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

Synthese hybrider Aminosäure-Nukleotide als Modell für eine RNA-Peptid-Welt

Felix Moritz Müller
aus
Worms, Deutschland

2023

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 Hilfsmittel erarbeitet.

München, den 16.06.2023

……

Felix Moritz Müller

Dissertation eingereicht am	21.06.2023
1. Gutachter	Prof. Dr. Thomas Carell
2. Gutachter	Dr. Pavel Kielkowski
Mündliche Prüfung am	03.08.2023

Gewidmet meiner Familie

„Und morgen wird die Sonne wieder scheinen
und auf dem Wege, den ich gehen werde,
wird uns, die Glücklichen sie wieder einen
inmitten dieser sonnenatmenden Erde...“
Morgen! John Henry Mackay, 1894

Danksagung

Mein erster Dank gilt meinem Doktorvater *Prof. Dr. Thomas Carell*. Ich möchte mich für diese sehr lehrreiche und in vielerlei Hinsicht spannende Zeit bedanken. In meiner Zeit im Arbeitskreis wurde ich vor einige Herausforderungen gestellt, an denen ich wachsen und mich entwickeln konnte. Das internationale und diverse Arbeitsumfeld war eine bereichernde Erfahrung für mich. Es war ebenso eine freie Arbeitsatmosphäre in der man seinen Überlegungen frei nachgehen konnte. Es war definitiv eine Zeit, die ich nicht missen möchte.

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

Dr. Markus Müller danke ich für anregende Diskussionen und seine unvergleichbaren Mühen die praktische Arbeit in den Laboren am Laufen zu halten. Die Arbeit in den Laboren wäre ohne dich nicht so nicht möglich.

Frau *Slava Gärtner* und *Dr. Nada Raddaoui* danke ich für die Hilfe bei allen bürokratischen 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 Korrekturlesen meiner Dissertation danke ich *Dr. Markus Müller* und *Johann de Graaff*.

Mein Dank gilt auch den großartigen Mitarbeitern, ohne die erfolgreiche Forschung nicht möglich gewesen wäre. An dieser Stelle möchte ich großen Dank an *Dr. Luis Escobar* richten, der mich auf einem Großteil meines wissenschaftlichen Weges begleitet hat und sowohl Mentor, als auch Freund war. Deine gewissenhafte und systematische Arbeitsweise war eine wertvolle Bereicherung. Ich möchte die Kaffeepausen mit dir auf dem Dach nicht missen. An dieser Stelle möchte ich ebenso *Johann de Graaff* danken, der als Sitznachbar immer ein offenes Ohr hatte. Sowohl bei fachlichen Problemen und auch bei privaten hast du immer eine passende Antwort parat und hast bei regelmäßigen Besuchen im Kino, beim Frisbeespielen an der Isar oder beim Lösen von Escape-Games immer für einen tollen Ausgleich zum Laboralltag gesorgt. Auch dein Engagement im Labor wo du für die Organisation von Praktika und die Funktionstüchtigkeit der Labore beigetragen hast, schätze ich sehr wert.

Ich möchte auch bei den aktuellen Kollegen der Subgroup um *Felix Xu*, *Ewa Węgrzyn*, *Jamie Chan*, *Alexander Pichler, Johannes Singer, Dr. Ivana Mejdrova* und *Kathrin Halter* danken, die durch ihre tatkräftige Arbeit die Forschung erst ermöglicht haben. Ich freue mich sehr, dass die begonnene Forschung in guten Händen ist und bin sehr gespannt, was die Zukunft für euch bringt.

Besonderer Dank gilt meinen beiden Masterstudenten *Alexander Pichler* und *Johannes Singer*, die mittlerweile Mitarbeiter im Arbeitskreis sind. Ihr habt großartige Arbeit geleistet und es war mir immer eine Freude euch zu unterstützen. Ich bin froh, dass ihr nun Teil des Arbeitskreises seid und wünsche euch viel Erfolg bei eurer weiteren Forschung.

Auch danke ich meinen Praktikanten *Chuyi Liu*, *Hans Hurmiz*, *Jakob Ewerhart* und *Ruolan Zhong* für euren großen Einsatz während eurer Bachelorarbeiten und Praktika. Es hat Spaß gemacht euch zu betreuen und ich konnte viel von euch lernen.

Zudem möchte ich mich bei allen Laborkollegen bedanken, die die Zeit im Arbeitskreis bereichert haben. Hierbei gilt der Dank unter anderem den Kollegen, *Jonas Feldmann*, *Johann de Graaff*, *Simon Veth*, *Stefan Wiedermann* und *Florian Schelter* aus dem alten Labor F4.012, die mich zu Beginn meiner Dissertation begleitet haben.

Außerdem danke ich meinen Spielpartnern im AK-Tischtennis. Es hat mir sehr geholfen einen Ausgleich zum Arbeitsalltag zu finden und es war mir eine Freude mit euch zu spielen, wenn auch zum Leidwesen für die Leute, die zu der Zeit Mittagessen waren.

Ich danke meinen Freunden sowohl in München als auch in Worms, die ihr auch schwierige Zeiten während meiner Doktorarbeit erträglicher gemacht habt und immer ein offenes Ohr hattet. Ich bin wirklich froh über den Zusammenhalt, den wir teils bereits seit der Schulzeit verspüren.

Zuletzt möchte ich den wichtigsten Personen in meinem Leben, meiner Familie danken. Ihr bietet immer ein unbeschwertes Umfeld und dafür möchte ich euch danken. Ohne euch und eure grenzenlose Unterstützung sowohl emotional als auch finanziell, wäre diese Arbeit nicht möglich gewesen. Ich weiß es sehr zu schätzen, dass ein Großteil von euch hier in der Nähe in München ist und man immer jemanden hat, mit dem man Zeit verbringen kann. Auch die Unterstützung aus meiner Heimat ist für mich in dieser Zeit immer spürbar gewesen und ich bin froh, dass ich jederzeit auch mit kurzfristiger Ankündigung zurück nach Worms kommen konnte.

Publikationsliste

Im Rahmen der Promotion erstellte Publikationen:

- M. Nainytė, F. Müller; G. Ganazzoli; C.Y. Chan; A. Crisp; D. Globisch; T. Carell, Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. *Chem. Eur. J.* **2020**, *26*, 14856–14860.

- 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.

- J.N. Singer†, F.M. Müller†, E. Węgrzyn, C. Hölzl, H. Hurmiz, C. Liu, L. Escobar, T. Carell, Loading of Amino Acids onto RNA in a Putative RNA-Peptide World. *Angew. Chem. Int. Ed.* **2023**, *62*, e202302360.

† Die Autoren haben zu gleichen Teilen zur Publikation beigetragen.

Weitere Publikationen:

- C.R. Dialer; S. Stazzoni; D.J. Drexler; F.M. Müller; S. Veth; A. Pichler; H. Okamura; G. Witte; K.P. Hopfner; T. Carell, A Click-Chemistry Linked 2′3′-cGAMP Analogue. *Chem. Eur. J.* **2019**, *25*, 2089–2095.

Konferenzbeiträge

08/2021	Molecular Origins of Life, München, Deutschland, Posterpräsentation.
06/2022	XVIIIth Symposium on Chemistry of Nucleic Acid Components, Cesky Krumlov, Tschechische Republik, Posterpräsentation.
06/2022	Molecular Origins of Life, Munich, München, Deutschland, Posterpräsentation.
10/2022	SFB1032 Annual Workshop, München, Deutschland, Vortrag.
10/2022	NASA PCE_3 Virtual Workshop; Nano-to-Cosmic Studies of Complex Systems, virtuell, Vortrag.

Inhaltsverzeichnis

Abstract

Research about the origin of life is concerned on the one hand with the abiotic origin of the building blocks of life and on the other hand with how these molecules have formed the first self-sustaining systems, which could further develop up to life as we know it today. Life is based on the interplay between deoxyribonucleic acids (DNA), ribonucleic acids (RNA) and proteins, which creates the central dogma of molecular biology. In this model, RNA is of central importance because it is the link between information storage (the genotype) and the expression of a function (the phenotype). Interestingly, RNA as a molecule possesses the property of being able to both store information and catalyze reactions. This dualism places RNA at the centre of theories about the origin of life. The RNA world hypothesis assumes that life arose from self-replicating RNA molecules, which in the course of evolution developed into increasingly complex structures with diverse catalytic properties. However, how the transition from a pure RNA world to a system involving DNA, RNA, and Proteins occurred, similar to what is found in biology today, is one of the major unanswered questions in the science of the origin of life. A possible answer can be found in the protein biosynthesis. The translation requires a complex machinery of information-containing messenger RNA (mRNA), amino acid-carrying transfer RNA (tRNA) and catalytically active ribosomal RNA (rRNA). The growing peptide chain is transferred between transfer RNAs (tRNAs), which carry a respective amino acid at the 3′-end in the form of an aminoacyl ester (**Figure *1*a**). It is assumed that this process developed gradually in the course of evolution and was preceded by more primitive forms of peptide synthesis. This hypothesis is supported on the one hand by the variety amino acid modified nucleotides that have been identified in current organisms. On the other hand, the multitude of possible reactants that could be present in a prebiotic environment suggests that early abiotic evolution was far more diverse than a pure RNA world composed just out of the four canonical nucleotides suggests. The consideration of an extended chemical space by a co-evolution of RNA and peptides would lead to an expansion of the repertoire of functional groups, which could have been a driving force of evolution. The current interplay between DNA, RNA and proteins, as described by the central dogma of molecular biology, could then develop from an RNA-peptide world by means of selection pressures.

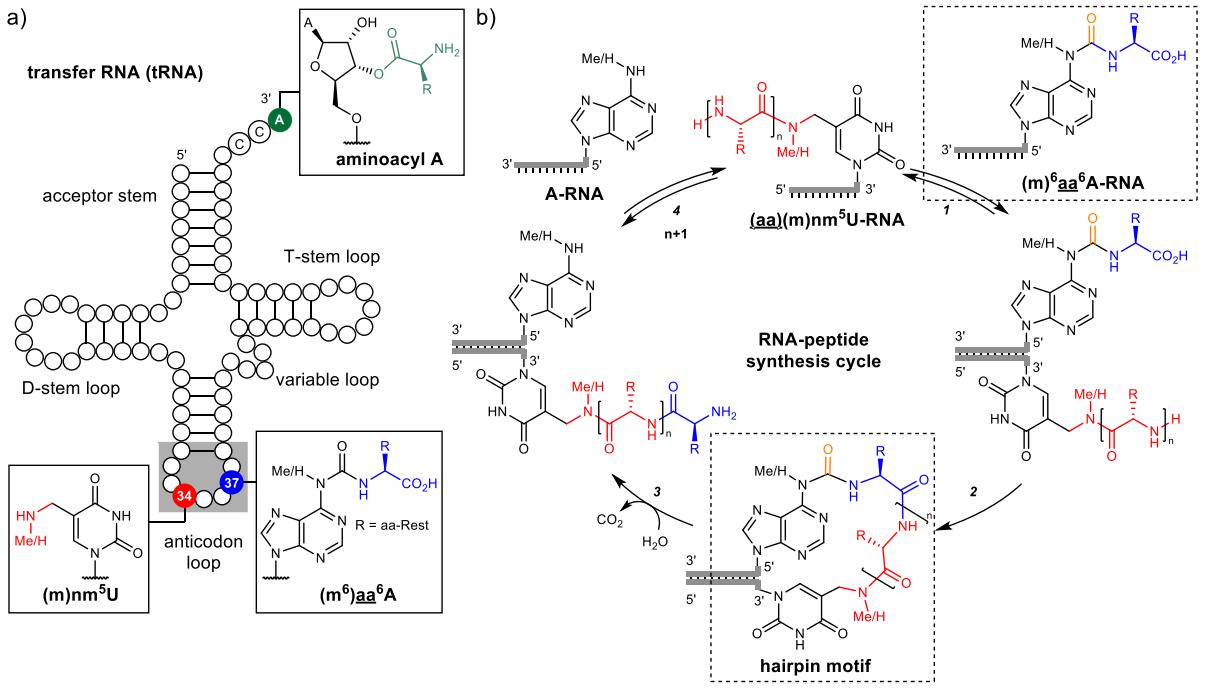


Figure 1 a) Structure of tRNA loaded with an amino acid at the ribose 3′ end. Representation of the modifications (m)nm^{5}U34 and (m^6)aa^{6}A37 in the anticodon loop. b) Prebiotic decoration of RNA with amino acids and a consecutive RNA-based transfer cycle of amino acids using non-canonical RNA modifications.

