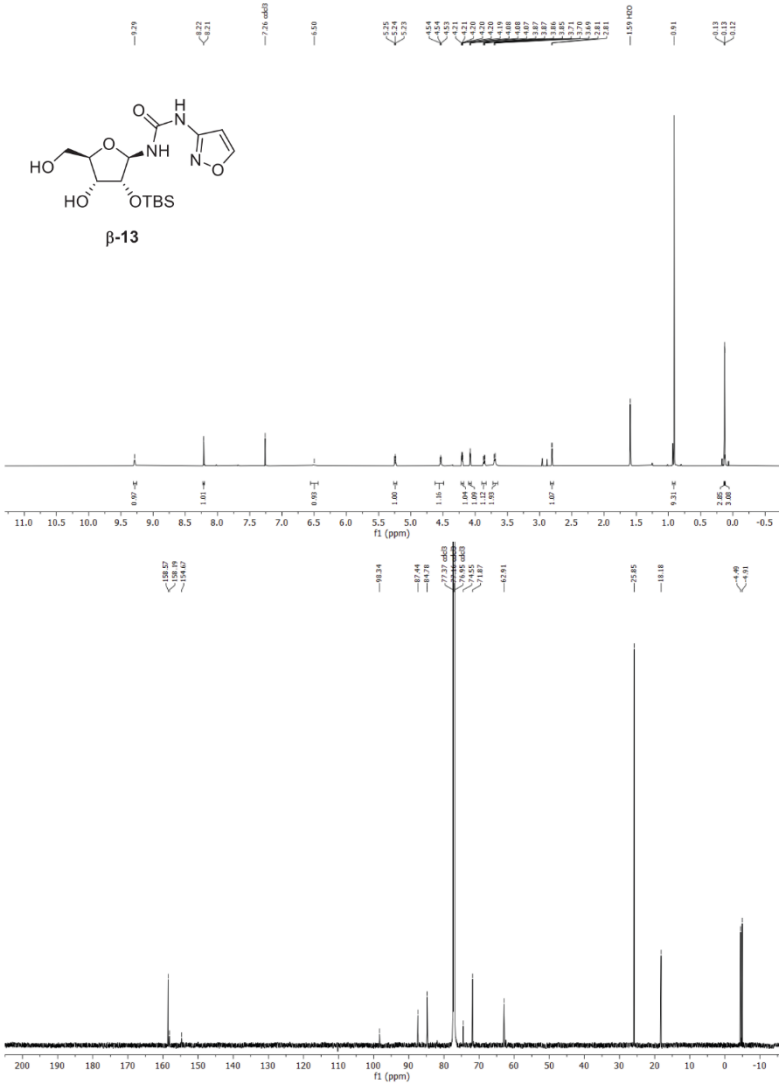
β-12

^{1}H and ^{13}C{^{1}H} NMR spectra of compound α-13

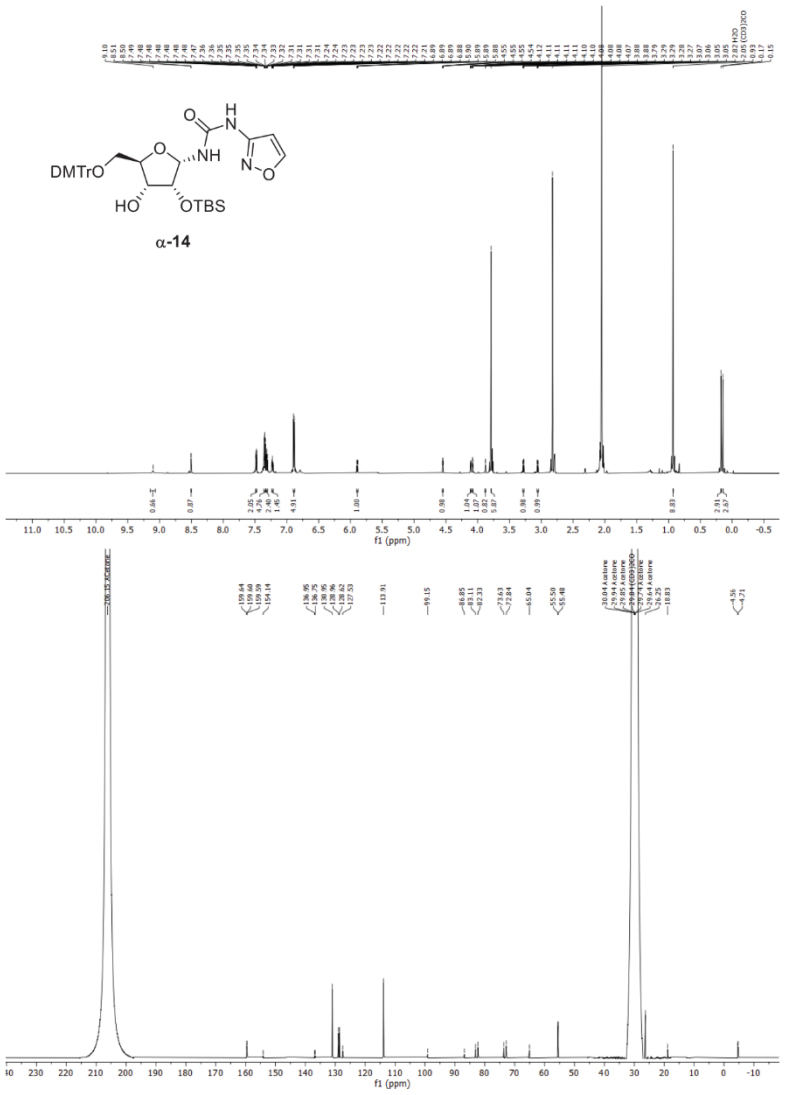


α-13

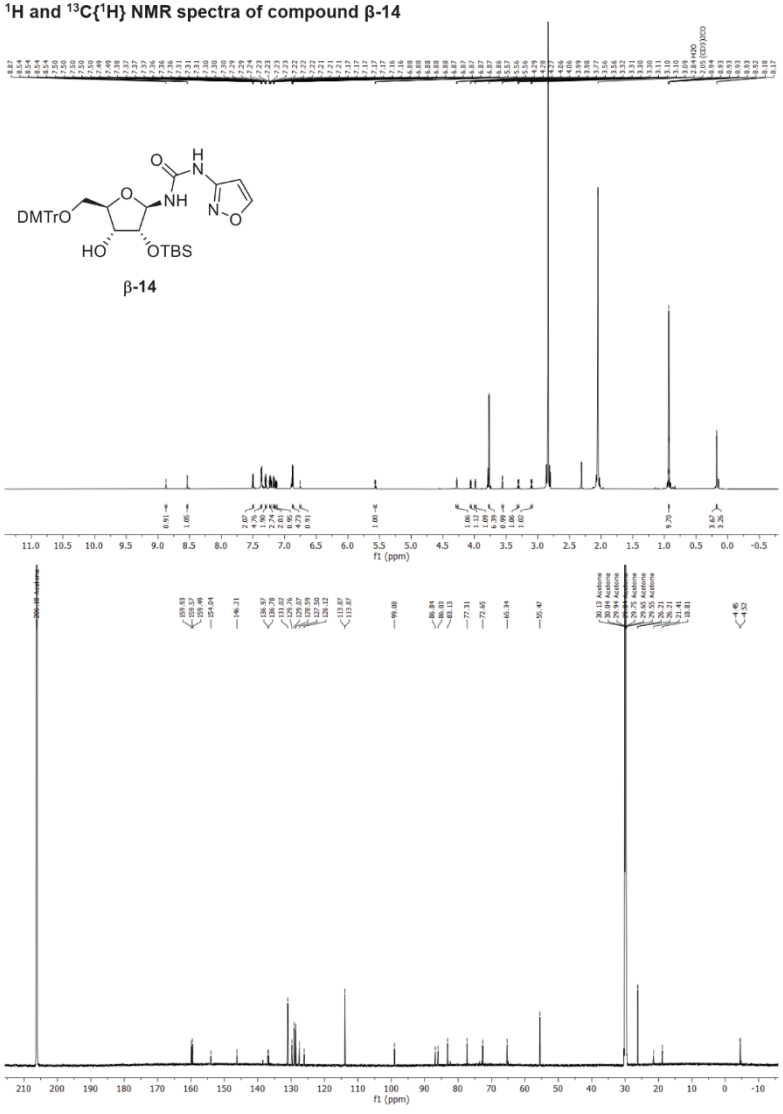
^{1}H and ^{13}C{^{1}H} NMR spectra of compound β-13