In this work, approaches were investigated in which amino acids are covalently linked to the nucleobase of various purines and pyrimidines in contrast to aminoacyl esters. Such conjugates can be found in today's tRNAs as post-transcriptional modifications. To this end, the formation of these modifications was investigated under prebiotic conditions and it was shown that *N*-methylcarbamoyl adenine can be converted into the corresponding amino acid-modified *N*^{6}-carbamoyl adenine derivatives ((m^6)aa^{6}A) in both nucleosides and oligonucleotides. In addition, synthetic routes were developed to produce the corresponding phosphoramidite building blocks, which were subsequently used to demonstrate a primitive form of RNA-based peptide synthesis using the non-canonical nucleoside (methyl)aminomethyl uridine ((m)nm^{5}U). This shows that RNA has the ability to self-decorate with peptides at specific nucleobases (**Figure *1*b**). The interaction of RNA and peptides leads to an expansion of the repertoire of functional groups, which could have been a driving force leading to an extended chemical space of ribozymes. In this regard, in addition to RNA-based peptide synthesis, it was possible to demonstrate that the covalent coupling of two RNA fragments *via* a peptide bond between two nucleobases can lead to an increased catalytic efficiency. In summary, prebiotically plausible modifications of nucleobases represent possible forms of a primordial RNA-peptide world that must be considered in the context of the origin of life.

Zusammenfassung

Die Forschung zur Entstehung des Lebens befasst sich zum einen mit der abiotischen Bildung der Bausteine des Lebens und zum anderen wie sich aus diesen Molekülen die ersten selbsterhaltenden Systeme bildeten, welche sich weiterentwickeln konnten, bis hin zu dem Leben wie wir es heute kennen. Letzteres basiert auf dem Wechselspiel zwischen Desoxyribonukleinsäuren (DNA), Ribonukleinsäuren (RNA) und Proteinen, welches das zentrale Dogma der molekularen Biologie bildet. In diesem ist die RNA von zentraler Bedeutung, da sie die Schnittstelle zwischen Informationsspeicherung (dem Genotyp) und der Ausprägung einer Funktion (dem Phänotyp) darstellt. Interessanterweise besitzt RNA die Eigenschaft sowohl Informationen speichern als auch Reaktionen katalysieren zu können. Dieser Dualismus rückt die RNA als Molekül in den Mittelpunkt der Theorien zur Entstehung des Lebens. Die RNA-Welt-Hypothese geht davon aus, dass das Leben aus selbst-replizierenden RNA-Molekülen entstand, die sich im Laufe der Evolution zu immer komplexeren Strukturen mit vielfältigen katalytischen Eigenschaften entwickelten. Wie allerdings der Übergang von einer reinen RNA-Welt hin zu einem System, basierend auf DNA, RNA und Proteinen, ähnlich wie es in der heutigen Biologie vorzufinden ist von Statten ging, ist eine der größten offenen Fragen in der Wissenschaft zur Entstehung des Lebens. Mögliche Hinweise zur Beantwortung dieser Frage lassen sich in der Proteinbiosynthese erkennen. Die Translation basiert auf einer komplexen Maschinerie aus informationsenthaltender messenger-RNA (mRNA), Aminosäure-tragender transfer-RNA (tRNA) und katalytisch aktiver ribosomaler-RNA (rRNA). Die wachsende Peptidkette wird zwischen transfer-RNAs (tRNAs) übertragen, die am 3′-Ende eine entsprechende Aminosäure in Form eines Aminoacyl-Esters besitzen (**Abbildung *1*a**). Es ist anzunehmen, dass sich dieser Prozess schrittweise im Laufe der Evolution entwickelte und diesem primitivere Formen der Peptidsynthese vorausgingen. Diese Hypothese wird gestützt durch die Vielfalt an Aminosäure-modifizierten Nukleotiden, die in heutigen Lebensformen identifiziert wurden. Zudem deutet die Vielzahl an möglichen Reaktanten, die in einer präbiotischen Umgebung vorhanden sein können darauf hin, dass die frühe abiotische Evolution sich weitaus vielfältigeren Molekülen bediente, als denen einer reinen RNA-Welt, bestehend aus nur den vier kanonischen Nukleotiden. Die Berücksichtigung eines erweiterten chemischen Spielraums durch eine Co-Evolution von RNA und Peptiden würde zu einer Ausweitung des Repertoires an funktionellen Gruppen führen, was eine Triebkraft der Evolution gewesen sein könnte. Aus dieser RNA-Peptid-Welt könnte sich dann im Laufe der Evolution über einwirkende Selektionsdrücke das heutige Zusammenspiel zwischen DNA, RNA und Proteinen, wie es das zentrale Dogma der molekularen Biologie beschreibt, entwickeln.

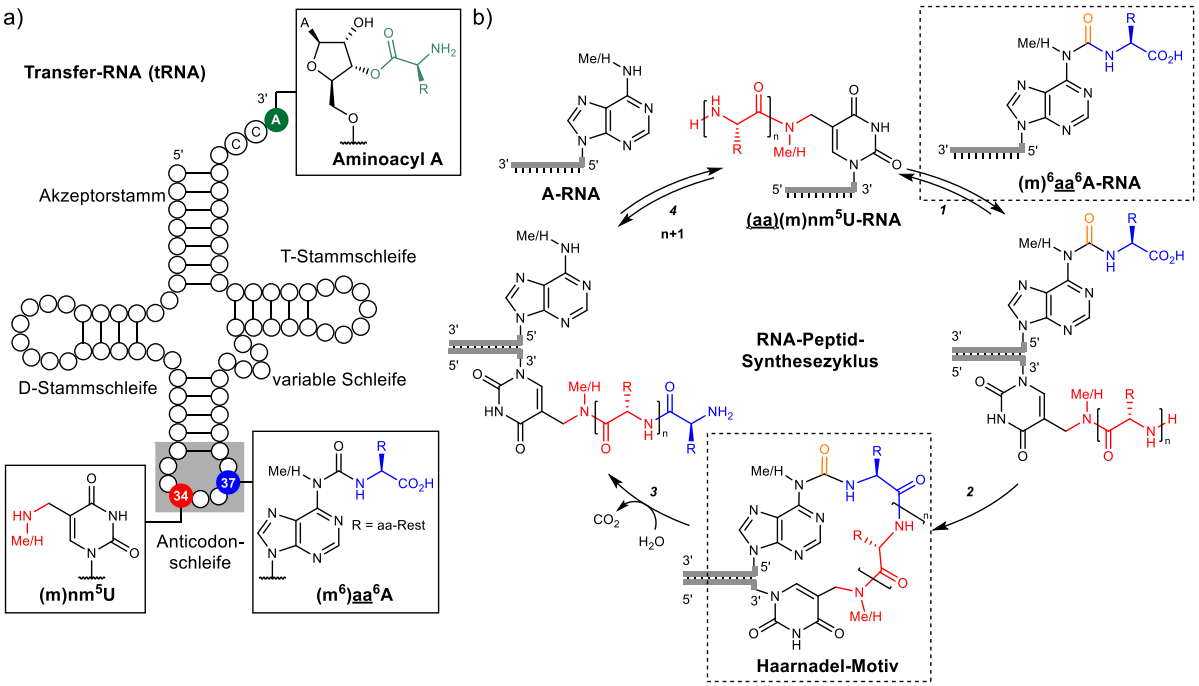


Abbildung 1 a) Struktur der tRNA beladen mit einer Aminosäure an der Ribose am 3′-Ende. Darstellung der Modifikationen (m)nm^{5}U34 und (m^6)aa^{6}A37 in der Anticodonschleife. b) Präbiotische Modifizierung von RNA mit Aminosäuren und ein konsekutiver RNA-basierter Transferzyklus von Aminosäuren mit Hilfe nicht-kanonischer RNA-Modifikationen.

In dieser Arbeit wurden Ansätze untersucht, in denen Aminosäuren kovalent mit der Nukleobase verschiedener Purine und Pyrimidine, im Gegensatz zu Aminoacylestern, verknüpft sind. Solche Konjugate lassen sich in tRNAs als post-transkriptionelle Modifikationen wiederfinden. Dazu wurde die Bildung dieser Modifikationen unter präbiotischen Bedingungen untersucht und es konnte gezeigt werden, dass *N*-Methylcarbamoyl-Adenin sowohl in Nukleosiden als auch Oligonukleotiden in die entsprechenden Aminosäure-modifizierten *N*^{6}-Carbamoyl-Adenosinderivate ((m^6)aa^{6}A) umgewandelt werden kann. Außerdem wurden Syntheserouten entwickelt, um die entsprechenden Phosphoramidit-Bausteine herzustellen, welche anschließend genutzt wurden, um eine primitive Form einer RNA-basierten Peptidsynthese mit der Hilfe des nicht-kanonischen Nukleosids (Methyl)aminomethyl-Uridin ((m)nm^{5}U) zu demonstrieren. Hierbei konnte gezeigt werden, dass RNA die Fähigkeit besitzt, sich selbst mit Peptiden zu dekorieren (**Abbildung *1*b**). Das Zusammenspiel von RNA und Peptiden führt somit zu einer Erweiterung des Repertoires an funktionellen Gruppen, was eine Triebkraft gewesen sein könnte, die zu einer erhöhten katalytischen Aktivität von Ribozymen führte. Diesbezüglich konnte neben der RNA-basierten Peptidsynthese demonstriert werden, dass die kovalente Kopplung zweier RNA-Fragmente über eine Peptidbindung zwischen zwei Nukleobasen zu einer Erhöhung der katalytischen Effizienz führen kann. Zusammenfassend stellen präbiotisch plausible Modifikationen der Nukleobasen mögliche Formen einer uranfänglichen RNA-Peptid-Welt dar, die im Kontext der Entstehung des Lebens berücksichtigt werden muss.

1. Einleitung

Der Ursprung des Lebens ist nach der Entstehung der Erde vor rund 4 Milliarden Jahren eine der großen ungelösten Fragen der Menschheit. Das Leben selbst ist in seiner ganzen Vielfalt Gegenstand umfangreicher naturwissenschaftlicher Forschung. Obwohl wir immer mehr über das Leben in seiner heutigen Form wissen, ist unser Wissen über den Ursprung noch immer wage. Trotz zahlreicher Meilensteine in der Molekularbiologie in den letzten Jahrzehnten, von der Aufklärung der molekularen Struktur unseres Erbguts im Jahr 1953, postuliert von *James Watson* und *Francis Crick*,[1] bis hin zu Projekten zur Sequenzierung des menschlichen Genoms im Jahr 2001[2], bleibt die Rekonstruktion der Entstehung des Lebens eine faszinierende Herausforderung. Einer der Pioniere in der Evolutionsforschung, *Charles Darwin*, hat in seinen Studien gezeigt, dass sich Lebensformen über natürliche Selektion den entsprechenden Lebensräume anpassen.[3] Bevor jedoch die biologische Evolution einsetzte, stellt sich in der Forschung zur Präbiotik die Frage, wie aus unbelebter Materie die ersten selbsterhaltenden Systeme entstanden sind und ob die Prinzipien der *Darwin´schen* Evolutionstheorie auch auf unbelebte Materie übertragbar ist.[4][5][6]

Die Disziplin der Phylogenetik beschäftigt sich mit der Abstammung aller Lebewesen. Phylogenetischen Studien implizieren, dass der Ursprung allen Lebens auf einen hypothetischen Urvorfahren, den *last universal common ancestor* (LUCA) zurückgeführt werden kann.[7][8] Durch molekulargenetische Analysen der Genome wurden bestimmte Gene identifiziert, die in allen drei Domänen des Lebens[9] konserviert sind. Dieser *Top-Down*-Ansatz ermöglicht es die biologische Evolution nachzuvollziehen und die Verteilung konservierter Gene über die Reiche des Lebens hinweg zu analysieren. Diese Analysen zeigen, dass vor allem Gene konserviert blieben, die für die Prozessierung von Ribonukleinsäuren (RNA) bei der Transkription und Translation verantwortlich sind.[10][11] Jede Lebensform auf diesem Planeten basiert auf dem Zusammenspiel von polymeren Biomolekülen, d.h. Desoxyribonukleinsäuren (DNA), Ribonukleinsäuren (RNA) und Proteinen, die das Leben definieren. Das zentrale Dogma der molekularen Biologie wurde 1970 von *Francis Crick* postuliert.[12] Es beschreibt, dass die genetische Information im Prozess der Replikation vervielfältigt und in der Transkription von der DNA in die RNA umgeschrieben wird, die dann in der Translation anhand des genetischen Codes in Proteine übersetzt wird (**Abbildung *2*a**).[12] Nukleinsäuren fungieren dabei als Träger der Erbinformation (Genotyp) und Proteine übernehmen enzymatische Prozesse (Phänotyp), um sowohl die RNA zu replizieren als auch die gespeicherte Information zu entschlüsseln und in funktionsfähige Proteine zu übersetzen.