^{1}H and ^{13}C{^{1}H} NMR spectra of compound α-14

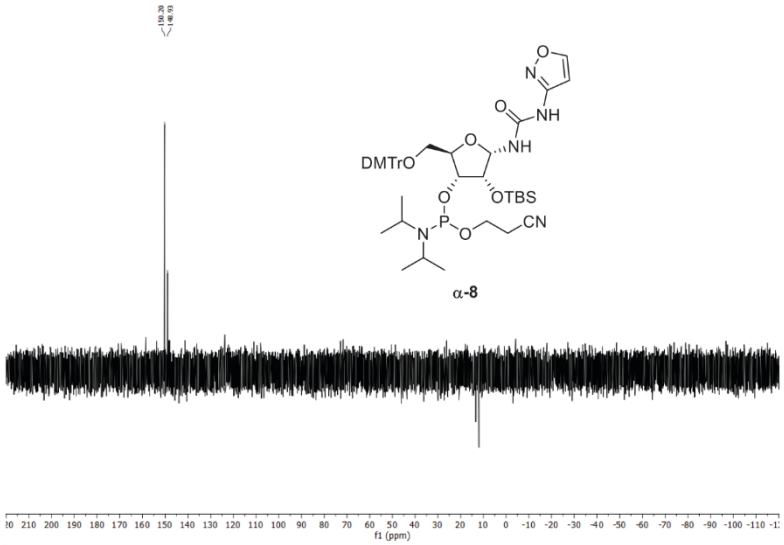


^{1}H and ^{13}C{^{1}H} NMR spectra of compound β-14



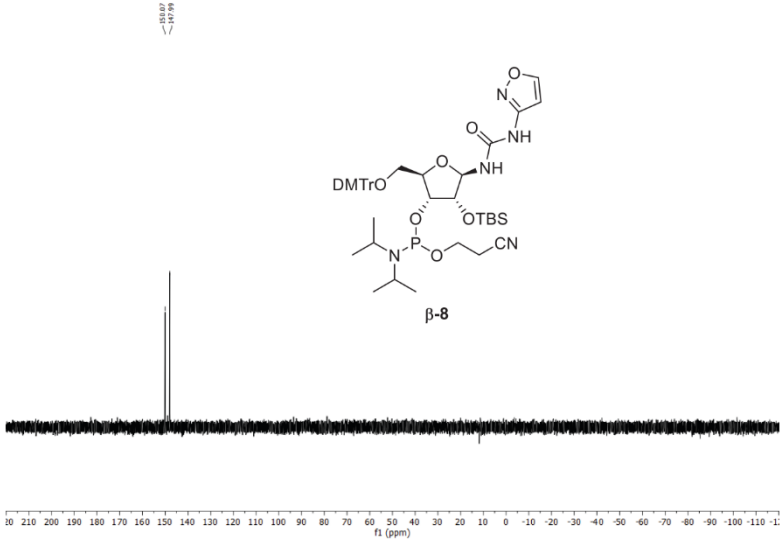
β-14

^{31}P {^{1}H} NMR spectra of compound α-8



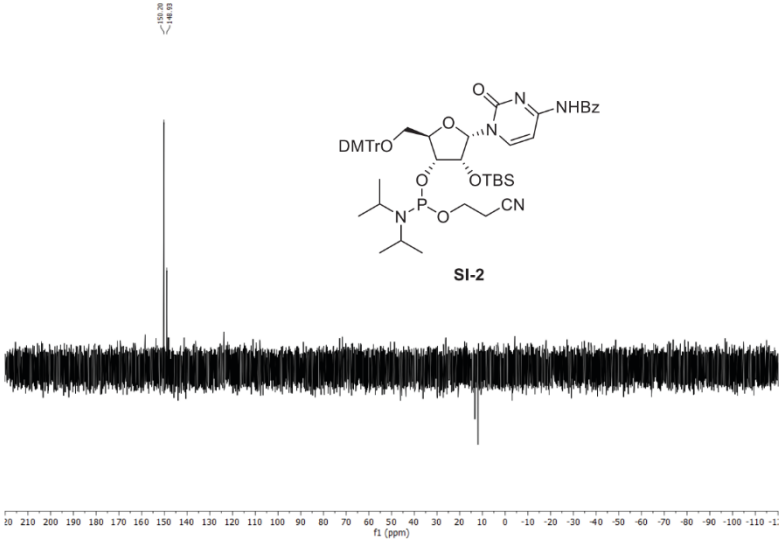
α-8

^{31}P {^{1}H} NMR spectra of compound β-8



β-8

^{31}P {^{1}H} NMR spectra of compound β-8



SI-2

References

[1] H. Okamura, A. Crisp, S. Hübner, S. Becker, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2019**, *58*, 18691-18696.
[2] F. Debart, B. Rayner, G. Degols, J. L. Imbach, *Nucleic Acids Res.* **1992**, *20*, 1193-1200.
[3] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76.
[4] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325-1327.
[5] A. T. Brunger, *Nature Protocols* **2007**, *2*, 2728-2733.
[6] M. P. Williamson, in *Modern Magnetic Resonance* (Ed.: G. A. Webb), Springer Netherlands, Dordrecht, **2006**, pp. 409-412.

Anhang III

ChemistryEurope

Supporting Information

An Aminoisoxazole-Based Proto-RNA

Felix Xu, Stefan Wiedemann, Jonas Feldmann, Sidney Becker, and Thomas Carell*

Table of Contents

General Experimental Methods

Chemicals were purchased from Sigma-Aldrich, TCI, Fluka, ABCR, Carbosynth, Carbolution or Acros Organics and used without further purification. The solvents were of reagent grade or purified by distillation. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F_{254} TLC plates from Merck KGaA. Flash column chromatography was performed on Silica gel 60 (40-63 µm) silica gel from Macherey-Nagel. Visualization of the developed TLC plates was achieved through UV-absorption or through staining with Hanessian's stain, *p*-anisaldehyde or KMnO4 stain. Reactions were conducted under a positive pressure of dry nitrogen in oven-dried glassware, and at ambient room temperature, unless otherwise specified. NMR spectra were recorded on Bruker Avance III HD 400 (400 MHz), Bruker Ascend 500 (500 MHz) or Bruker Avance III (800 MHz) spectrometers. ^{1}H NMR shifts were calibrated to the residual solvent resonances: DMSO-d_6 (2.50 ppm), D_2O (4.79 ppm), CD_3OD (4.87 ppm), Acetone-d_6 (2.05 ppm) or $CDCl_3$ (7.26 ppm). ^{13}C NMR shifts were calibrated to the residual solvent: DMSO-d_6 (39.52 ppm), CD_3OD (49.00 ppm), $CDCl_3$ (77.16 ppm), Acetone-d_6 (29.84 ppm). All NMR spectra were analyzed using the program MestReNOVA 10.0.1 from Mestrelab Research. NMR data are reported as follows: chemical shift (multiplicity, coupling constants where applicable, number of hydrogens, assignment). Splitting is reported with the following symbols: s = singlet, d = doublet, t = triplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublets of doublets, m = multiplet. S. L. Normal resolved mass spectra were measured on a LTQ FT-ICR by Thermo Finnigan GmbH. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of the Ludwig-Maximilians-Universität München on the following spectrometers (ionization mode in brackets): MAT 95 (EI) and MAT 90 (ESI) from Thermo Finnigan GmbH, unless otherwise specified. IR spectra were recorded on a Shimadzu IRSpirit FT-IR instrument or a PerkinElmer Spectrum BX II FT-IR system. Substances were applied as a film or directly as solids on the ATR unit. 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 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 H_2O/MeCN 20/80 (pH 5.5)). The gradient for isoxazole or pyrimidine nucleosides containing compounds and nucleosides was 0 → 25 min, 0% → 10% buffer B. The elution was monitored at 210,223, 243 and 260 nm. 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. The synthetic standards for the co-injection experiments were synthesized in our lab according to reported literature[1] or purchased. Semi-preparative reverse-phase HPLC separations were performed using a 1260 Infinity II Manual Preparative LC System from Agilent (G7114A detector) equipped with the column VP 250/10 Nucleodur 100-5 C18ec from Macherey Nagel. A flow rate of 5 mL/min with varying gradients between 0-10% to 10-20% of buffer B in 45 min was applied for the purifications. The following buffer system was used: 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)).

SUPPORTING INFORMATION WILEY-VCH

Prebiotic Synthesis and Reference Molecule Synthesis

5-Aminoisoxazole (3)

3

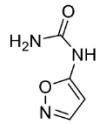
From cyanoacetylene:

To a solution of hydroxylamine (50 wt% in H_2O, 6.72 µL, 110 µmol, 1.1 eq.) and $NaHCO_3$ (8.40 mg, 100 µmol, 1.0 eq.) in water (1 mL) cyanoacetylene (6.26 µL, 100 µmol, 1.0 eq.) was added. The reaction mixture was shaken at 25 °C at 750 rpm for 24 h in an Eppendorf ThermoMixer®. A sample (10 µL) was taken and diluted with water (990 µL) to 1 mL for LCMS analysis (4 µL injection volume). Compound **3** was formed with 62% yield. The yield was determined by LC-MS measurement with the calibration curve prepared using synthetically produced 5-aminoisoxazole.

Synthetic reference:

Hydroxylamine (50 wt% in H_2O, 10.2 g, 9.40 mL, 153 mmol, 1.1 eq.) was added to 3-ethoxyacrylonitrile (mixture of *cis* and *trans,* 13.2 g, 14.0 mL, 136 mmol 1.00 eq.) and stirred for 2 h at 65 °C. The crude product was purified by flash column chromatography (*i*Hex:EF 3:2) to obtain the product **3** (5.92 g, 70.3 mmol, 52%) as slightly yellowish needles. For LC-MS measurements, a fraction of the product was further purified by reversed phase HPLC. The NMR spectrum of the isolated compound was identical to that of a commercial sample.

1-(Isoxazol-5-yl)urea (4)



4

Prebiotic origin:

A solution of 5-Aminoisoxazole **3** (10 µL, 10 µmol, 1 M, 1.00 eq.) was added to a mixture of NaOCN (2.60 mg, 40 µmol, 4.00 eq.) and $ZnSO_4 \cdot 7\,H_2O$ (11.5 mg, 40 µmol, 4.00 eq.). The mixture was shaken at 65 °C and 450 rpm for 14 h in an Eppendorf ThermoMixer® open to the air to allow water to evaporate. The resulting residue was dissolved in water to a final volume of 2 mL. A sample (100 µL) was taken and diluted with water (900 µL) to 1 mL for LCMS analysis (4 µL injection volume). Compound **4** was formed in 49% yield. The yield was determined by LC-MS measurement with the calibration curve prepared using synthetically produced 1-(isoxazol-5-yl)urea **4**.

Prebiotic purification:

A solution of 5-aminoisoxazole **3** (100 µL, 100 µmol, 1 M, 1.00 eq) was added to a mixture of NaOCN (26.0 mg, 400 µmol, 4.00 eq) and $ZnSO_4 \cdot 7H_2O$ (115 mg, 400 µmol, 4.00 eq.). The mixture was shaken at 65 °C and 450 rpm for 14 h in an

Eppendorf ThermoMixer® open to the air to allow water to evaporate. The residue was dissolved in water (10 mL) and Na_2CO_3 (42.4 mg, 400 µmol, 4.0 eq.). The precipitate was filtered off and the filtrate was transferred into a 250 mL beaker and left at RT for 7 days to allow crystallization by concentration. The supernatant was removed and the formed solid residue was dissolved in D_2O. **4** (0.80 mg, 6.30 µmol, 7%) has been identified by comparison of the ^{1}H-NMR spectra with synthetic material and the yield determined by 1,4-dioxane as internal standard (**Figure S1**).

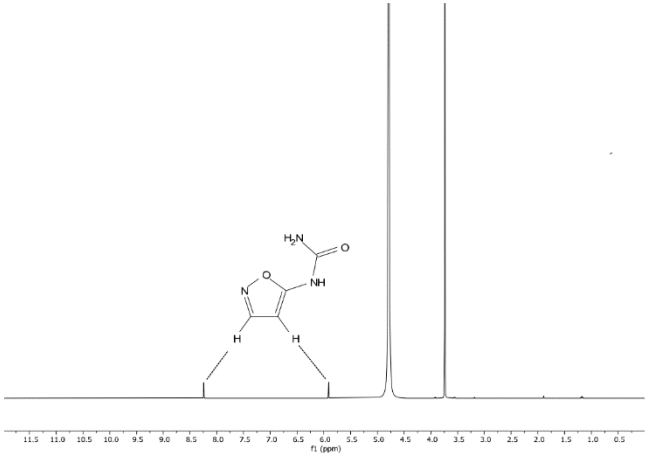


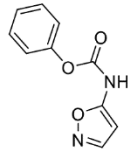
Figure S1. ^{1}H-NMR spectrum (400 MHz, {D_2O}, 0.0–12.0 ppm) of prebiotically enriched 1-(Isoxazol-5-yl)urea **4**. The peak at 3.74 ppm refers to 1,4-dioxane, which has been used as an internal standard.