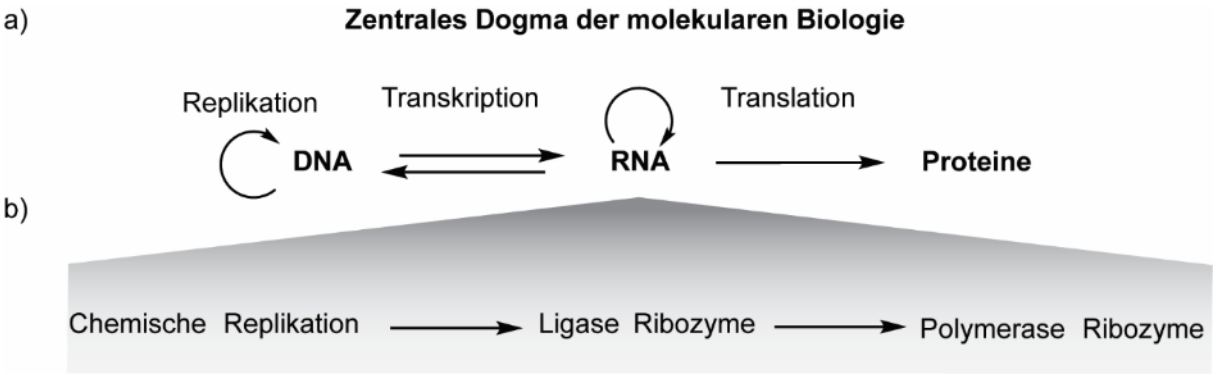


Abbildung 2. a) Zentrales Dogma der molekularen Biologie und b) Mögliche Stufen der chemischen Evolution, die von diversen Stufen der Replikation zur heutigen Biologie führten.

1.1 Die RNA-Welt-Theorie

Die potentielle Fähigkeit der Nukleinsäuren, Informationen zu speichern führte bereits in den 1960er Jahren zu der Hypothese, dass die RNA, die sich selbst replizieren und Reaktionen katalysieren kann, das zentrale Molekül bei der Entstehung des Lebens gewesen sein könnte, noch bevor die Proteinsynthese auftrat.[13][14][15] In den 1980ern konnte auch experimentell nachgewiesen werden, dass RNA nicht nur Träger der genetischen Information sein kann, sondern auch katalytisch aktiv sein kann.[16][17] Nach den Entdeckungen von *Thomas Cech*[18] und *Sidney Altman*,[19] dass RNA selbst als Katalysator auftreten kann, prägte *Walter Gilbert* 1986 den Begriff der RNA-Welt-Theorie,[20] die postuliert, dass die chemische Evolution möglicherweise allein auf RNA als Molekül zurückgeführt werden kann. *Cech* und Mitarbeiter fanden heraus, dass in *self-splicing introns* von *Tetrahymena* der Prozess durch RNA katalysiert wird.[21][22][23][24][25][26][27][28] Ebenso entdeckten Altman und Mitarbeiter, dass die RNA-Komponente von RNAse P für die katalytische Funktion verantwortlich ist.[29][30] Somit konnte experimentell nachgewiesen werden, dass die RNA beide Funktionen, Genotyp und Phänotyp, in einem Molekül vereinen kann. Diese Fähigkeit der RNA könnte dazu geführt haben, dass sie zu Beginn der Evolution Prozesse durchführte, die später von effizienteren Katalysatoren basierend auf Proteinen übernommen wurden.[31] Ausgehend von der RNA-Welt-Hypothese stellt sich in der Präbiotik die Frage, wie sich die ersten Bausteine des Lebens unter Bedingungen, die auf einer frühen Erde vorlagen, über abiotische Wege formten und polymere Strukturen bilden konnten. Diese Polymere könnten dann in einem folgenden Schritt als Templat für die eigene chemische oder gar ribozymatische Replikation gedient haben (**Abbildung 2b**). Durch weitere einwirkende Selektionsdrücke im Laufe der Evolution könnten sich weitere Ribozyme mit komplexeren Eigenschaften gebildet haben, die sich in Lipid-Vesikel einhüllten, woraus sich das heutige Zusammenspiel zwischen DNA, RNA und Proteinen, wie es das zentrale Dogma der molekularen Biologie beschreibt, entwickelte.[32][33][34][35]

1.1.1 Aufbau kanonischer und nicht-kanonischer Nukleotide

Die besondere Eigenschaft sowohl Informationen zu speichern, als auch Reaktionen katalysieren zu können, beruht auf der komplexen Struktur der Nukleinsäuren. Sie bestehen aus Nukleobasen, der chiralen D-Ribose und dem Phosphat-Rückgrat, über das die einzelnen Nukleoside an der 3′- und 5′-Position der Ribose verknüpft sind (**Abbildung *3*a**). Die Abfolge der jeweiligen Nukleotide bestimmt die Information, welche anhand des genetischen Codes bei der Translation in eine Aminosäuresequenz übersetzt werden kann. Dies beruht auf der Fähigkeit antiparallele Doppelhelices zu bilden, in denen die Purin-Nukleobasen Adenin (A) und Guanin (G) entsprechend mit den Pyrimidin-Nukleobasen Uracil (U) und Cytosin (C) Wasserstoffbrückenbindungen ausbilden können und somit eine molekulare Erkennung stattfinden kann (**Abbildung *3*b**).[36]

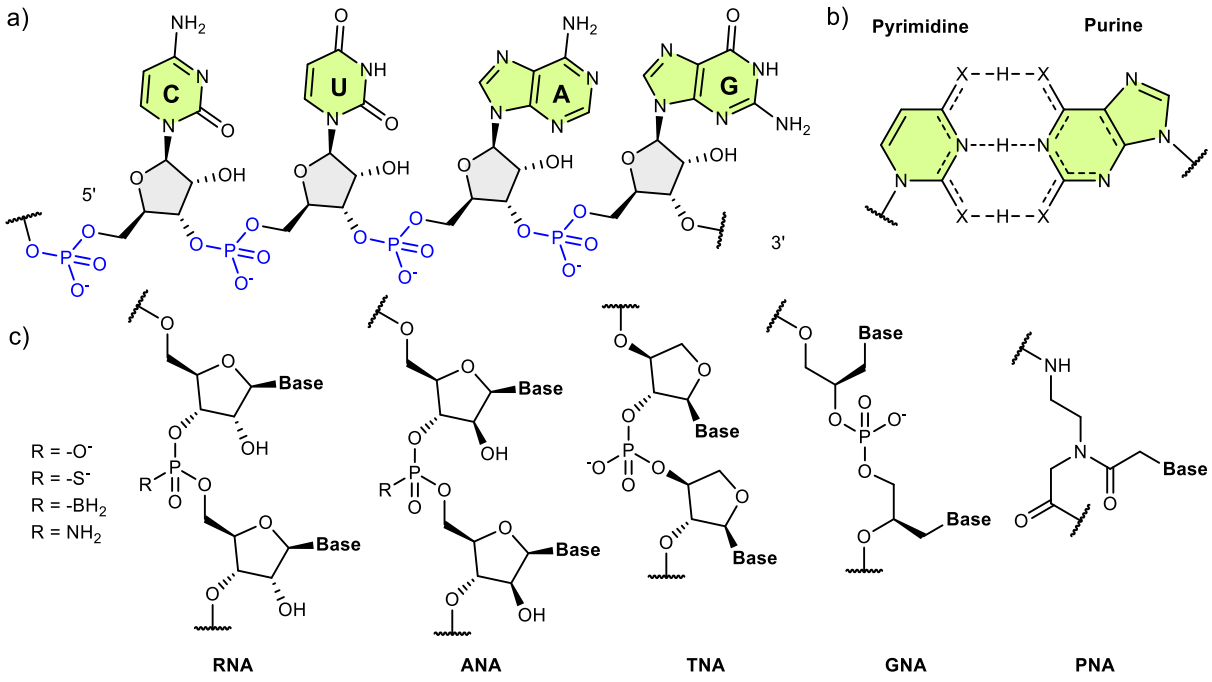


Abbildung 3. a) Molekularer Aufbau der RNA aus den vier Nukleobasen Cytosin (C), Uracil (U), Adenin (A) und Guanin (G) in grün, Ribose in grau und dem Phosphat-Rückgrat in blau. b) Allgemeine Darstellung der Basenpaarungen zwischen Pyrimidinen und Purinen. X = Heteroatom.[36] c) Modifikationen des Phosphatrückgrats und der Zuckereinheit. RNA = Ribonukleinsäuren; ANA = Arabinonukleinsäuren; TNA = Threonukleinsäure; GNA = Glycolnukleinsäure; PNA = Peptidnukleinsäure.

In einem *Bottom-up*-Ansatz stellt sich die Frage nach den ursprünglichen Ausgangsmaterialien und Bedingungen, unter denen die ersten Bausteine des Lebens über eine abiotische Synthese entstanden sind.[37] In den ersten Experimenten von *Urey* und *Miller*, in denen die Bedingungen einer möglichen frühen Erde simuliert wurden, wurde beobachtet, dass in Gasentladungsexperimenten unter einer bestimmten Atmosphäre aus CH_4, NH_3, H_2O und H_2 die ersten Moleküle des Lebens in Form von Aminosäuren entstehen können.[38][39][40] Wie die

ersten Nukleoside als Bausteine der RNA-Welt entstanden sein könnten, konnte in nachfolgenden Experimenten der präbiotischen Chemie in einigen wichtigen Durchbrüchen gezeigt werden. Diese Ergebnisse wurden in der Literatur vielfach diskutiert.[41][37][42][43] Forschungsgruppen konnten demonstrieren, dass sowohl Purin-[44][45][46] als auch Pyrimidin-Nukleoside[47][48][49] und auch deren Nukleotide[48] unter präbiotisch plausiblen Bedingungen[50][51] einer frühen Erde entstehen können.

Den Rahmen für die Auswahl präbiotischer Ausgangsmaterialien bieten zum einen Untersuchungen an extraterrestrischem Material wie z.B. Kometen bzw. Asteroiden[52][53] oder auch spektrale Analysen des Weltraums[54] bzw. der Umgebung von Sternen[55][56] oder Planeten. Unter anderem konnten Edukte für die Synthese von Kohlenhydraten wie z.B. Ribose im interstellaren Raum nachgewiesen werden.[54] In der Formosereaktion[57] können aus Formaldehyd und Glycolaldehyd größere Kohlenhydrateinheiten aufgebaut werden. Die Stabilität der Ribose und ihre Akkumulation unter präbiotischen Bedingungen galt lange Zeit als limitierender Faktor.[58] Jedoch konnte gezeigt werden, dass diese Einschränkungen in Anwesenheit von Boratmineralien umgangen und die Ribose stabilisiert werden kann.[59][60] Bisher wurde jedoch noch kein präbiotischer Syntheseweg für Nukleoside entdeckt, bei dem ausschließlich *β*-Ribofuranoside erhalten wurden. Stattdessen wurden neben den in der heutigen Biologie beobachteten kanonischen Nukleosiden auch diverse Regio- und Konfigurationsisomere der Nukleoside nachgewiesen[42], was darauf hindeutet, dass die Zusammensetzung der chemischen Bausteine der frühen Erde weitaus vielfältiger war, als sich aus einem *Top-Down* Ansatz[61] mit einem Blick auf die Molekularbiologie vermuten lässt.

Es konnte gezeigt werden, dass sowohl kanonische als auch nicht-kanonische Nukleoside in Nass- und Trockenzyklen gebildet werden können, welche auch als modifizierte Bausteine in heutigen Lebensformen vorkommen.[62] Daraus lässt sich vermuten, dass das Leben aus einer komplexen Mischung an Molekülen entstanden ist, von denen manche bis heute in Lebewesen nachgewiesen werden können.[63][64] Auf der Grundlage dieser Beobachtungen wurden einige alternative präbiotische Szenarien beschrieben, in denen die Synthese und Eigenschaften nicht-kanonischer Nukleotide untersucht wurden.[65] Dazu gehören zum einen Modifikationen der Nukleobase[66][67] und auch Modifikationen der Zuckereinheit. Es konnte gezeigt werden, dass Nukleinsäuren, die auf Glycol,[68] Threose[69][70][71] oder Arabinose[72][73][74] basieren, stabile Duplexe mit unmodifizierter DNA oder RNA bilden können. Ebenso konnte gezeigt werden, dass die codierenden Eigenschaften der Nukleobasen wenig von der Beschaffenheit des verbindenden Rückgrats abhängen,[75] *d.h.* weder von der Zuckereinheit[76][77] noch von der chemischen Zusammensetzung des Phosphatrückgrats in Form von Phosphorthioaten[78][79], Boranophosphaten[80], Phosphoramidaten[81][82] oder von Substitutionen der gesamten verknüpfenden Struktur z.B. durch Peptidnukleinsäuren[83][84] (**Abbildung *3*c**). Letztere wurden aufgrund ihrer höheren Stabilität gegenüber Hydrolyse im Vergleich zu

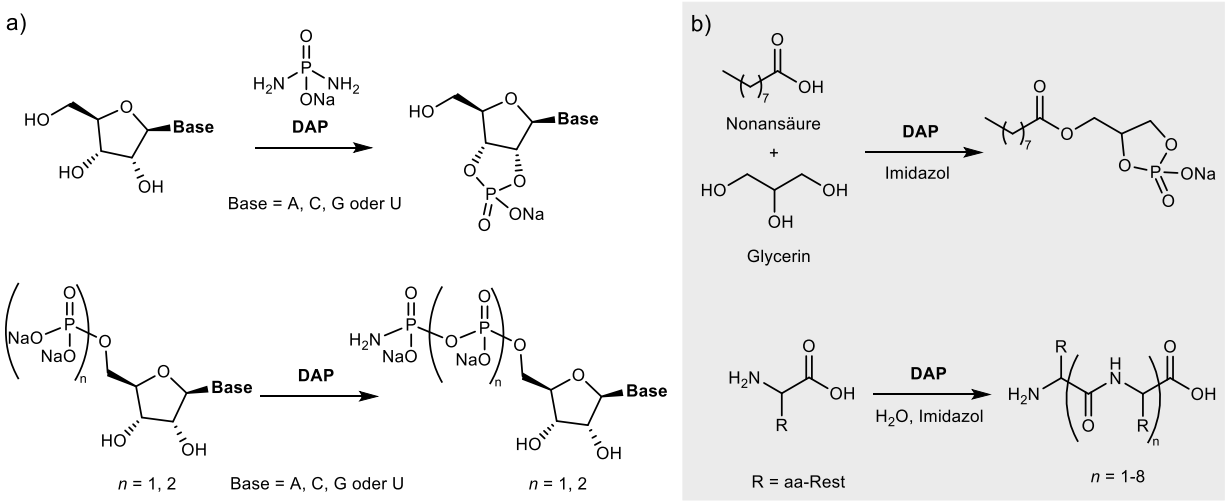
Phosphordiesterbindungen als mögliche Kandidaten für eine präbiotische Evolution in Betracht gezogen.[85][86][87] Somit konnte demonstriert werden, dass diese alternativen Nukleinsäuren ebenso geeignet sind, genetische Information zu speichern und zu replizieren, bis hin zur Entwicklung katalytischer Eigenschaften.[88] In Anbetracht der RNA-Welt-Theorie stellt sich jedoch die Frage, wie sich letztendlich aus dieser heterogenen Mischung präbiotisch plausibler Nukleotide über eine *Darwin´sche* chemische Evolution eine molekulare Spezies in Form von RNA anreicherte. In dieser Hinsicht, wurden Oligomerisierungsstudien durchgeführt, die darauf hindeuten, dass sowohl Template als auch Monomere mit modifizierten Nukleobasen- und Zuckereinheiten in nicht-enzymatischen Primer-Extensionsexperimenten weiterhin als RNA repliziert werden können, wobei eine kinetische Überlegenheit der RNA im Vergleich zu Threo- oder Arabinonukleinsäuren festgestellt wurde.[89] Auf Grund der drei Hydroxygruppen eines Ribonukleosids ist nicht nur die molekulare Beschaffenheit des verknüpfenden Rückgrats variabel, sondern auch die Verbindung der Monomere in Form von Konstitutionsisomeren. Natürlich vorkommende RNA ist über die 3′-5′-Hydroxygruppen der Ribose über Phosphordiester miteinander verknüpft, wobei aus präbiotischer Sicht auch eine Verknüpfung über die 2′- und 5′-Hydroxygruppen denkbar ist. Studien hierzu haben gezeigt, dass sich die Bevorzugung der 3′-5′-Bindung gegenüber der 2′-5′-Bindung durch eine präbiotisch plausible Acetylierung der 2′-Position entwickelt haben könnte, da die 2′-5′-Verknüpfung unter basischen Bedingungen schneller hydrolysiert.[90]

1.1.2 Phosphorylierung von Nukleosiden

In der heutigen Biologie wird die genetische Information durch enzymatische Replikations- oder Transkriptions-Prozesse vervielfältigt bzw. übertragen, wobei Nukleosid-Triphosphate als aktivierte Monomere verwendet werden.[31] Dazu sind komplexe Enzyme erforderlich, deren Existenz in einer frühen RNA-Welt in Frage gestellt wird. Um Ribozyme als Katalysatoren für die Replikation von Information zu nutzen, muss zunächst eine nicht-enzymatische Form der Oligomerisierung von monomeren Bausteinen stattgefunden haben.[91] Daher spielt neben der präbiotischen Synthese der Nukleoside auch die Phosphorylierung und die anschließende chemische Verknüpfung von Monomeren zu Oligomeren eine ebenso wichtige Rolle bei der Entschlüsselung der Entstehung des Lebens. Um die Bildung oligomerer Nukleotide zu realisieren, werden reaktive Ausgangsstoffe benötigt, die in einer Kondensationsreaktion die Abspaltung von Wasser ermöglichen.[92]

Diverse Möglichkeiten der Phosphorylierung mit Orthophosphaten, der häufigsten Form von Phosphatsalzen auf der Erdoberfläche, wurden im präbiotischen Kontext beschrieben.[93] Erste Untersuchungen von *Orgel* und Mitarbeitern im Jahre 1968 befassten sich mit der Reaktion

von Nukleosiden mit Phosphatsalzen und Kondensationsreagenzien in wässriger Lösung. Phosphorylierte Nukleoside konnten jedoch aufgrund der konkurrierenden Reaktion von Wasser mit dem aktivierten Phosphat nur in geringen Ausbeuten nachgewiesen werden.[94] Durch Zugabe von Harnstoff und bei erhöhten Temperaturen konnte die Ausbeute auch in Abwesenheit von Kondensationsreagenzien gesteigert werden.[95][96] Um die Hydrolyse der aktivierten Phosphate zu vermeiden, wurde auch Formamid als mögliches präbiotisches Lösungsmittel in Betracht gezogen. So konnten Nukleoside durch Reaktion mit Phosphatsalzen (KH_2PO_4)[97] oder Phosphatmineralen[98] in Formamid zu den entsprechenden Nukleotiden umgesetzt werden. Neben der Phosphorylierung durch Orthophosphate, wurde auch die Verwendung reaktiverer Phosphatspezies in Betracht gezogen.[99][100][93] Dabei wurden Anhydride des Phosphats, d.h. Pyrophosphate oder Metaphosphate, die im Vulkanismus nachgewiesen werden können, untersucht.[101] Einerseits können Trimetaphosphate Nukleoside an der 2′- oder 3′-Position phosphorylieren[102] und andererseits können sie auch Aminosäuren aktivieren und so zu einer Peptidsynthese führen.[103] Außerdem kann Trimetaphosphat als Ausgangsmaterial mit Ammoniak zu Diamidophosphatsalzen (DAP) reagieren,[104] die unter anderem Zuckermoleküle phosphorylieren können.[105] Es wurde auch gezeigt, dass DAP alle vier kanonischen Ribonukleoside in hohen Ausbeuten von 27-89% zu 2′-3′-zyklischen Nukleosid-Monophosphate (2′-3′-cNMP) und 5′-Monophosphate zu den entsprechenden Amidodiphosphaten und Amidotriphosphaten umsetzen kann (**Schema *1*a**).[106]



Schema 1. a) Phosphorylierung von Nukleosiden und Nukleotiden mit Diamidophosphat (DAP). b) Imidazol-katalysierte Aktivierung von Carbonsäuren mit DAP zu Phospholipidvorstufen und Oligopeptiden.[106]

Analog zu den natürlich vorkommenden Nukleosid-Triphosphaten (NTPs) werden diese sowohl von proteinbasierten als auch von Ribozym-Polymerasen als Substrate erkannt, was einen möglichen Übergang von der abiotischen hin zur biotischen Polymerisation darstellen kann.[107] Neben der Phosphorylierung von Nukleosiden und Nukleotiden, reagieren auch

Aminosäuren mit DAP, wobei diese zu Peptiden oligomerisieren und Glycerin zusammen mit Fettsäuren in der Gegenwart von DAP zu einfachen Phospholipidvorstufen reagiert (**Schema *1*b**).[106] Solch eine Form der Aktivierungschemie von Phosphaten und Carbonsäuren deutet darauf hin, dass im Fall einer Co-Existenz von Nukleotiden, Aminosäuren und Fettsäuren simultan eine Entwicklung zu höherer molekularer Vielfalt stattgefunden haben könnte.

1.1.3 Oligomerisierung aktivierter Nukleotide

Nach der Phosphorylierung der Nukleoside stellt sich die Frage, wie diese anschließend aktiviert werden und in einem nicht-enzymatischen Prozess Oligomere bilden.[63][35] Es konnte gezeigt werden, dass 2′-3′-zyklische Nukleotide[108] und 3′-5′-zyklische Nukleotide[109] als Ausgangsmaterial für Oligomerisierung verwendet werden können, wobei jedoch nur eine geringe Effizienz erzielt wurde. Plausibler erscheinen daher Nukleotide mit einem aktivierten 5′-Phosphat, zumal bei der enzymatischen DNA- und RNA-Synthese 5′-Triphosphate als Substrate für die Polymerisation verwendet werden (**Abbildung *4*a**).