Synthetic reference:

5-Aminoisoxazole **3** (2.00 g, 23.8 mmol, 1.10 eq.) was dissolved in dry MeCN (28 mL) under inert atmosphere. The reaction mixture was cooled down to 0 °C and trichloroacetyl isocyanate (3.98 g, 2.50 mL, 21.1 mmol, 1.00 eq.) was added dropwise. The ice bath was removed, and the reaction mixture was stirred for 1.5 h at rt. The reaction was quenched with methanol (15 mL) and the solvents were removed *in vacuo*. After co-evaporation with EtOH (2 x 25 mL) the residue was dissolved in methanolic ammonia (28 mL, 2 M) and stirred for 1.5 h at rt. The solvent was removed *in vacuo* and the crude product purified by flash column chromatography (silica, DCM:MeOH 24:1). The product **4** (2.36 g, 18.6 mmol, 88%) was obtained as a white solid.

mp: 165 °C; R_f = 0.30 (*i*Hex:EtOAc = 2:1); **^{1}H NMR** (400 MHz, DMSO-d_6): δ (ppm) = 10.05 (s, 1H, HN), 8.26 (d, J = 1.9 Hz, 1H, 3-H), 6.25 (s, 2H, H_2N), 5.88 (d, J = 1.9 Hz, 1H, 4-H); **^{13}C NMR** (101 MHz, DMSO-d_6): δ (ppm) = 162.7 (C5), 152.9 (C=O), 152.0 (C3), 84.1 (C4); **IR** (cm^{-1}): ṽ = 3464 (w), 3261 (w), 3195 (m), 3021 (w), 1734 (vs), 1699 (m), 1568 (vs), 1560 (vs), 1373 (s), 1326 (s), 1240 (m), 1189 (s), 1122 (m), 1045 (s), 980 (w), 911 (s), 864 (m), 853 (m), 805 (m), 770 (s); **HRMS** (ESI): calcd. for $C_4H_4N_3O_2^-$ [M-H]$^-$:126.0309, found: 126.0309.

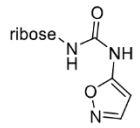
Phenyl isoxazol-5-ylcarbamate



5-Aminoisoxazole **3** (6.00 g, 71.4 mmol, 1.00 eq.) was dissolved in dry MeCN (120 mL) and dry pyridine (5.48 g, 5.60 mL, 69.4 mmol, 0.97 eq.). The reaction mixture was cooled down to 0 °C and phenyl chloroformate (13.6 g, 9.00 mL, 86.8 mmol, 1.22 eq.) was added dropwise. The ice bath was removed, and the reaction mixture stirred for 15 h at rt. Water (120 mL) was added and the reaction mixture was kept at 5 °C overnight. The formed precipitate was filtered off and washed with ice-cold water to give the product (9.63 g, 47.1 mmol, 66 %) as a yellow solid.

mp 195 °C; **^{1}H NMR** (400 MHz, $CDCl_3$): δ (ppm) = 8.18 (dd, J = 1.9 Hz, J = 0.8 Hz, 1H, 3-H), 7.82 (s, 1H, NH), 7.46 – 7.39 (m, 2H, Ar3-H), 7.33 – 7.26 (m, 1H, Ar4-H), 7.22 – 7.18 (m, 2H, Ar2-H), 6.19 (s, 1H, 4-H); **^{13}C NMR** (101 MHz, $CDCl_3$): δ (ppm) = 160.2 (C5), 152.2 (C3), 150.2 (C=O), 149.5 (Ar1), 129.8 (Ar3), 126.6 (Ar4), 121.4 (Ar2), 87.2 (C4); **IR** (cm^{-1}): ṽ = 3203 (vw), 3003 (vw), 1760 (s), 1612 (m), 1560 (s), 1494 (m), 1476 (m), 1340 (w), 1246 (s), 1221 (m), 1190 (vs), 1030 (m), 995 (w), 916 (s), 781 (s), 715 (s), 686 (vs); **HRMS** (ESI): calcd. for $C_{10}H_7N_2O_3^-$ [M-H]$^-$: 203.0462, found: 203.0460.

1-(isoxazol-5-yl)-*N'*-ribosyl-urea (α/β-5, α/β-6)



α/β-**5**, α/β-**6**

Prebiotic origin:

A solution of **4** (500 µL, 0.05 mmol, 100 mM) was mixed with ribose (125 µL, 0.25 mmol, 2 M) and boric acid (50 µL, 0.025 mmol, 500 mM). The mixture was kept in an oven for 14 h at 80 °C. The sample was dissolved in H_2O (1 mL) and a sample (10 µL) was taken and diluted with water (990 µL) to 1 mL for LCMS analysis (4 µL injection volume). To confirm the structural integrity, the different pyrano-isomers were isolated by reversed phase HPLC in pure form.

Prebiotic origin and isomerization:

A solution of **4** (500 µl, 0.05 mmol, 100 mM) is mixed with ribose (83.3 µL, 0.25 mmol, 3 M) and boric acid (50 µL, 0.035 mmol, 500 mM). The mixture was kept in an oven for 14 h at 95 °C. The residue was dissolved in an aqueous borax solution (1.00 mL, 500 mM) and shaken for 3 h at 750 rpm at 75°C in an Eppendorf ThermoMixer®. A sample (10 µL) was taken and diluted with water (990 µL) to 1 mL for LCMS analysis (4 µL injection volume) (**Figure S1**).

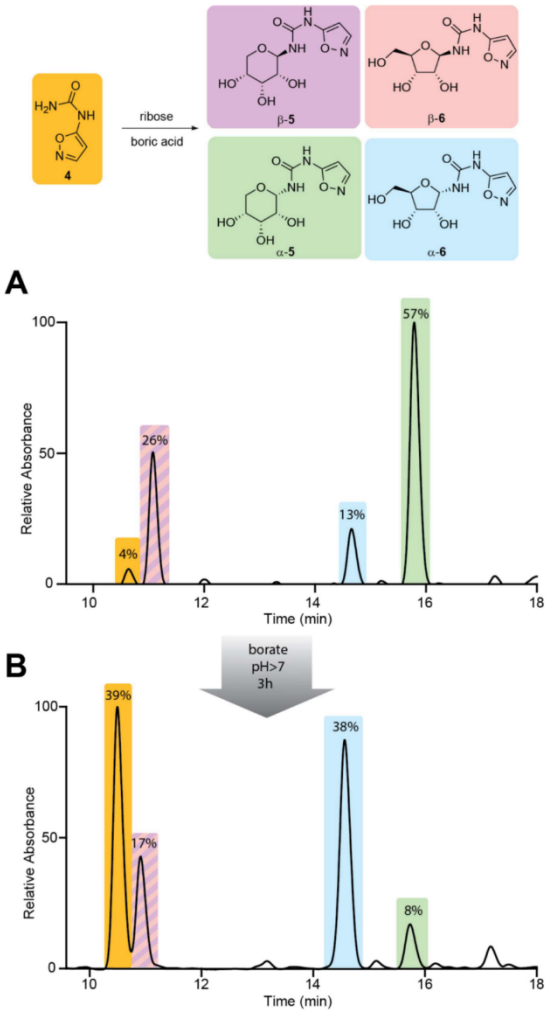


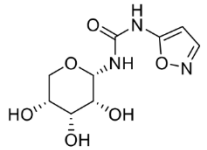
Figure S2. LC-MS analysis for the formation of α/β-**5** and α/β-**6**. The UV-chromatograms at 243 nm are shown. The β-anomers could not be distinguished. A) Boric acid catalyzed ribosylation of **4** affords the four expected α and β anomers of isoxazolyl-urea ribosides **5** and **6**. B) The mixture obtained in **A** was heated in 500 mM borax at 75 °C for 3 h. The pyranoside products (α/β-**5**) are converted into the furanoside products (α/β-**6**), which is seen for the α anomers and expected to happen for the β anomers. The second reaction can be performed at different conditions to result different ratios for α/ β and furanoside/pyranoside products.

Synthetic reference:

2,3,5-Tri-*O*-acetyl-β-D-ribofuranosyl azide[2] (7.07 g, 14.5 mmol, 1.00 eq.) was dissolved in THF (40 mL) and Palladium on carbon (10%, 25 mg, 23.5 µmol, 0.002 eq.) was added. The mixture was stirred for 19 h at rt under an H_2-atmosphere and

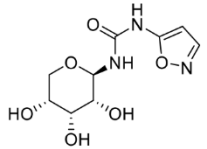
subsequently filtered over celite and the solvent was removed *in vacuo*. The residue was dissolved in dry MeCN (120 mL) together with phenyl isoxazol-5-ylcarbamate (3.26 g, 16.0 mmol, 1.19 eq.). The mixture was stirred for 72 h at rt and for an additional 48 h at 45 °C. The solvent was removed *in vacuo* and the residue was purified by flash column chromatography (silica. DCM:MeOH 99:1 → 49:1). The product was dissolved in methanolic ammonia (84 mL, 2 M) and stirred for 14 h at rt. The solvent was removed *in vacuo* and the residue was purified by flash column chromatography (silica, DCM:MeOH 9:1). Not completely deprotected fractions were collected and stirred in methanolic ammonia (42 mL. 2 M) for 16 h, The solvent was removed *in vacuo* and the residue was again purified by flash column chromatography (silica, DCM:MeOH 9:1). The products were combined to give **6** (645 mg, 2.49 mmol, 17%) as a yellowish foam in a 2.25:1 mixture of α- and β-furanosides. The isomers were separated by reversed phase HPLC to confirm the structural integrity of the different furano-compounds.

α-pyranosyl-isomer (α-5)



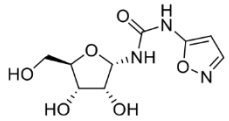
^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm) = 8.40 (s, 1H, NH), 8.33 (d, *J* = 1.9 Hz, 1H, 3-H), 7.48 (d, ^3J = 9.0 Hz, 1H, NH), 5.95 (d, *J* = 1.9 Hz, 1H, 4-H), 5.02 (dd, *J* = 8.9 Hz, *J* = 3.3 Hz, 1H, 1'-H), 3.74 (t, *J* = 2.7 Hz, 1H, 3'-H), 3.60 – 3.52 (m, 3H, 2'-H, 4'-H, 5'-H), 3.35 – 3.31 (m, 1H, 5''-H). **^{13}C NMR** (101 MHz, DMSO-d_6): δ (ppm) = 162.5 (C5), 152.0 (C3), 151.8 (C=O), 84.4 (C4), 77.7 (C1'), 69.8 (C3'), 68.8 (C4'), 67.4 (C2'), 61.6 (C5').

β-pyranosyl-isomer (β-5)



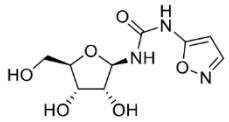
^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm) = 8.41 (s, 1H, NH), 8.32 (d, *J* = 1.9 Hz, 1H, 3-H), 7.67 (d, *J* = 8.9 Hz, 1H, NH), 5.95 (d, *J* = 1.8 Hz, 1H, 4-H), 4.89 (t, *J* = 8.9 Hz, 1H, 1'-H), 3.88 (q, *J* = 2.7 Hz, 1H, 3'-H), 3.49 (ddt, *J* = 11.8 Hz, *J* = 6.5 Hz, *J* = 2.3 Hz, 1H, 4'-H), 3.46 – 3.34 (m, 2H, 5'-H, 5''-H), 3.23 (dd, *J* = 8.9 Hz, *J* = 2.7 Hz, 1H, 2'-H). **^{13}C NMR** (101 MHz, DMSO-d_6): δ (ppm) = 163.1 (C5), 152.8 (CO), 152.4 (C3), 84.8 (C4), 78.2 (C1'), 71.2 (C3'), 70.1 (C2'), 67.5 (C4'), 64.6 (C5').