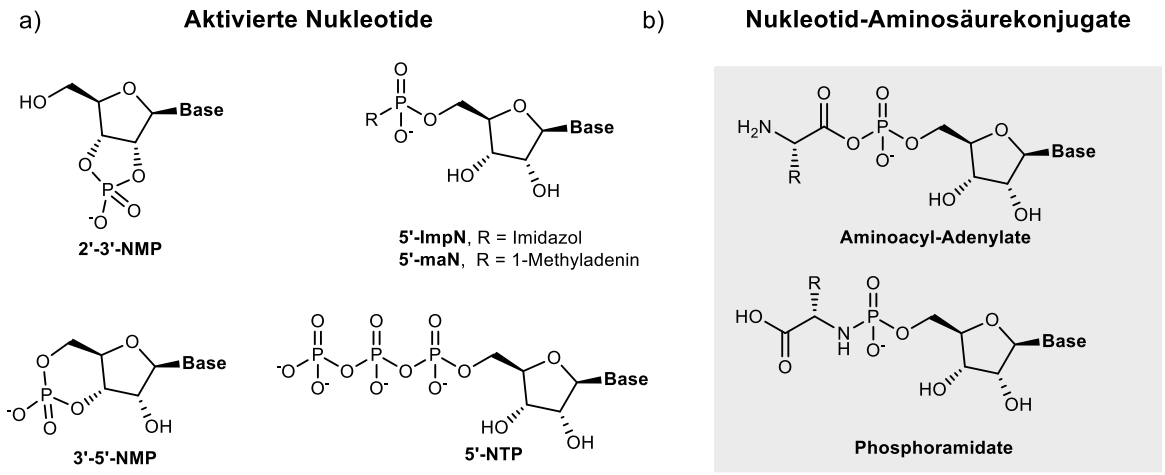


Abbildung 4. a) Aktivierte Nukleotide als Ausgangsmaterialien für die Bildung von Oligonukleotiden. b) Synthese RNA-Aminosäurekonjugaten.

Die zuvor erwähnten 5′-Phosphoramidate in Form von Phosphorimidazoliden sind plausible Kandidaten für eine Oligomerisierung in wässriger Lösung. Diese können aus den entsprechenden Polyphosphaten und Imidazol gewonnen werden.[110] Bei der Oligomerisierung durch Imidazol-aktivierte Nukleotide ergeben sich allerdings neue Herausforderungen, da im Nukleotid verschiedene Nukleophile vorhanden sind, die mit der aktivierten Spezies reagieren können. So bilden sich neben der kanonischen 3′-5′-Verknüpfung, vor allem Pyrophosphate und nicht-kanonische 2′-5′-Phosphodiester.[111] Dieses Problem der Regiospezifität konnte durch Adsorption der aktivierten Nukleotide an die Oberfläche des Minerals Montmorillonite umgangen werden, wodurch die Bildung der 3′-5′-

Bindung bevorzugt wurde. In einem Ansatz von *Orgel* und Mitarbeitern konnten somit präbiotische Oligomere in einer Länge von bis zu 50 Nukleotiden erhalten werden.[112][113] Neben 5′-Phosphorimidazoliden wurden auch über 1-Methyladenin aktivierte 5′-Phosphate verwendet, welche eine ähnliche Selektivität von ~70% für die kanonische 3′-5′-Bindung für die Oligomerisierung von aktiviertem Adenosin zeigten.[114]

Neben der *de novo* Bildung von Oligonukleotiden spielt die Replikation bestehender Sequenzen eine wichtige Rolle, um den Erhalt evolutionärer Vorteile zu gewährleisten. In Studien zur Oligomerisierung aktivierter Guanosin-Nukleotide in Anwesenheit von Poly-Cytidin-Templaten wurde gezeigt, dass die Verwendung von 2-Methylimidazol[115] als Abgangsgruppe oder der Zusatz von Zn^{2+} zu der Reaktion,[116] die Regioselektivität in Richtung der 3′-5′-Bindung verschieben kann. *Clemens Richert* und Mitarbeitern konnten zeigen, dass durch *in situ* Aktivierung von Nukleotiden mit Carbodiimiden, d.h. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimid (EDC), und in Gegenwart eines Katalysators in Form verschiedener Imidazolderivate ein nahezu quantitativer Einbau eines komplementären Nukleotids erfolgen kann.[117] Ähnlich wie bei der *de novo* Synthese von Oligonukleotiden, bei der bereits synthetisierte 5′-Phosphorimidazolide verwendet wurden,[118] konnte auch bei der *in situ* Aktivierung zur chemischen Replikation eines Templat-Strangs die Bildung gänzlich neuer Oligomere beobachtet werden. Interessanterweise konnte auch gezeigt werden, dass in Anwesenheit von Aminosäuren unter den gleichen Reaktionsbedingungen wie bei der *in situ* RNA-Replikation kovalente Konjugate aus RNA und Aminosäuren in Form von Phosphoramidaten entstehen (**Abbildung *4*b**).[119][120][121][122][123] Das Kondensationsreagenz EDC ist zwar präbiotisch nicht plausibel, aber Carbodiimide generell sind das Tautomer des Cyanamids, das unter präbiotischen Bedingungen[124] gebildet werden kann und auch im Rahmen der präbiotischen Synthese von Nukleosiden untersucht wurde (**Abbildung *5***).[48]

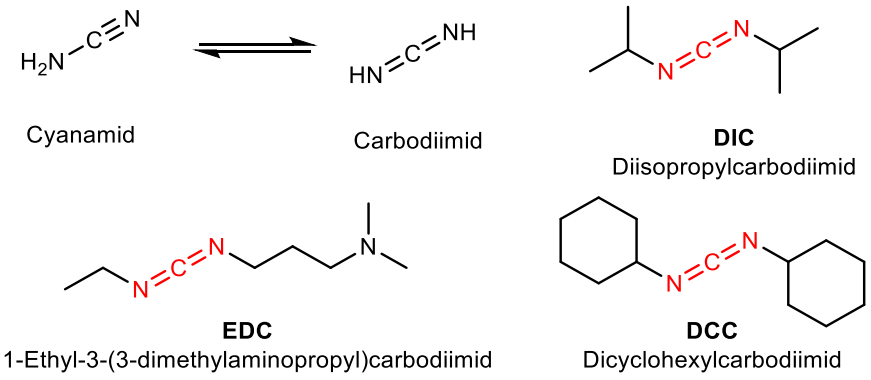


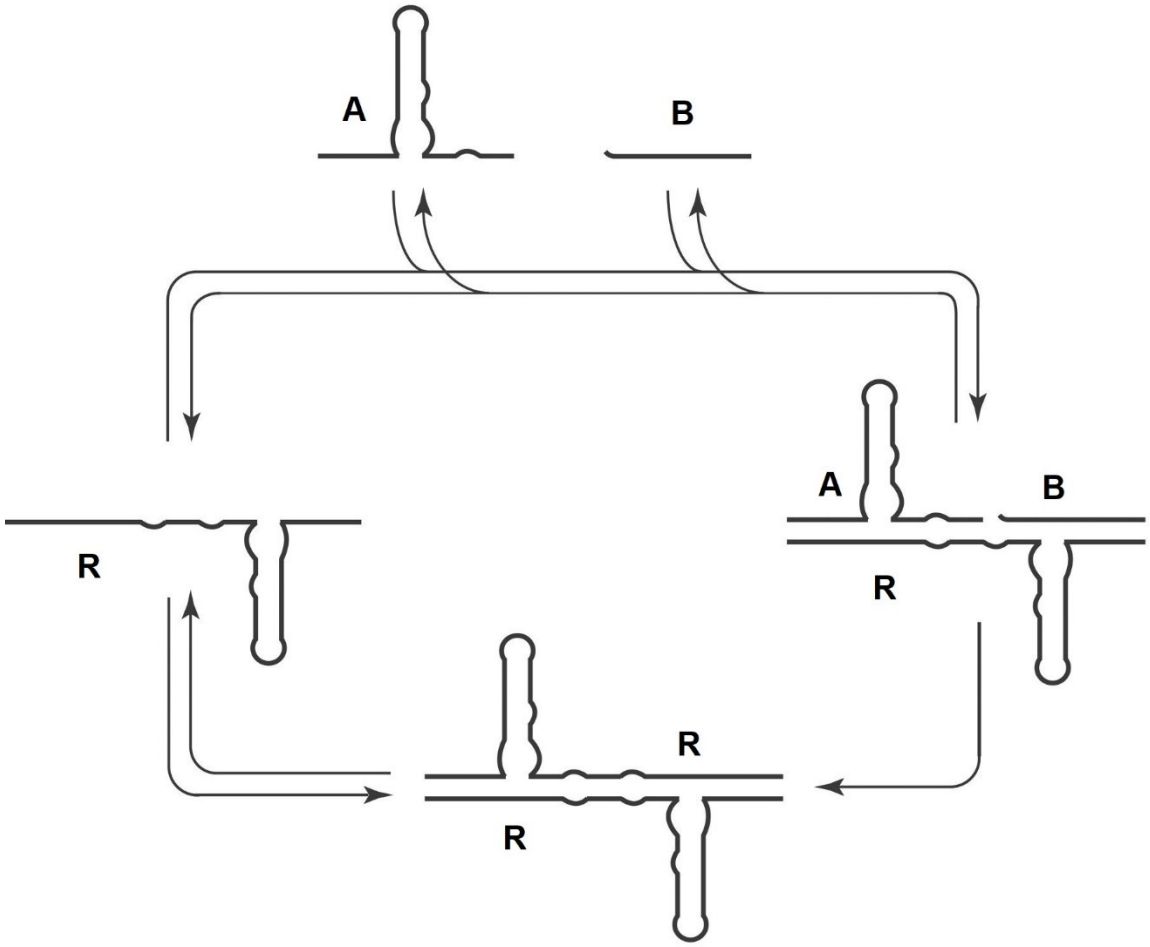
Abbildung 5. Tautomerie zwischen Cyanamid und Carbodiimid und auf Carbodiimiden basierende Aktivierungsreagenzien.

Ebenso wurde festgestellt, dass Cyanamid und ATP zur Peptidkopplung[125] oder Dicyclohexylcarbodiimid (DCC) und AMP zur Synthese von Aminoacyl-Adenylaten verwendet werden können.[126] Betrachtet man die chemische Vielfalt[127] und die vergleichsweise unspezifischen Bedingungen bei der abiotischen Kondensation, ist die alleinige Fokussierung auf die heute in biotischen Systemen genutzten Bausteine und Reaktionsmechanismen vermutlich eine zu starke Vereinfachung. So kommt die Frage auf, ob in der RNA-Welt-Theorie, die im Wesentlichen von den kanonischen Nukleotiden, d.h. Adenosin, Uridin, Guanosin und Cytidin ausgeht, wichtige evolutionäre Vorteile übersehen werden, die die Entstehung des Lebens erst ermöglicht haben.[128][129]

1.1.4 Selbstreplizierende Ribozyme

In der RNA-Welt-Hypothese wird davon ausgegangen, dass die RNA als zentrales Molekül bei der Entstehung des Lebens fungierte, da sie sowohl Informationen speichern, als auch Reaktionen katalysieren kann.[18][19][20][130] Auf Grund der Eigenschaft von Nukleinsäuren komplexe tertiäre Strukturen ausbilden zu können, können sie in Reaktionen durch Stabilisierung von Übergangszuständen oder durch Förderung einer bestimmten Substratorientierung als Katalysator wirken. Solche katalytisch aktiven RNA-Moleküle werden auch als Ribozyme bezeichnet, eine Wortbildung aus Ribose und Enzym. In einem Evolutionsschritt, nach der Bildung der monomeren Bausteine, die dann chemisch polymerisieren konnten, stellt sich die Frage, ob RNA seine eigene Replikation katalysieren kann. Ein solcher Schritt hin zu selbst-replizierenden Systemen, die sich aufgrund ihrer katalytischen Eigenschaften anreichern können, ist eine Voraussetzung der molekularen Evolution.

Bartel und *Szostak* konnten durch *in vitro* Evolution zeigen, dass aus einer Mischung zufälliger RNA-Sequenzen funktionsfähige Ribozyme erzeugt werden können. So konnten künstliche Ligase-Ribozyme isoliert werden, die einen ersten Schritt hin zu einer selbst-replizierenden RNA darstellen können.[131][132] Forscher um *Gerald Joyce* untersuchten das R3C-Ligase Ribozym. Dieses ist eine weitere Form eines selbst-replizierenden Ribozyms, das zwei Substrate verbindet, um eine exakte Kopie des Ribozyms zu erzeugen.[133] In diesem Fall wurde eines der beiden Substrate als 5′-Triphosphat verwendet. Ein Hauptproblem dieses selbst-replizierenden Ribozyms war jedoch die Produktinhibierung, aufgrund der thermodynamisch höheren Stabilität der Hybridisierung des verknüpften Produkts **R** im Vergleich zu den beiden Substraten **A** und **B** (**Schema *2***).