α-furanosyl-isomer (α-6)



^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm) = 10.54 (s, 1H, NH), 8.33 (d, *J* = 2.0 Hz, 1H, 3-H), 7.07 (d, *J* = 9.5 Hz, 1H, NH), 5.95 (d, *J* = 1.9 Hz, 1H, 4-H), 5.50 – 5.44 (m, 2H, 1'-H, 2'-OH), 5.10 (d, *J* = 5.8 Hz, 1H, 3'-OH), 4.70 (t, *J* = 5.7 Hz, 1H, 5'-OH), 3.94 (q, *J* = 4.7 Hz, 1H, 2'-H), 3.88 (td, *J* = 6.0, 4.6 Hz, 1H, 3'-H), 3.72 (ddd, *J* = 6.2 Hz, *J* = 4.8 Hz, *J* = 3.2 Hz, 1H, 4'-H), 3.48 (ddd, *J* = 11.8 Hz, *J* = 5.5 Hz, *J* = 3.3 Hz, 1H, 5'-H), 3.40 – 3.29 (m, 1H, 5''-H); **^{13}C NMR** (101 MHz, DMSO-d_6): δ (ppm) = 162.3 (C5), 152.1 (C3), 151.5 (C=O), 84.4 (C4), 82.3 (C4'), 80.5 (C1'), 71.1 (C3'), 70.1 (C2'), 61.6 (C5').

β-furanosyl-isomer (β-6)



^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm) = 8.42 (s, 1H, NH), 8.31 (d, *J* = 1.9 Hz, 1H, 3-H), 8.16 (d, ^{3}*J* = 9.3 Hz, 1H, NH), 5.95 (d, *J* = 1.9 Hz, 1H, 4-H), 5.20 (dd, *J* = 9.3 Hz, *J* = 5.1 Hz, 1H, 1'-H), 3.87 (t, *J* = 4.7 Hz, 1H, 3'-H), 3.77 (t, *J* = 5.1 Hz, 1'-H), 3.66 (q, *J* = 4.3 Hz, 1H, 4'-H), 3.46 (dd, *J* = 11.7 Hz, *J* = 4.0 Hz, 1H, 5'-H), 3.38 (dd, *J* = 11.7, 4.7 Hz, 1H. 5''-H). **^{13}C NMR** (101 MHz, DMSO-d_6): δ (ppm) = 162.8 (C5), 152.3 (C=O), 151.9 (C3), 84.8 (C1'), 84.4 (C4), 83.5 (C4'), 74.1 (C2'), 70.5 (C3'), 62.1 (C5').

Prebiotic formation of uridine (α/β-7, α/β-8)

From 1-(isoxazol-5-yl)-N'-ribosyl-urea α/β-**6**:

The second reaction step was carried out under inert atmosphere. All solutions were degassed for 1 h with argon before usage. A solution of ribose isoxazole α/β-**6** as a 2.25:1 mixture (25.9 mg, 0.10 mmol, 1.00 eq.) in water (0.5 mL) was added to a mixture of sodium carbonate (5.30 mg, 0.05 mmol, 0.50 eq.), DTT (23.1 mg, 1.50 mmol, 1.50 eq.) and FeS_2 (6.00 mg, 0.50 mmol, 0.50 eq.) in a 15 mL falcon tube. The tube was sealed with a PTFE sealing tape and shaken for 4 h at 750 rpm at 100°C in an Eppendorf ThermoMixer®. After cooling to rt and subsequent centrifugation a sample (10 µL) was taken and diluted with H_2O (990 µL) to 1 mL. This diluted sample was used for LC-MS (4 µL) analysis according to the information given in the general experimental methods.

One-Pot from **4**:

The reaction was carried out inert atmosphere. All solutions were degassed for 1 h with argon before usage. A solution of **4** (500 µl, 0.05 mmol, 100 mM) is mixed with ribose (83.0 µL, 0.25 mmol, 3 M) and boric acid (25.0 µl, 13.0 µmol, 500 mM). The mixture was kept in an oven for 14 h at 95 °C. The residue was dissolved in 75 mM Na_2CO_3 (0.5 mL, pH 10) and added to DTT (11.6 mg, 75.0 µmol, 1.5 eq). $(NH_4)_2Fe(SO_4)_2$ solution (10 µL, 75.0 nmol, 0.001 eq, 7.5 mM) was added and the mixture sealed with a PTFE sealing tape and shaken for 4 h at 750 rpm at 100°C in an Eppendorf ThermoMixer®. After cooling to rt and subsequent centrifugation a sample (10 µL) was removed and diluted with H_2O (990 µL) to 1 mL. This diluted sample was used for LC-MS analysis according to the general information (4 µL injection volume) (**Figure S3**).

SUPPORTING INFORMATION WILEY-VCH

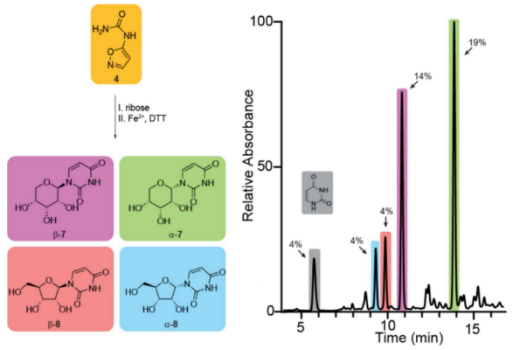
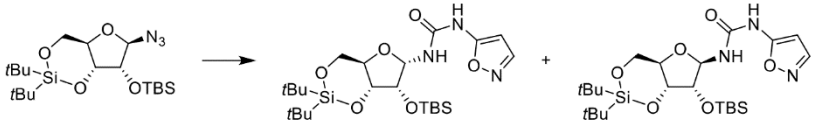


Figure S3. LC-MS analysis for the formation of **α/β**-7, **α/β**-8 and uracil in a one-pot reaction from **4**.

Synthesis and Characterization of the Phosphoramidite Building Blocks

N-isoxazol-5-yl-*N'*-(1'-(2'-*O*-*tert*-butyldimethylsilyl-3',5'-O-(di-*tert*-butylsilandiyl)-β-D-ribofuranosyl))-urea (10)



9 α-**10** β-**10**

9 (1.50 g, 3.49 mmol, 1.00 eq.) and 5-aminoisoxazole (646 mg, 7.68 mmol, 2.20 eq.) were dissolved in dry toluene (35 mL). Palladium on carbon (10%, 185 mg, 175 µmol, 0.05 eq.) and XPhos (166 mg, 349 µmol, 0.1 eq.) were added to the reaction mixture. The reaction mixture was stirred at 60 °C under CO-atmosphere (5 bar) for 48 h. The catalyst was filtered off over a pad of $MgSO_4$ and washed with toluene. The filtrate was concentrated under reduced pressure and the crude product was purified by flash column chromatography (silica, *i*Hex:EtOAc = 5:1) to yield **α-10** (563 mg, 1.10 mmol, 31%) and **β-10** (254 mg, 494 µmol, 14%) as colorless solids

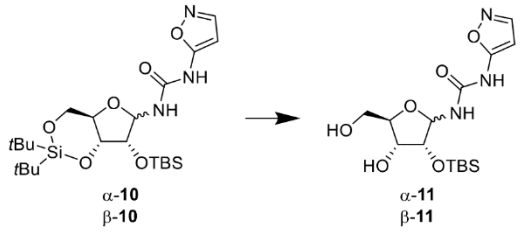
α-10:

***R*f** = 0.52 (*i*Hex:EtOAc = 5:1); **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.65 (s, 1H, NH), 8.22 (d, *J* = 1.9 Hz, 1H, 3-H), 6.73 (d, *J* = 10.1 Hz, 1H, NH), 6.10 (d, *J* = 1.9 Hz, 1H, 4-H), 5.98 – 5.96 (m, 1H, 1'-H), 4.57 (t, *J* = 4.5 Hz, 1H, 2'-H), 4.30 (dd, *J* = 9.3 Hz, *J* = 5.2 Hz, 1H, 5'-H), 4.06 (dd, *J* = 9.6, *J* = 4.6 Hz, 1H, 3'-H), 3.92 (td, *J* = 10.0 Hz, *J* = 5.2 Hz, 1H, 4'-H), 3.80 (dd, *J* = 10.3 Hz, *J* = 9.3 Hz, 1H, 5''-H), 1.07 (s, 9H, Si-C-CH_3), 1.04 (s, 9H, Si-C-CH_3), 0.97 (s, 9H, Si-C-CH_3), 0.23 (s, 3H, Si-CH_3), 0.20 (s, 3H, Si-CH_3); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 163.0 (C5), 152.5 (C3), 85.9 (C4), 82.9 (C1'), 78.5 (C3'), 73.2 (C4'), 72.1 (C2'), 68.9 (C5'), 27.9 (Si-C-CH_3), 27.5 (Si-C-CH_3), 26.6 (Si-C-CH_3), 23.18 (Si-C), 20.91 (Si-C), 19.49 (Si-C), -3.96 (Si-CH_3), -4.69 (Si-CH_3) ; **IR** (cm^{-1}): $\tilde{\nu}$ = 3341 (w), 2953 (m), 2930 (m), 2857 (w), 2361 (w), 1708 (m), 1606 (m), 1532 (m), 1523 (s), 1484 (m), 1472 (m), 1446, (m), 1405 (w), 1363 (m), 1301 (m), 1250 (s), 1222 (m), 1176 (s), 1153 (m), 1100 (m), 1032 (s), 994 (m), 911, (m), 835 (s), 780 (m), 702 (m); **HRMS** (ESI): calcd. for $C_{23}H_{44}N_3O_6Si_2^+$ $[M+H]^+$: 514.27632, found: 514.27688.

β-10:

***R*f** = 0.36 (*i*Hex:EtOAc = 5:1); **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.34 (s, 1H, NH), 8.21 (d, *J* = 1.9 Hz, 1H, 3-H), 7.01 (d, *J* = 12.7 Hz, 1H, NH), 6.08 (d, *J* = 1.9 Hz, 1H, 4-H), 5.47 (d, *J* = 9.2 Hz, 1H, 1'-H), 4.35 (dd, *J* = 9.2 Hz, *J* = 5.1 Hz, 1H, 5'-H), 4.31 (d, *J* = 4.7 Hz, 1H, 2'-H), 4.04 (dd, *J* = 9.5 Hz, *J* = 4.7 Hz, 1H, 3'-H), 3.95 (ddd, *J* = 10.5 Hz, *J* = 9.5 Hz, *J* = 5.1 Hz, 1H, 4'-H), 3.83 (dd, *J* = 10.5, 1.4 Hz, 1H, 5''-H), 1.06 (s, 9H, Si-C-CH_3), 1.04 (s, 9H, Si-C-CH_3), 0.95 (s, 9H, Si-C-CH_3), 0.18 (s, 3H, Si-CH_3), 0.16 (s, 3H, Si-CH_3); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 163.0 (C5), 152.4 (C3), 90.1 (C1'), 86.0 (C4), 77.6 (C3'), 77.0 (C2'), 74.1 (C4'), 69.2 (C5'), 27.9 (Si-C-CH_3), 27.5 (Si-C-CH_3), 26.3 (Si-C-CH_3), 23.2 (Si-C), 20.9 (Si-C), 19.0 (Si-C), -4.16 (Si-CH_3), -4.80 (Si-CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 3306 (w), 2952 (m), 2929 (m), 2857 (w), 2361 (w), 1708 (m), 1605 (m), 1531 (m), 1507 (s), 1483 (m), 1464 (m), 1445, (m), 1405 (w), 1363 (m), 1300 (m), 1249 (s), 1222 (m), 1175 (s), 1153 (m), 1099 (m), 1030 (s), 992 (m), 909, (m), 833 (s), 780 (m), 701 (m); **HRMS** (ESI): calcd. for $C_{23}H_{44}N_3O_6Si_2^+$ $[M+H]^+$: 514.27632, found: 514.27654.

N-isoxazol-5-yl-*N'*-(1'-(2'-*O*-*tert*-butyldimethylsilyl)-β-D-ribofuranosyl)-urea (11)



α-**10** α-**11**
β-**10** β-**11**

α-10 or **β-10** (1.00 eq.) was dissolved in a mixture of DCM and pyridine (10/1). HF (70% in pyridine, 5.00 eq.) was added dropwise at 0 °C. The resulting mixture was stirred for 2 h and afterwards quenched by the addition of methoxytrimethylsilane. The solution was diluted with DCM and washed with sat. $NaHCO_3$. The aqueous phase was extracted with DCM (3x). The combined organic layers were washed with brine, dried over Na_2SO_4, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (silica, DCM:MeOH = 100:5) to yield **11** as a colorless foam.

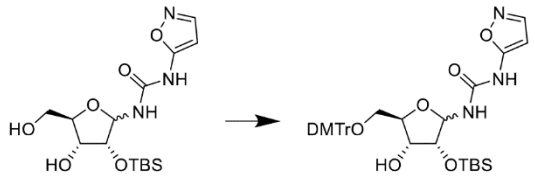
α-11:

Yield: 83%; ***R*f** = 0.27 (DCM:MeOH = 100:5); **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.79 (s, 1H, NH), 8.20 (dd, *J* = 7.8, 1.9 Hz, 1H, 3-H), 6.74 (d, *J* = 9.6 Hz, 1H, NH), 6.08 (d, *J* = 1.9 Hz, 1H, 2-H), 5.69 (dd, *J* = 9.4, 5.8 Hz, 1H, 1'-H), 4.39 (dd, *J* = 5.7 Hz, *J* = 4.9 Hz, 1H, 2'-H), 4.14 – 4.11 (m, 1H, 3'-H), 3.98 (d, *J* = 3.1 Hz, 1H, 3'-OH), 3.95 (td, *J* = 3.9 Hz, *J* = 2.5 Hz, 1H, 4'-H), 3.82 (dd, *J* = 6.8 Hz, *J* = 5.4 Hz, 1H, 5'-OH), 3.59 – 3.51 (m, 1H, 5''-H), 0.91 (s, 9H, Si-C-CH_3), 0.15 (s, 3H, Si-CH_3), 0.12 (s, 3H, Si-CH_3); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 163.4 (C5), 152.4 (C3), 85.5 (C4), 85.0 (C4'), 82.1 (C1'), 73.2 (C3'), 72.6 (C2'), 63.3 (C5'), 26.2 (Si-C-CH_3), 18.80 (Si-C), -4.70 (Si-CH_3), -4.84 (Si-CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 2850 (m), 1612 (w), 1473 (s), 1362 (m), 1338 (w), 1255 (m), 1193 (m), 1145 (m), 1092 (w), 1003 (m), 911 (m), 836 (s), 777 (s); **HRMS** (ESI): calcd. for $C_{15}H_{27}N_3NaO_6Si^+$ $[M+Na]^+$: 396.15613, found: 374.15597.

β-11:

Yield: 77%; ***R*f** = 0.25 (DCM:MeOH = 100:5); **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.54 (s, 1H, NH), 8.21 (d, *J* = 1.9 Hz, 1H, 3-H), 6.67 (d, *J* = 9.0 Hz, 1H, NH), 6.09 (d, *J* = 1.9 Hz, 1H, 4-H), 5.42 (dd, *J* = 9.1 Hz, *J* = 4.4 Hz, 1H, 1'-H), 4.25 (t, *J* = 4.7 Hz, 1H, 2'-H), 4.11 (ddd, *J* = 6.9 Hz, *J* = 5.0 Hz, *J* = 3.2 Hz, 1H, 3'-H), 4.02 (s, 1H, 5'-OH), 3.87 (dt, *J* = 4.3 Hz, *J* = 3.0 Hz, 1H, 4'-H), 3.66 (dt, J *J* = 12.1 Hz, *J* = 3.5 Hz, 1H, 5'-H), 3.62 – 3.56 (m, 1H, 5''-H), 3.47 (d, *J* = 5.5 Hz, 1H, 3'-OH), 0.93 (s, 9H, Si-C-CH_3), 0.16 (s, 3H, Si-CH_3), 0.16 (s, 3H, Si-CH_3); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 163.1 (C5), 152.4 (C3), 87.0 (1C'), 85.8 (C4), 84.9 (C4'), 77.4 (C2'), 72.1 (C3'), 62.6 (5'), 26.2 (Si-C-CH_3), 18.8 (Si-C), -4.5 (Si-C-CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 2930 (m), 2858 (m), 1700 (m), 1611 (w), 1549 (s), 1473 (s), 1362 (w), 1339 (w), 1255 (m), 1194 (m), 1039 (m), 1005 (m), 911 (m), 836 (s), 779 (s); **HRMS** (ESI): calcd. for $C_{15}H_{27}N_3NaO_6Si^+$ $[M+Na]^+$: 396.15613, found: 374.15601.

_N_-isoxazol-5-yl-_N'_-(1'-(5'-_O_-(4,4'-dimethoxytrityl)-2'-_O-tert_-butyldimethylsilyl)-β-D-ribofuranosyl)-urea (12)



α-**11**
β-**11**

α-**12**
β-**12**

α-11 or **β-11** (1.00 eq.) was dissolved in pyridine and DMT-Cl (1.40 eq.) was added. The resulting mixture was stirred at room temperature for 13 h. The Reaction was quenched by the addition of a few drops of methanol. Afterwards the solvent was evaporated *in vacuo* and the crude product was purified by column chromatography (silica, *i*Hex:EE = 2.5:1 + 0.1% pyridine) to yield **12**as a slightly yellow foam.

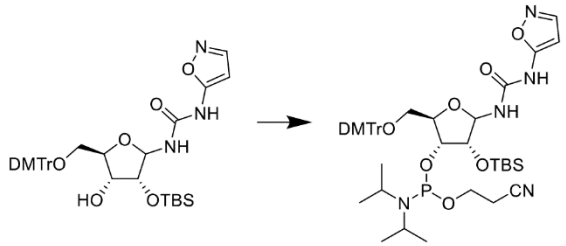
α-12:

Yield: 87%; R_f = 0.25 (*i*Hex:EE = 2.5:1 + 0.1% pyridine); **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.81 (s, 1H, NH), 8.21 (d, *J* = 1.9 Hz, 1H, 3-H), 7.49 – 7.46 (m, 2H, Ar-H), 7.36 – 7.30 (m, 6H, Ar-H), 7.23 (ddt, *J* = 7.7 Hz, *J* = 6.9 Hz, *J* = 1.2 Hz, 1H, Ar-H), 6.91 – 6.88 (m, 4H, Ar-H), 6.79 (d, *J* = 9.5 Hz, 1H, NH), 6.11 (d, *J* = 1.9 Hz, 1H, 4-H), 5.87 – 5.84 (m, 1H, 1'-H), 4.57 (dd, *J* = 5.8 Hz, *J* = 4.8 Hz, 1H, 2'-H), 4.08 (dtt, *J* = 5.8 Hz, *J* = 3.6 Hz, *J* = 2.2 Hz, 2H, 3'-H, 4'-H), 4.04 (t, *J* = 3.1 Hz, 1H, 3'-OH), 3.79 (s, 6H, O-CH_3), 3.28 (dd, *J* = 10.2 Hz, *J* = 3.7 Hz, 1H, 5'-H), 3.05 (dd, *J* = 10.2 Hz, *J* = 3.5 Hz, 1H, 5''-H), 0.92 (s, 9H, Si-C-CH_3), 0.17 (s, 3H, Si-CH_3), 0.14 (s, 3H, Si-CH_3); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 159.6 (C5), 152.4 (C3), 146.2 (ArC), 136.9 (ArC), 136.7 (ArC), 131.0 (ArC), 130.9 (ArC), 128.9 (ArC), 128.6 (ArC), 127.6 (ArC), 113.9 (ArC), 113.9 (ArC), 86.9 (Ar-C_3), 85.5 (C4), 83.5 (C3'), 82.3 (C1'), 73.7 (C4'), 72.6 (C2'), 64.9 (C5'), 55.5 (O-CH_3), 26.2 (Si-C-CH_3), 18.8 (Si-C), -4.6 (Si-CH_3), -4.8 (Si-CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 2930 (m), 2858 (w), 2360 (m), 1608 (m), 1558 (s), 1508 (s), 1477 (s), 1446 (w), 1362 (w) 1302 (w), 1251 (s), 1177 (s), 1101(w), 1079 (w), 1033 (s), 1000 (m), 911 (m), 835 (s), 779 (s); **HRMS** (ESI): calcd. for $C_{36}H_{44}N_3O_8Si^-$ [M-H]$^-$: 674.29032, found: 674.29083.