Schema 2. Selbst-replizierendes Ribozym. Das Ligase-Ribozym R3C-Ligase **R** katalysiert die Verknüpfung der beiden Fragmente **A** und **B** wodurch eine Kopie des Ribozyms **R** entsteht. Aus N. Paul et al. *Proc. Natl Acad. Sci.* **2002**, *99*, 12733-12740 und weiter bearbeitet.[133]

Diese Produktinhibierung konnte jedoch in einem kreuzkatalytischem System, bestehend aus 4 Fragmenten, umgangen werden[134] und zudem die Reaktionsraten gesteigert werden.[135][136] Außerdem können Schwankungen der Umgebungsbedingungen zu einer Dissoziation der komplementären Oligonukleotide führen. Nicht nur die Temperatur kann die Hybridisierung der Stränge beeinflussen, sondern auch pH-Schwankungen können die Dissoziation des Produktstrangs begünstigen.[137] Neben den R3C Ligasen, wurden auch modifizierte Varianten des *Azoarcus Self-Splicing Introns* (SSI) untersucht, wobei ebenfalls kreuzkatalytische Netzwerke entdeckt wurden, in denen sich das Ribozym aus 3 bis 5 Fragmenten selbst zusammensetzt.[138][139][140] Somit konnte mehrfach gezeigt werden, dass sich Ribozyme mit Ligaseaktivität selbst replizieren und eine höhere Komplexität entwickeln können. Ebenso konnte demonstriert werden, dass sich solche Ribozyme auch ursprünglich über nicht-enzymatische Ligation zusammensetzen konnten, was die beiden zuvor beschriebenen Ansätze der chemischen und der ribozymatischen Replikation mit einander verbindet.[141][142]

In den letzten Jahrzehnten wurden zahlreiche Meilensteine erreicht, die zum Verständnis der Entstehung des Lebens beigetragen haben. Diese reichen von der plausiblen Rekonstruktion der abiotischen Synthese der monomeren Bausteine des Lebens, bis hin zur Entdeckung der ersten selbst-replizierenden Systeme. Die RNA ist dabei von zentraler Bedeutung, da sie Genotyp und Phänotyp in einem einzigen Molekül vereint. Die RNA-Welt-Hypothese wird auch dadurch gestützt, dass die RNA in der heutigen Biologie als Katalysator bei der Synthese von Proteinen wirkt. Das Ribosom selbst ist also ein Ribozym und könnte Aufschluss darüber geben, wie der Übergang von einer reinen RNA-Welt zur Interaktion von RNA und Proteinen, wie wir sie heute bei der Translation beobachten, stattgefunden haben könnte.[143][144]

1.2 Die transfer-RNA und modifizierte Nukleoside

Bei der Translation wird die in Nukleinsäuren gespeicherte Information anhand des genetischen Codes in eine Abfolge von Aminosäuren übersetzt. Auf diese Weise wird der Genotyp in eine aktive Funktion, den Phänotyp, überführt und stellt somit die Schnittstelle zwischen der RNA-Welt-Hypothese und dem heutigen Leben dar.[143][144][145] Im Zentrum dieses Prozesses steht die katalytisch aktive Einheit der Peptidyltransferreaktion, das Ribosom.[146][147][148] Phylogenetische Studien haben gezeigt, dass ein bestimmter Satz an Genen notwendig ist, um eine funktionsfähige Zelle zu bilden. In dieser Studie zur Annäherung an den hypothetischen *last universal common ancestor* (LUCA)[7][8] wurden Gene identifiziert, welche hauptsächlich für die Translation und für grundlegende Aspekte der Transkription benötigt werden. Dies deutet darauf hin, dass die Evolution der Translation aus einer reinen RNA-Welt, wenn es denn jemals eine reine RNA-Welt gegeben hat, ein Schlüsselprozess bei der Entstehung des Lebens war.[10][11]

Die Aufklärung der molekularen Struktur des Ribosoms in den 2000er Jahren mit Hilfe der Röntgendiffraktometrie war einer der grundlegenden Durchbrüche, die zum Verständnis des Translationsmechanismus beigetragen haben.[149][150] Die Proteinbiosynthese basiert auf einer komplexen Maschinerie aus informationsenthaltender messenger-RNA (mRNA), Aminosäure-tragender transfer-RNA (tRNA) und katalytisch aktiver ribosomaler-RNA (rRNA). Das Ribosom selbst ist dabei eine hybride Struktur, die in Bezug auf die molekulare Masse zu zwei Dritteln aus RNA und zu einem Drittel aus verschiedenen Proteinen besteht, die sich auf zwei unterschiedlich große ribosomale Untereinheiten verteilen.[151] Diese Untereinheiten unterscheiden sich je nach Lebensform und werden durch den Sedimentationskoeffizienten *S* charakterisiert. Bei Prokaryoten setzt sich das 70S Ribosom beispielsweise aus der 30S Untereinheit und der 50S Untereinheit zusammen.[152][153] Die kleine Untereinheit dient der Interaktion der tRNA-Anticodonschleife mit dem mRNA-Codon, wodurch die Sequenz der

wachsenden Peptidkette festgelegt wird. In der großen Untereinheit findet die Peptidyltransferreaktion statt, bei der durch den nukleophilen Angriff des Amins der Aminoacyl-tRNA an der A-Stelle auf den C-Terminus der Peptidyl-tRNA an der P-Stelle eine Übertragung des wachsenden Peptids erfolgt, woraus eine Synthese vom *N*- zum *C*-Terminus des letztendlichen Proteins resultiert (**Abbildung 6**).[147][150]

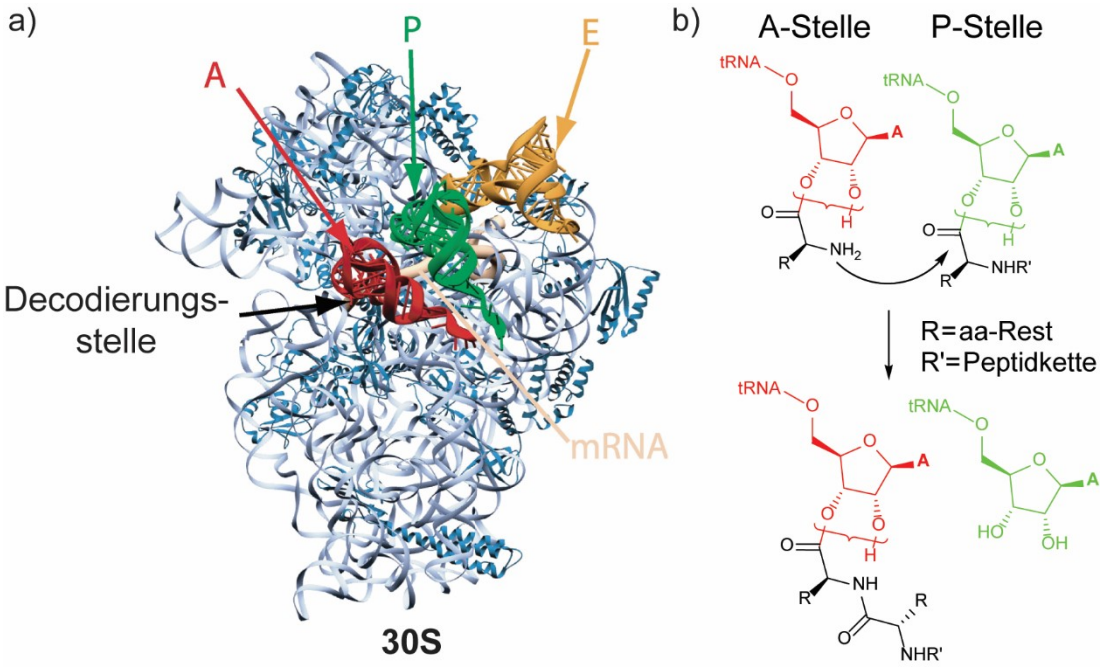
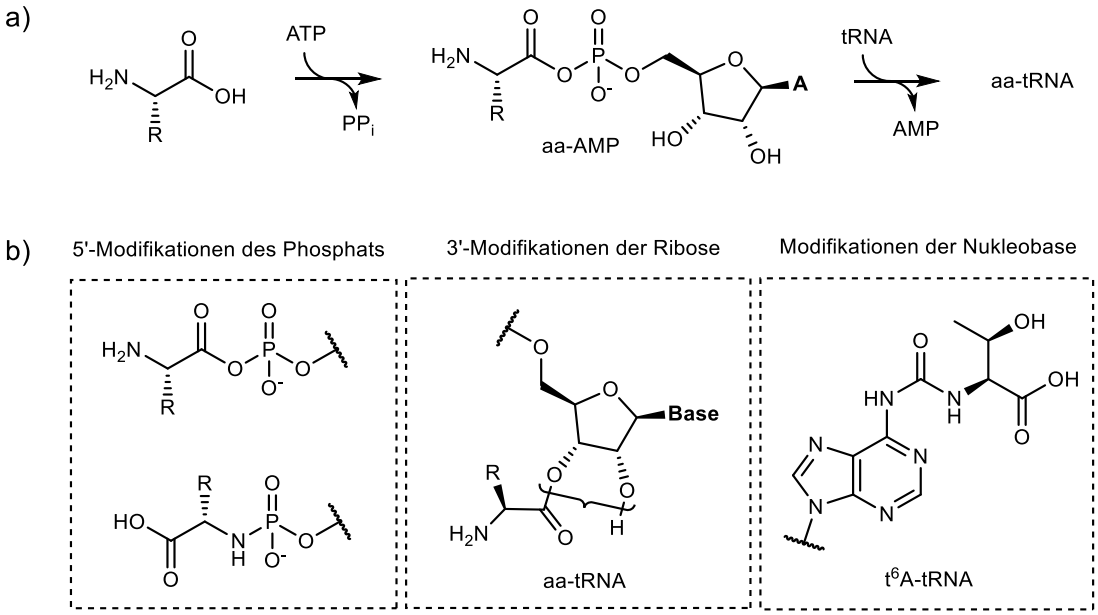


Abbildung 6. a) Kristallstruktur der 30S-Untereinheit des Ribosoms mit tRNAs gebunden an der Aminoacyl-Stelle (A-Stelle, rot), Peptidyl-Stelle (P-Stelle, grün) und Exit-Stelle (E-Stelle). Aus V. Ramakrishnan, *Cell* **2002**, *108*, 557-572 und weiterbearbeitet.[150] b) Peptidyltransfer zwischen zweier tRNAs in der A-Stelle (rot) und der P-Stelle (grün).