β-12:

Yield: 78%; R_f = 0.21 (*i*Hex:EE = 2.5:1 + 0.1% pyridine); **^{1}H NMR** (800 MHz, Acetone-d_6): δ (ppm) = 9.42 (s, 1H, NH), 8.22 (d, *J* = 1.9 Hz, 1H, 3'-H), 7.50 – 7.48 (m, 2H, Ar-H), 7.37 – 7.34 (m, 4H, Ar-H), 7.31 – 7.28 (m, 2H, Ar-H), 7.23 – 7.20 (m, 1H, Ar-H), 6.89 – 6.86 (m, 4H, Ar-H), 6.73 (d, *J* = 9.4 Hz, 1H, NH), 6.12 (d, *J* = 1.9 Hz, 1H, 4-H), 5.54 (dd, *J* = 9.4 Hz, *J* = 5.2 Hz, 1H, 1'-H), 4.25 (td, *J* = 5.1 Hz, *J* = 1.3 Hz, 1H, 2'-H), 4.10 – 4.06 (m, 1H, 3'-H), 4.00 (q, *J* = 4.1 Hz, 1H, 4'-H), 3.77 (d, *J* = 2.6 Hz, 6H, O-CH_3), 3.57 (d, *J* = 5.3 Hz, 1H, 3'-OH), 3.28 (dt, *J* = 10.1 Hz, *J* = 3.2 Hz, 1H, 5'-H), 3.12 (dd, *J* = 10.1 Hz, *J* = 4.5 Hz, 1H, 5''-H), 0.94 (s, 9H, Si-C-CH_3), 0.18 (s, 3H, Si-CH_3), 0.18 (s, 3H, Si-CH_3); **^{13}C NMR** (201 MHz, Acetone-d_6): δ (ppm) = 159.6 (C5), 152.5 (C3), 146.2 (ArC), 136.9 (ArC), 136.8 (ArC), 131.0 (ArC), 131.0 (ArC), 129.1 (ArC), 128.6 (ArC), 127.5 (ArC), 113.9 (ArC), 86.8 (C1'), 86.2 (Ar-C_3), 85.9 (C4), 83.1 (C4'), 77.1 (C2'), 72.5 (C3'), 65.5 (C5'), 55.48 (O-CH_3), 26.20 (Si-C-CH_3), 18.81 (Si-C), -4.45 (Si-CH_3), -4.51 (Si-CH_3); **IR** (cm^{-1}): $\tilde{\nu}$ = 2930 (w), 2858 (w), 1608 (m), 1541 (s), 1508 (s), 1446 (w), 1362 (w) 1301 (w), 1250 (s), 1175 (s), 1031(s), 910 (m), 834 (s), 779 (s); **HRMS** (ESI): calcd. for $C_{36}H_{44}N_3O_8Si^-$ [M-H]$^-$: 674.29032, found: 674.29077.

_N_-isoxazol-5-yl-_N'_-(1'-(5'-_O_-(4,4'-dimethoxytrityl)-3'-_O_-[2-cyanoethoxy(diisopropylamino)phosphino]-2'_O-tert_-nutyldimethylsilyl)-β-D-ribofuranosyl)-urea (13)



α-**12**
β-**12**

α-**13**
β-**13**

α-12 or **β-12** (1.00 eq.) was pre-dried by lyophilization from benzene (3x) and dissolved in DCM at 0 °C under argon atmosphere. DIPEA (4.00 eq.) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (2.50 eq.) were added dropwise. The resulting solution was stirred at room temperature for 4 h and subsequently quenched by the addition of sat. $NaHCO_3$ solution. The aqueous layer was extracted with DCM (3x). The combined organic layers were washed with brine, dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by flash column chromatography (silica, *n*Hex:EE = 2:1 + 0.1% pyridine, HPLC grade solvents) to yield **13** as mixture of diastereoisomers, as a colorless foam.

α-13:

Yield: 85%; R_f = 0.36 (*n*Hex:EE = 2:1 + 0.1% pyridine); **^{31}P NMR** (202 MHz, Acetone-d_6): δ (ppm) = 150.26, 148.05. **HRMS** (ESI): calcd. for $C_{45}H_{63}N_5O_9PSi^+$ [M+H]$^+$: 876.41272, found: 876.41382.

β-13:

Yield: 73%; R_f = 0.34 (*n*Hex:EE = 2:1 + 0.1% pyridine); **^{31}P NMR** (202 MHz, Acetone-d_6): δ (ppm) = 150.01, 148.23. **HRMS** (ESI): calcd. for $C_{45}H_{62}N_5NaO_9PSi^+$ [M+Na]$^+$: 898.39521, found: 898.39525.

Synthesis and Purification of Oligonucleotides

Phosphoramidites of canonical ribonucleosides (Bz-A-CE, iBu-G-CE, Ac-C-CE and U-CE) were purchased from Sigma-Aldrich. Oligonucleotides were synthesized on a 1 µmol scale using high Load Glen UnySupport™ as solid supports using an RNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with a standard phosphoramidite chemistry. Oligonucleotides were synthesized in DMT-OFF mode using DCA as a deblocking agent in CH_2Cl_2, Activator 42® or ETT as activator in MeCN, Ac_2O as capping reagent in pyridine/THF and I_2 as oxidizer in pyridine/H_2O. The cleavage and deprotection of the CPG bound oligonucleotides were performed with a 1:1 aqueous solution mixture (0.6 mL) of 30% NH_4OH and 40% $MeNH_2$. The suspension was heated at 65 °C for 1 h. Subsequently, the supernatant was collected, and the beads were washed with water (3×0.3 mL). The combined aqueous solutions were concentrated under reduced pressure using a SpeedVac concentrator and subsequently lyophilized. After that, the crude was dissolved in DMSO (100 µL) and triethylamine trihydrofluoride (125 µL) was added. The solution was heated at 65 °C for 1.5 h. Finally, the oligonucleotides were precipitated by adding 3 M NaOAc in water (25 µL) and *n*-BuOH (1 mL). The mixture was kept at -80 °C for 2 h and centrifuged at 4 °C for 1 h. The supernatant was removed, and the white precipitate was lyophilized. The oligonucleotides were further purified by semi-preparative reverse-phase HPLC using a 1260 Infinity II Manual Preparative LC System from Agilent (G7114A detector) equipped with the column VP 250/10 Nucleodur 100-5 C18ec from Macherey Nagel. A flow rate of 5 mL/min with varying gradients between 0-15% and 0-40% of buffer B in 45 min was applied for the purifications. The following buffer system was used: buffer A: 100 mM NEt_3/HOAc (pH 7.0) in H_2O and buffer B: 100 mM NEt_3/HOAc in 80% (v/v) acetonitrile. The purified oligonucleotides were analyzed by analytical RP-HPLC on a 1260 Infinity II LC System from Agilent (G7165A detector) equipped with the column an EC 250/4 Nucleodur 100-3 C18ec from Macherey Nagel using a flow of 1 mL/min, a gradient of 0-15% or 0-40% of buffer B in 45 min was applied. Finally, the purified oligonucleotides were desalted using a C18 RP-cartridge from Waters. The absorbance of the synthesized oligonucleotides in H_2O solution were measured using an IMPLEN NanoPhotometer® N60/N50 at 260 nm. The extinction coefficients of the oligonucleotides were calculated using the OligoAnalyzer Version 3.0 from Integrated DNA Technologies. The structural integrity of the synthesized oligonucleotides was analyzed by MALDI-TOF mass measurement. For this purpose, the synthesized oligonucleotides (2-3 µL) were desalted on a 0.025 µm VSWP filter (Millipore), co-crystallized in a 3-hydroxypicolinic acid matrix (HPA, 1 µL) and measured on a Bruker Autoflex II.

HPL-chromatograms and MALDI Spectra of Synthesized Oligonucleotides

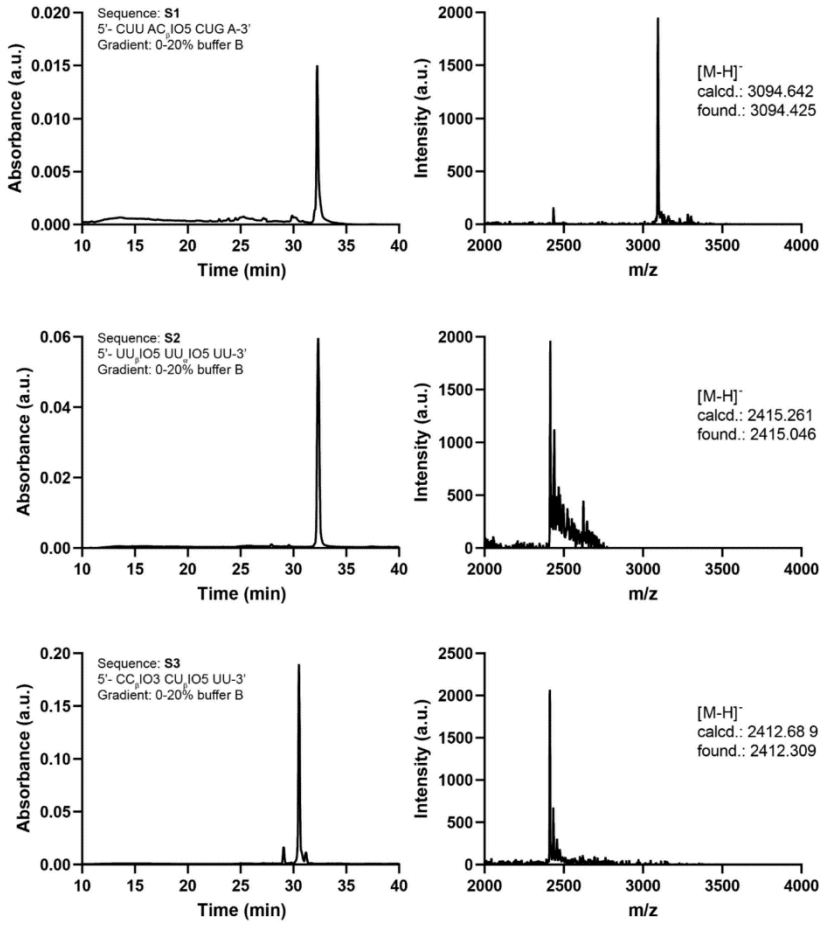


Figure S4. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the purified oligonucleotides.

In-strand Pyrimidine Formation Reactions

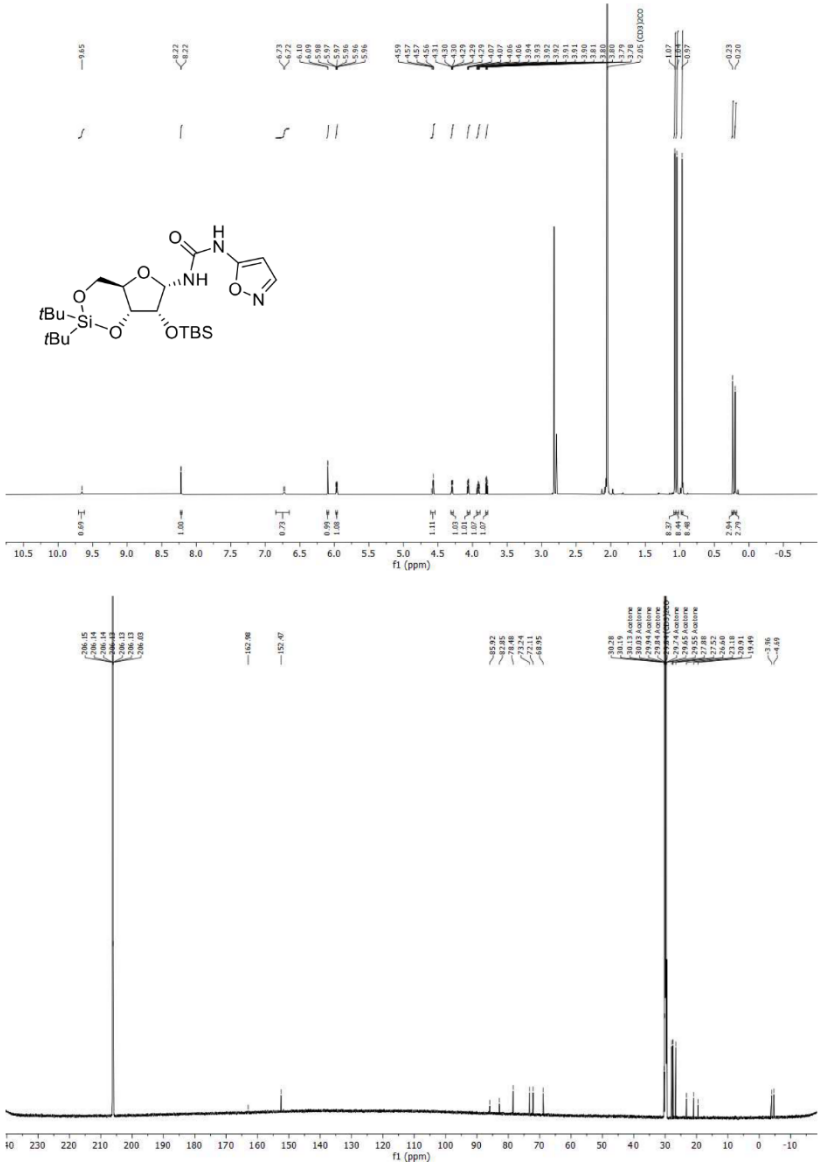
Stock solutions of boric acid (pH 9.7, 100 mM) and Na_2CO_3 (1000 mM) were prepared in water. The oligonucleotides were with mixed with buffer, Na_2CO_3, and water. The Fe^{2+} source (FeS or FeS_2) and DTT were added to the mixture. The final concentration of the components: 100 uM of Oligo, 50 mM Buffer, 100 mM Na_2CO_3, 100 mM Fe^{2+}, 300 mM DTT. The mixture was heated at 90 °C for 2 h in a *TAdvanced Thermocycler* by *Biometra*. After cooling to room temperature, the solids were removed by centrifugation and washed with water (2×0.2 mL), residual solids were removed using a syringe filter (0.20 um, PTFE-membrane). The reaction mixture was concentrated by lyophilization and subsequently analyzed by reverse-phase HPLC. The yields of the reactions were calculated by integration of the chromatographic peaks of the products and the use of the calibration curves of the synthetically prepared product.

Digestion and LC-HESI-MS analysis

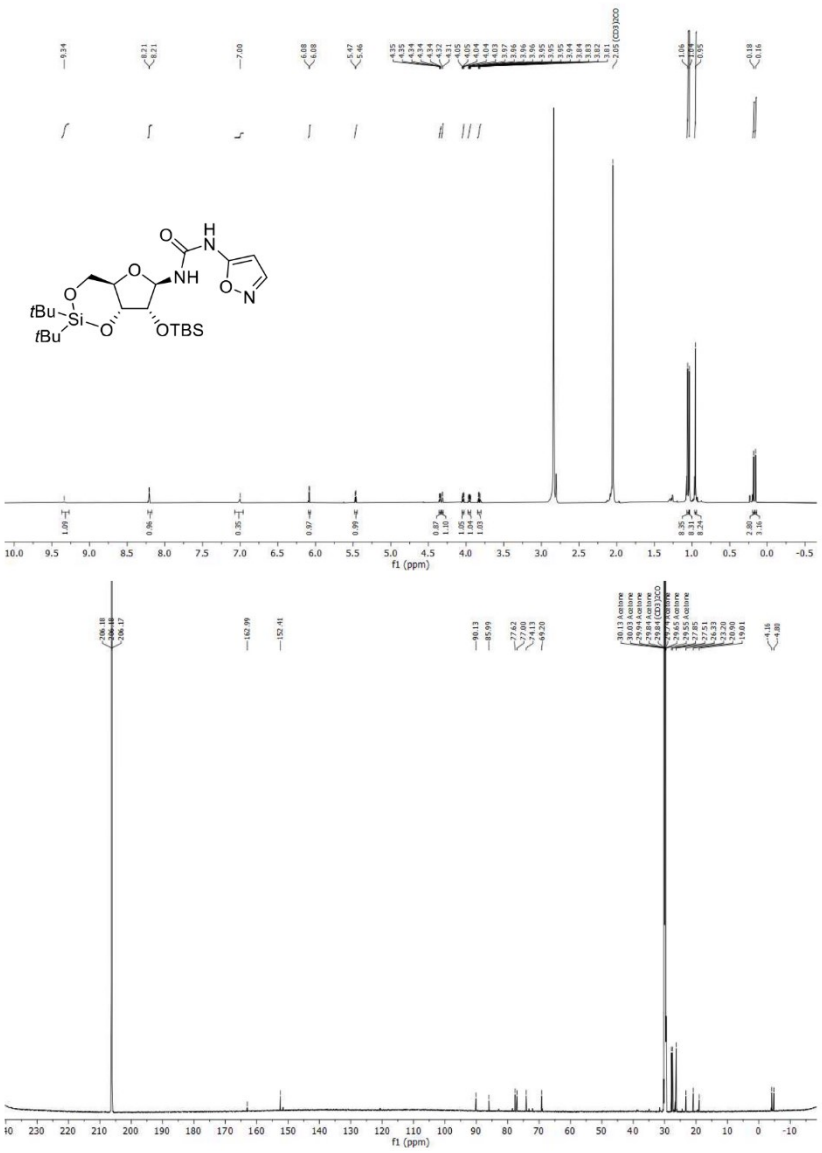
Reaction buffer 10X and *Enzyme mix* was bought as a Nucleoside Digestion Mix (M0649S) kit (*New England BioLabs Inc.*). The purified oligonucleotide (250-500 ng in 46 uL) was incubated with *Reaction buffer 10X* (5 uL) and *Enzyme mix* (1 uL) at 37°C for 1.5 h. The mixture was subsequently diluted to 90 uL and then analyzed by Thermo Scientific Vanquish system coupled to a Thermo Scientific QExactive HF mass spectrometer. All chromatographic separations were performed on an Interchim Uptisphere UP3HDO C18 column with a flow of 0.2 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 H_2O/MeCN 20/80 (pH 5.5)). The elution was monitored at 223, 243 and 260 nm (Thermo Scientific Vanquish Diode Array Detector). The chromatographic eluent was directly injected into the ion source without prior splitting. The synthetic standards for the co-injection experiments were synthesized in our lab according to reported literature[1] or purchased.

NMR Spectra

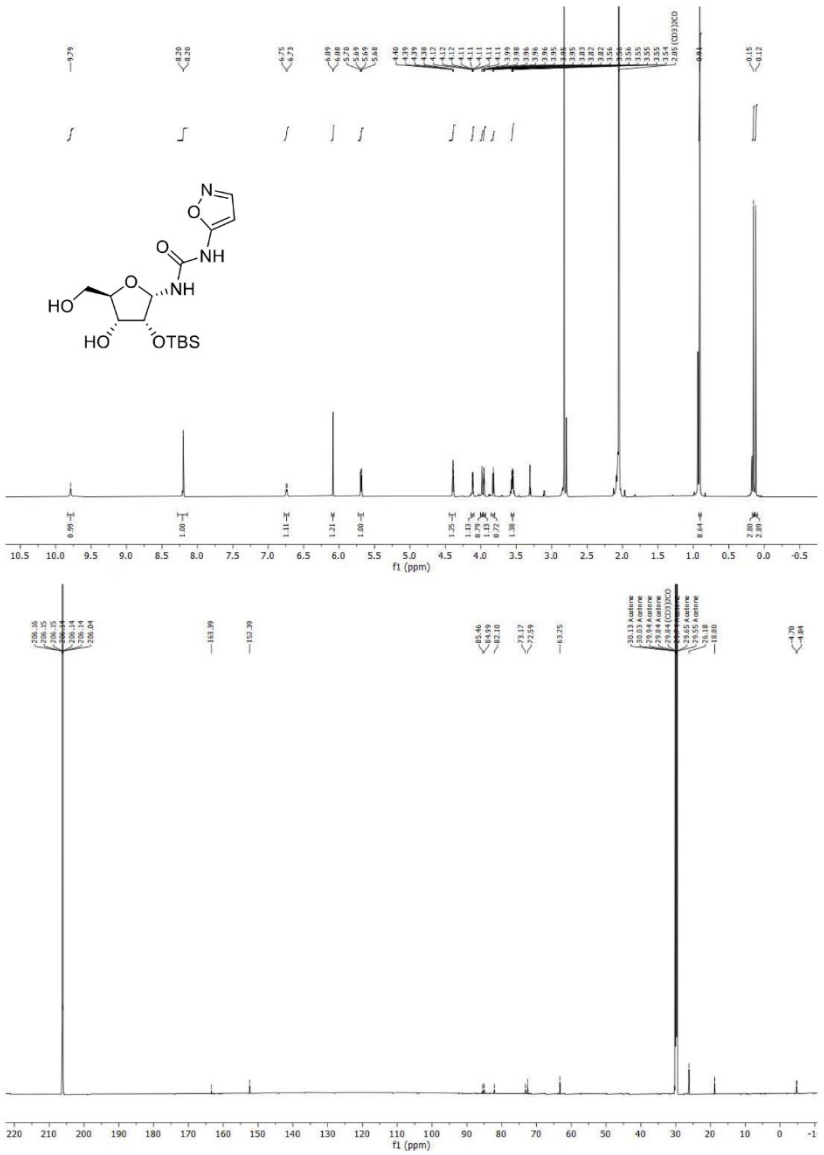
^{1}H and ^{13}C{^{1}H} NMR spectra of compound α-10



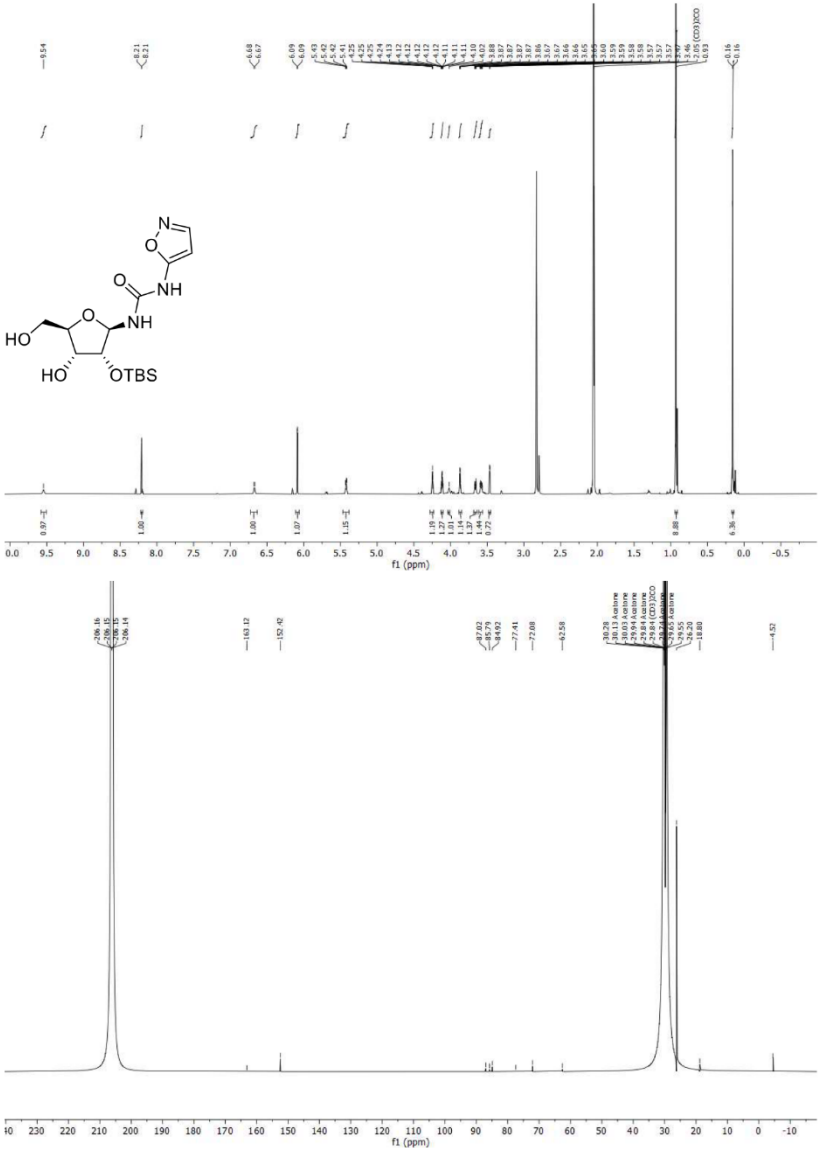
^{1}H and ^{13}C{^{1}H} NMR spectra of compound β-10