Bei der Suche nach dem Ursprung dieser Maschinerie stößt man unmittelbar auf das Dilemma, dass das Ribosom sowohl aus RNA als auch aus Proteinen besteht und daher auch Proteine benötigt werden, um Proteine synthetisieren zu können.[144][154][155][156] Es stellt sich die Frage ob es möglich ist, eine vereinfachte Form dieser Maschinerie zu rekonstruieren, die ausschließlich auf Ribonukleinsäuren basiert.[155][157][158] Die Entstehung der Proteinsynthese aus einer RNA-Welt heraus und damit die Einführung eines dualen Systems aus RNA und Peptiden wurde 1976 von *Francis Crick* als „notorisch schwieriges Problem“ (*notoriously difficult problem*) bezeichnet.[159][145] Das von ihm postulierte Modell einer uranfänglichen Translation geht von einem Codon-Pentett aus, das vollständig aus RNA besteht und ohne die Unterstützung von Proteinen auskommt. Stattdessen postulierte er, dass die Proto-tRNA eine Kavität besaß, in der die aktivierte Aminosäure koordiniert werden konnte. Somit wird vermutet, dass der Ursprung der codierten Proteinsynthese gänzlich ohne strukturgebende Proteine ablief, sondern lediglich mRNA und primitive Vorgänger der tRNA benötigt wurden.[159] Diese Hypothese wird durch die Entdeckung gestützt, dass das Ribosom seine katalytische Aktivität über die ribosomale RNA (rRNA) erhält. *Noller* und Mitarbeiter konnten

zeigen, dass die ribosomale RNA-Anteil des Ribosoms seine Aktivität als Peptidyltransferase (PTase) auch dann beibehält, wenn die ribosomalen Proteine der 50S Untereinheit entfernt werden, wenn auch mit deutlich reduzierter Aktivität.[160] Ebenso konnte gezeigt werden, dass die Abwesenheit von proteinbasierten Faktoren zwar die Effizienz der Translation reduziert, die Translation aber dennoch stattfinden kann.[161][162] Diese indirekten Nachweise, dass die rRNA als Katalysator wirkt, konnten später durch die Aufklärung der Kristallstruktur der 50S-Untereinheit des Ribosoms weiter bestätigt werden, wodurch das Ribosom selbst als Ribozym identifiziert werden konnte.[149][150]

Auf der Suche nach einem Vorgänger der heutigen tRNA, einer Proto-tRNA, stellt sich die Frage, wie diese aufgebaut war und wie die Aminosäuren daran gebunden waren.[163][164][165] In den heutigen tRNAs sind die Aminosäuren am 3′-Ende als Aminoacylester mit der Ribose verknüpft.[166] Die Aminosäuren werden in Form ihres Aminoacyl-Adenylats (aa-AMP) aktiviert[167] und mit Hilfe von Aminoacyl-tRNA-Synthetasen (aaRS) spezifisch an die entsprechende tRNA gebunden (**Schema *3a***).[168][169][170][171] Diese Strukturen gelten als hochkonservierte Motive,[154][172] deren Ursprung jedoch eine der größten ungelösten Fragen der präbiotischen Chemie darstellt.[164][165] In Anbetracht der RNA-Welt-Theorie wird vermutet, dass die Funktion der proteinbasierten aaRS in einem frühen Stadium der Evolution durch Ribozyme ausgeführt wurde.[173][174]



Schema 3 a) Aktivierung der Aminosäuren als Adenylat (aa-AMP) und anschließender Transfer auf das 3′-Ende der entsprechenden tRNA. b) Chemische Strukturen Aminosäure-modifizierter Nukleinsäuren am 5′-Phosphat, an der 2′-3′-Position der Ribose oder an der Nukleobase.

Vor diesem biochemischen Hintergrund wurden zahlreiche Studien durchgeführt, in denen Adenylate und Ester für den RNA-basierten Transfer von Aminosäuren untersucht wurden (siehe **Kapitel 1.4**). Diese Verbindungen erweisen sich jedoch als hydrolytisch instabil,

insbesondere in Umgebungen, die stark von den physiologischen Bedingungen abweichen.[175][176][177] Daraus resultiert eine kurze Halbwertszeit in wässriger Lösung, was RNA-Aminosäure-Konjugate, wie sie in der Translation verwendet werden, in einem frühen Stadium der Evolution als unwahrscheinlich macht.[178] Das 3′-Ende ist jedoch nicht die einzige Position in der tRNA, die mit Aminosäuren modifiziert sein kann. In ähnlicher Weise befinden sich in einigen modernen tRNAs in der Nähe der Anticodonschleife nicht-kanonische Nukleotide, die Aminosäuren über eine Harnstoffbrücke an der Nukleobase tragen können, die vergleichsweise stabiler gegenüber Hydrolyse ist (**Schema *3*b**).[179][180]

Generell ist die tRNA reich an nicht-kanonischen Nukleosiden und darunter insbesondere an Modifikationen der Nukleobase.[181][182] tRNAs sind ein zentrales Molekül in der Proteinbiosynthese und gehören zu den nicht-codierenden RNAs (non-coding RNA), die Substrate für zahlreiche post-transkriptionelle Modifikationen sind.[183] Nicht-kanonische Nukleoside spielen neben der Codierung auch eine wichtige Rolle bei der Stabilität und der Ausprägung von Sekundär- oder Tertiärstrukturen und spiegeln die Vielfalt der Möglichkeiten wider, wie RNA ausgeprägt sein kann.[184][185] Insbesondere innerhalb der Anticodonschleife befinden sich diverse Aminosäuremodifikationen.[186] Das N^6-exozyklische Amin von Adenosin und die 5-Position von Uridin können dabei vielseitig modifiziert sein.[186] Die Position 37 der tRNA, die in 3′-Richtung neben der 3. Base in der Anticodonschleife liegt, ist in >70% aller tRNAs modifiziert.[187] Eine der möglichen Modifikationen enthält die Aminosäure Threonin in Form von N^6-Threonylcarbamoyl-Adenosin (t^6A).[188] Neben t^6A wurden auch weitere N^6-Carbamoyl-Adenosin-Derivate entdeckt, die über eine Harnstoffbrücke gebundene Aminosäuren tragen, wie z.B. N^6-Hydroxynorvalincarbamoyl-Adenosin (hn^6A),[189] Hydroxythreonylcarbamoyl-Adenosin (ht^6A),[190] N^6-Glycinylcarbamoyl-Adenosin (g^6A)[191] oder auch Derivate an der Nukleobase von t^6A, wie das zyklische ct^6A[192] oder das methylierte m^6t^6A.[193] Ob ct^6A letztendlich in der Hydantoin- oder der Oxazolon-Isoform vorliegt, konnte bisher nicht eindeutig geklärt werden (**Abbildung 7**).[192][194]

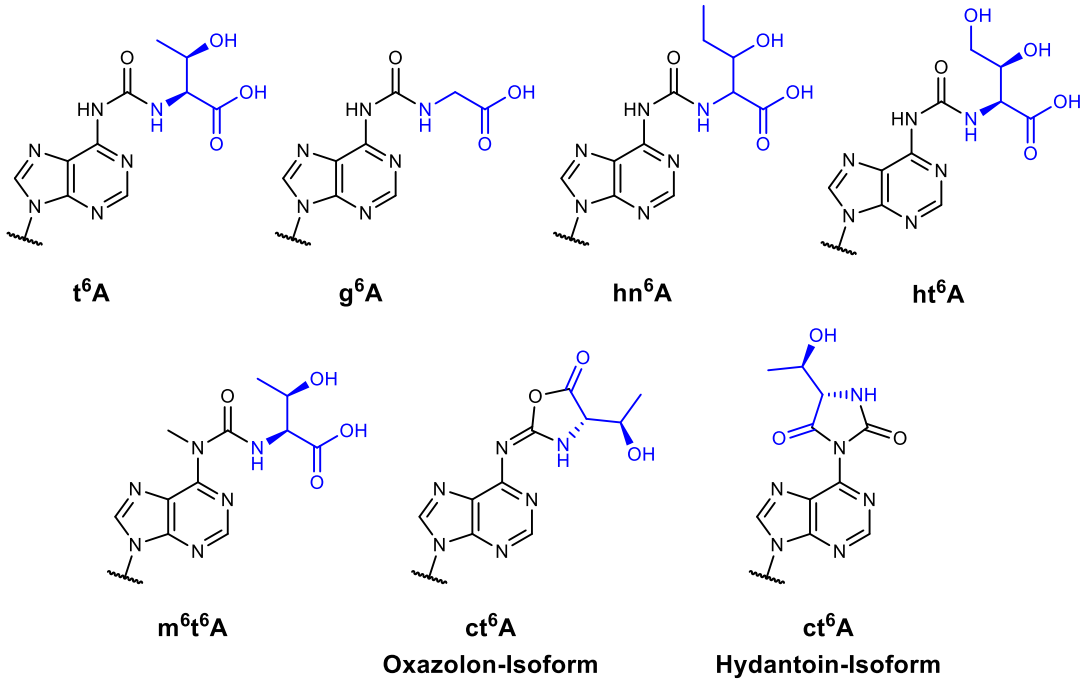


Abbildung 7. Strukturen der Nukleobasen diverser Aminosäure-modifizierter *N*^{6}-Carbamoyl-Adenosin-Derivate.

N^6-Threonylcarbamoyl-Adenosin (t^{6}A) kann ausschließlich in Position 37 der tRNA gefunden werden und tritt in allen tRNAs auf, die entsprechende ANN-Codons lesen.[187] Diese Modifikation dient der korrekten Ausbildung der Anticodonschleife. Die Harnstoffeinheit bildet dabei eine planare Struktur zum Adenin-Ring, welche über eine Wasserstoffbrücke zwischen der *N*^{1}-Position des aromatischen Rings und der NH-Gruppe der Aminosäure stabilisiert wird. Dadurch wird die Fläche für Basenstapelung erweitert. Die Threonyl-Einheit hingegen kann frei um die NH-Gruppe und den *α*-Kohlenstoff der Aminosäure rotieren. Auf Grund des sterischen Anspruchs der Threonyl-Modifikation findet keine Basenstapelung zwischen der benachbarten Base U36 statt, sondern mit der Nukleobase A1 des Codons, wodurch eine erhöhte Stabilität der Codon-Anticodon-Interaktion erreicht wird. Außerdem verhindert die t^{6}A-Modifikation die Bildung einer Basenpaarung zwischen A37:U33, was für eine fehlerfreie Ausprägung der Sekundärstruktur der Anticodonschleife essentiell ist (**Abbildung *8*a und b**).[185][195]

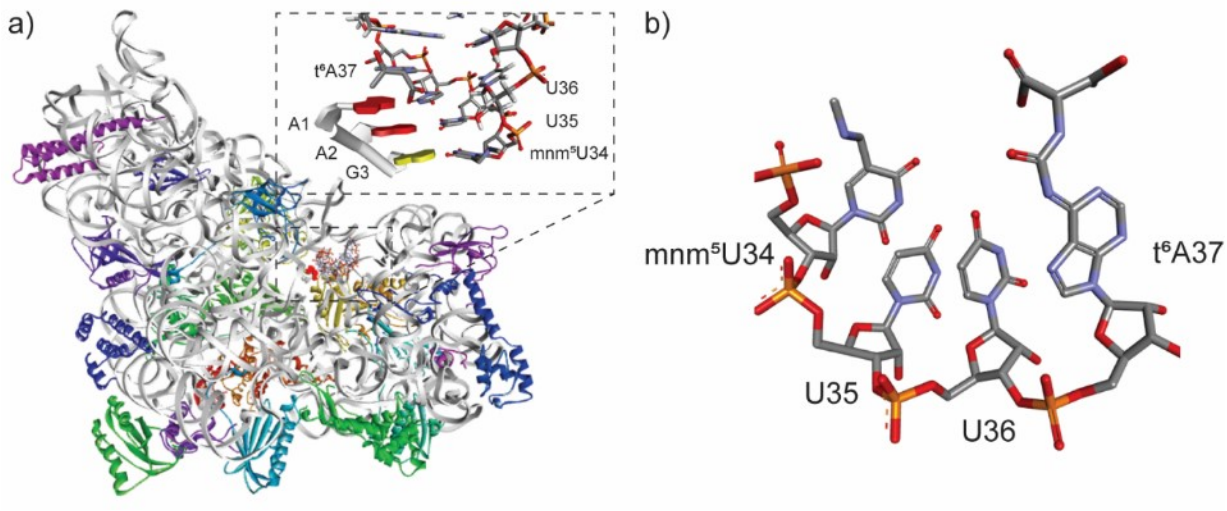
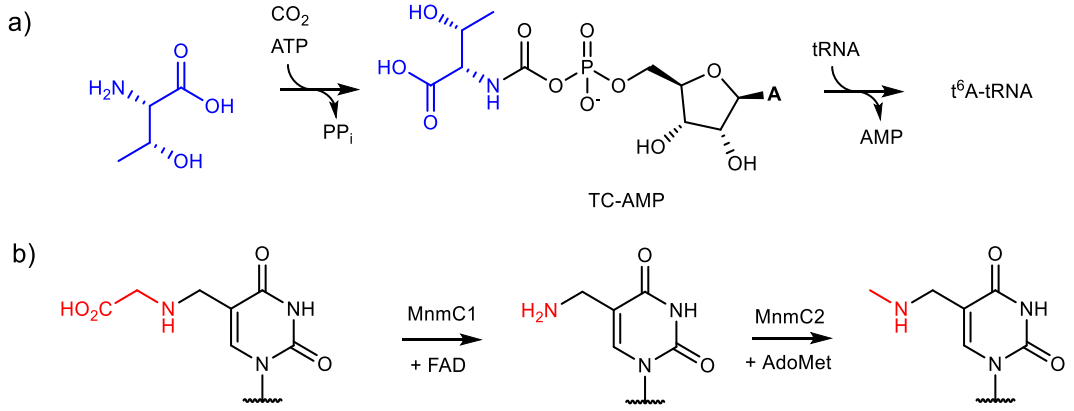


Abbildung 8. a) Kristallstruktur der Anticodonschleife der modifizierten $tRNA^{Lys}_{mnm^5UUU}$ gebunden an die ribosomale 30S Untereinheit in der A-(Aminoacyl-)Stelle. In der Vergrößerung sind die *Watson-Crick* Basenpaarungen A:U und die *Wobble* Basenpaarung G:mnm^{5}U zwischen der Anticodonschleife und des AAG Lysin-Codons dargestellt. b) Ausschnitt der Anticodonschleife mit den natürlichen Modifikationen mnm^{5}U34 und t^{6}A37. Die Kristallstrukturen wurden dem PDB-Eintrag 1XMO[185] entnommen und mit *Discovery Studio Visualizer v21.1.0* bearbeitet.