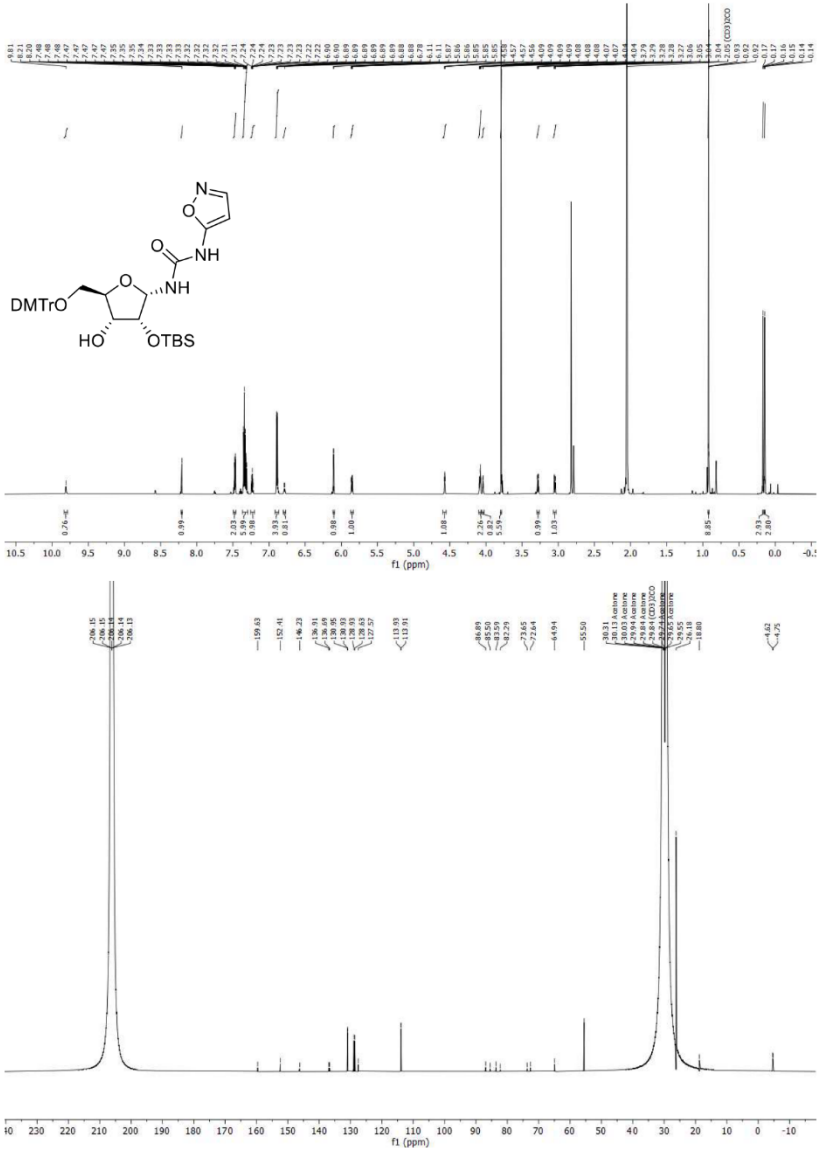
^{1}H and ^{13}C{^{1}H} NMR spectra of compound α-11



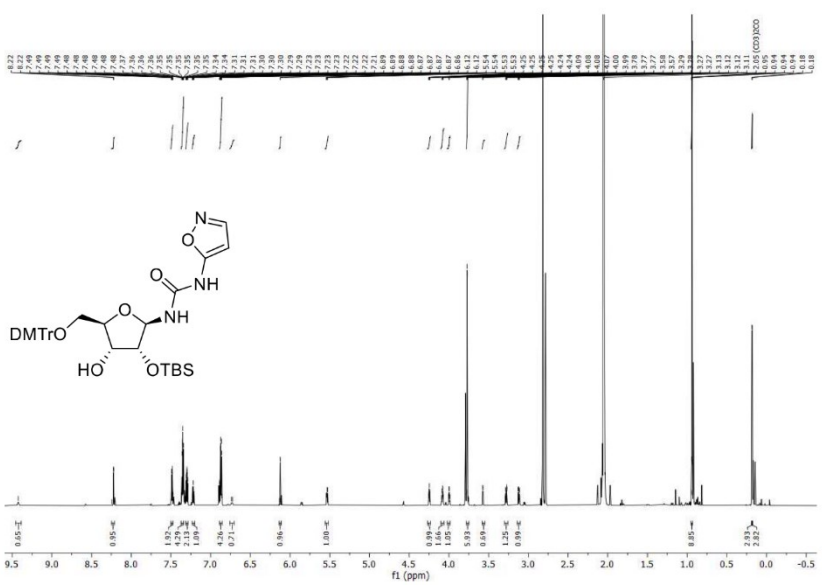
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound β-11

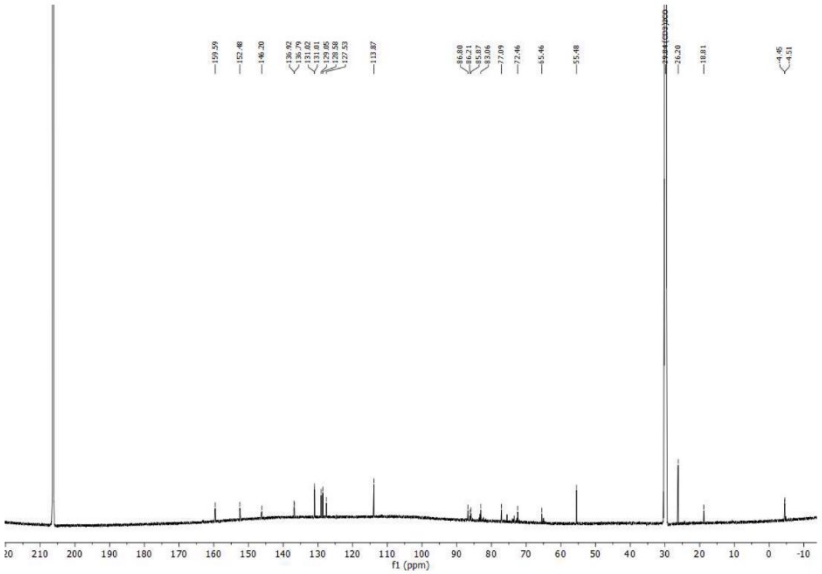


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound α-12

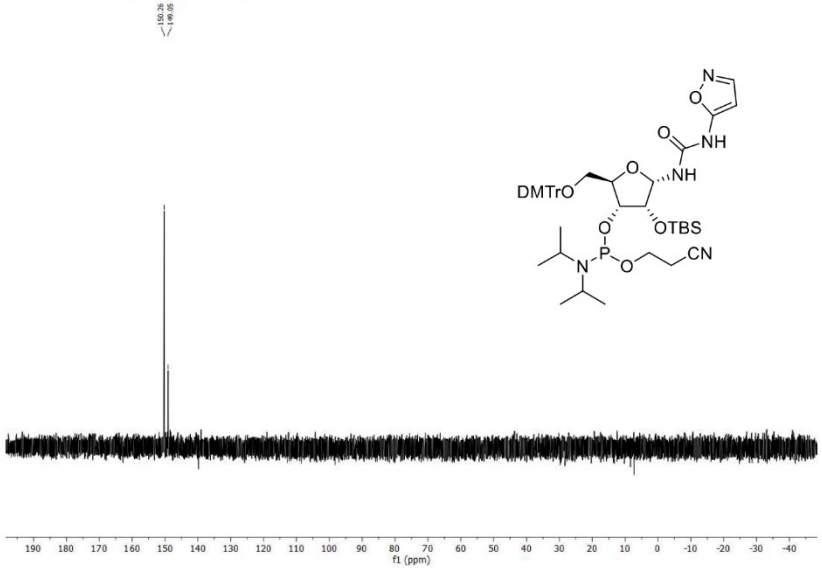


^{1}H and ^{13}C{^{1}H} NMR spectra of compound β-12

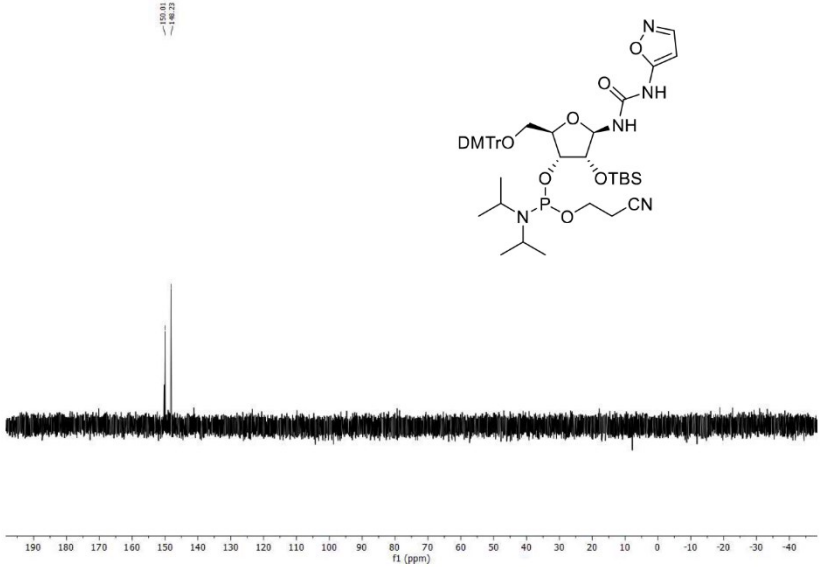




^{31}P{^{1}H} NMR spectra of compound α-13



^{31}P{^{1}H} NMR spectra of compound β-13



References

[1] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76-82.
[2] A. Štimac, J. Kobe, *Carbohydr. Res.* **1992**, *232*, 359-365.