Diese *N*^{6}-Carbamoyl-Modifikation wird post-transkriptionell durch spezifische Enzyme eingefügt, deren Untereinheiten im Fall des *kinase, endopeptidase and other proteins of small size* (KEOPS) Komplexes über Archaeen und Eukaryoten hinweg konserviert sind.[196] Die Kae1-Untereinheit, welche die Bildung von t^{6}A katalysiert, kann sogar in allen drei Domänen des Lebens gefunden werden.[10] In der Synthese von t^{6}A wird ausgehend von L-Threonin, Bicarbonat/CO_2 und ATP das Intermediat Threonylcarbamoyl-Adenosinmonophosphat (TC-AMP) hergestellt, welches anschließend unter der Abspaltung von Adenosinmonophosphat (AMP) die tRNA mit t^{6}A modifiziert. (**Schema *4a***).[197][198]

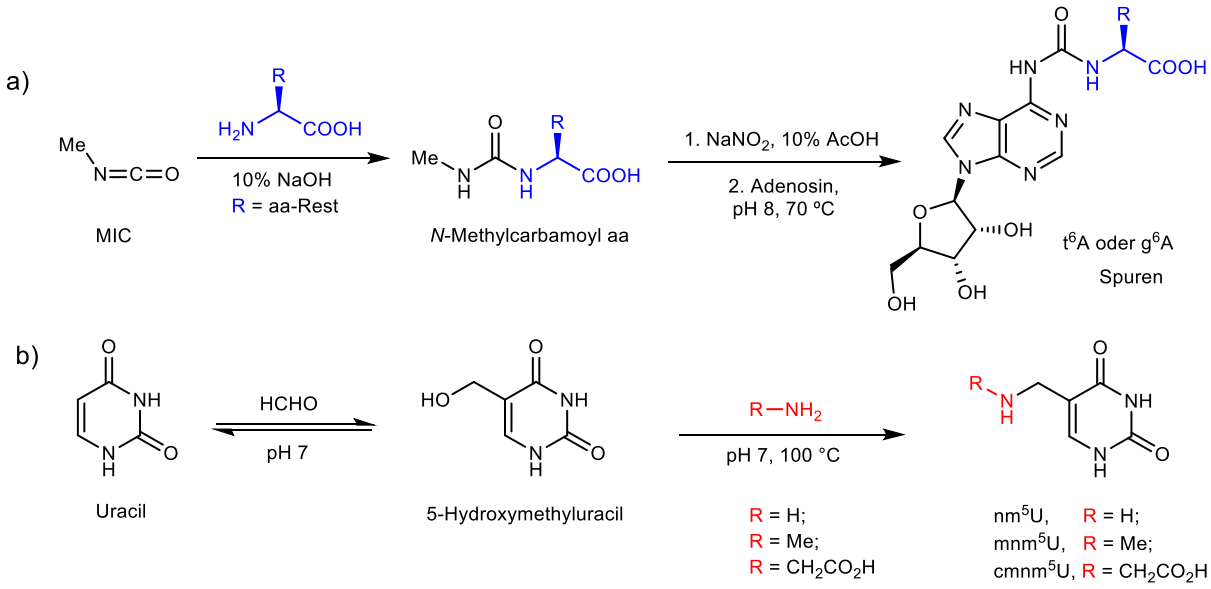


Schema 4. Enzymatische Synthese der tRNA-Modifikationen t^{6}A und mnm^{5}U ausgehend von L-Threonin bzw. 5-Carboxymethylaminomethyl-Uridin.[197][198][199]

Eine weitere Stelle, sogar direkt innerhalb des Anticodons, die mit Aminosäuren modifiziert sein kann, ist die *Wobble* Position 34, in der die folgenden Aminosäuremodifikationen

vorkommen können: Glutaminsäure als Glutamyl-Queuosin (GluQ)[200][201], Lysin als 2-Lysidin (k^2C),[202] und Glycin als 5-Carboxymethylaminomethyl-Uridin ($cmnm^5U$).[199] Diese Modifikationen spielen eine grundlegende Rolle bei der Decodierung der NNG/A Codons durch Stabilisierung der U:G *Wobble*-Basenpaarung (**Abbildung *8*a und b**).[203] Letztere Modifikation, $cmnm^5U$, wurde in *Escherichia coli* in der $tRNA^{Leu}$ nachgewiesen. Ebenso kann $cmnm^5U$ durch das bifunktionelle Protein MnmC über die FAD-abhängige Oxidasedomäne MnmC1 in 5-Aminomethyl-Uridin (nm^5U) und anschließend über die *S*-Adenosylmethionin-abhängige Domäne MnmC2 in das methylierte Derivat 5-Methylaminomethyl-Uridin (mnm^5U) umgewandelt werden (**Schema *4*b**).[199]

Aufgrund der Komplexität des gesamten Translationsprozesses ist es schwierig den Ursprung der codierten Peptidsynthese aus einer RNA-Welt heraus zu rekonstruieren. *Harry Noller* postulierte, dass in einer vereinfachten Proto-Translation im Laufe der Evolution ursprünglich nicht direkt funktionelle Proteine entstanden, sondern zufällige Peptide, die an die RNA kovalent gebunden waren und damit die chemische Diversität erhöhten.[145] In der Folge könnte dies zu einer erhöhten Stabilität[204][205][204] oder Funktionalität der RNA geführt haben.[128][206][207] Es wird vermutet, dass solche Modifikationen der Nukleobase bereits Teil eines ursprünglichen genetischen Codes gewesen sein könnten, in dem Aminosäuren in der Anticodon-Region der Proto-tRNA vorkamen und nun als mögliche molekulare Fossilien beobachtet werden.[208][209]



Schema 5. Präbiotische Routen zu a) Aminosäure-modifiziertem N^6-Carbamoyl-Adenosin und b) 5-modifiziertem Uracil.[210][211]

Diese Hypothese kann durch den Nachweis gestützt werden, dass einige dieser Modifikationen auch in präbiotisch plausibler Weise entstehen können. Es wurde gezeigt, dass sowohl g^6A als auch t^6A unter präbiotischen Reaktionsbedingungen erhalten werden können (**Schema *5*a**).[210] Aminosäuren können durch Reaktion mit Methylisocyanat (MIC) in einer

Formamidlösung zu den entsprechenden *N*-Methylcarbamoyl-Aminosäuren umgesetzt werden, die dann unter nitrosierenden Bedingungen mit dem exozyklischen Amin des Adenosins zu t^6A bzw. g^6A reagieren können. Neben *O*-Carbamoyl-modifizierten Nukleosiden wurde zudem auch die Bildung methylierter Spezies beobachtet, da bei der Bildung des *N*-Isocyanats der Aminosäure Diazomethan abgespalten wird, was einen möglichen Weg zu methylierten Nukleosiden darstellt. Auch die Nukleobasen der nicht-kanonischen Nukleoside mnm^5U, nm^5U und $cmnm^5U$ konnten unter präbiotischen Bedingungen erhalten werden. Die Reaktion wurde von *Miller* und Mitarbeitern ausschließlich mit Uracil untersucht, da die Addition von Formaldehyd im ersten Schritt eine höhere Reaktionsrate mit der Nukleobase im Vergleich zum Nukleosid aufweist. Dennoch ist es aus präbiotischen Gesichtspunkten ebenso plausibel, dass 5-Hydroxymethyl-Uridin aus dem entsprechenden unmodifizierten Nukleosid Uridin durch Addition von Formaldehyd entstehen kann. Im zweiten Schritt wurde von *Miller* und Mitarbeitern gezeigt, dass 5-Hydroxymethyl-Uracil mit verschiedenen Nukleophilen, u.a. Methylamin, Ammoniak oder Glycin, substituiert werden kann, wodurch die Nukleobasen-Modifikationen nm^5U, mnm^5U und $cmnm^5U$ erhalten wurden (**Schema *5*b**).[211]

1.3 Selbst-aminoacylierende Ribozyme

Die selektive Beladung von tRNAs oder analoger RNA-Sequenzen ist nicht nur im Zusammenhang mit der Entstehung des Lebens von Bedeutung,[212] sondern auch in der Biotechnologie, um den genetischen Code zu erweitern und künstliche Proteine mit modifizierten Eigenschaften zu erhalten.[170][213] Im Kontext der präbiotischen Chemie gilt die Darstellung aminoacylierter RNA als ein weiteres Bindeglied, um den Übergang von einer RNA-Welt zu den Ursprüngen der Translation zu rekonstruieren.[145][214] Dieser Dualismus aus RNA und Aminosäuren konnte erstmals in der nicht selektiven Aminoacylierung von RNA beobachtet werden. Dabei wurde die Aminosäure in einer Imidazol-katalysierten Reaktion von einem Aminoacyl-Adenylat auf eine der Hydroxygruppen der Ribose eines Uridin- oder Adenosin-Oligomers übertragen.[215][216] Sowohl aus präparativer, als auch biotechnologischer Perspektive haben Forschungsgruppen an der *in vitro* Evolution von Ribozymen gearbeitet, die selektiv RNA aminoacylieren können.[174][217] Dabei konnte gezeigt werden, dass Ribozyme die Eigenschaft haben können, RNA am 3′-Ende mit Aminosäuren zu modifizieren. Eine der bahnbrechenden Arbeiten in diesem Kontext wurde 1995 von *Yarus* und Mitarbeitern veröffentlicht, in der ein Ribozym aus nur 5 Nukleotiden, die Aminoacylierung an der 2′3′-Position eines 4-mers mit einem 3′-terminalen Uridin katalysierte (**Abbildung *9*a**).[218] Es konnte gezeigt werden, dass lediglich drei konservierte Nukleotide für die katalytische Aktivität notwendig waren.[219] Zwei Nukleotide des Ribozyms und ein Nukleotid des Substrats bilden

eine katalytisch aktive Stelle, wodurch die als Adenylat aktivierte Aminosäure (Phe-AMP oder Tyr-AMP) auf das Substrat übertragen werden kann.[220][217] Interessanterweise katalysierte das von *Yarus* und Mitarbeitern beschriebene Ribozym sowohl die Aminoacylierung von RNA als auch eine sukzessive Peptidkopplung, wodurch kurze Peptid-Oligomere erhalten wurden.[178][220][221] Diese geringe Spezifität spiegelt die Vielseitigkeit primitiver Ribozyme wider, die ein wichtiger Motor der Evolution gewesen sein kann.[222][223][173]

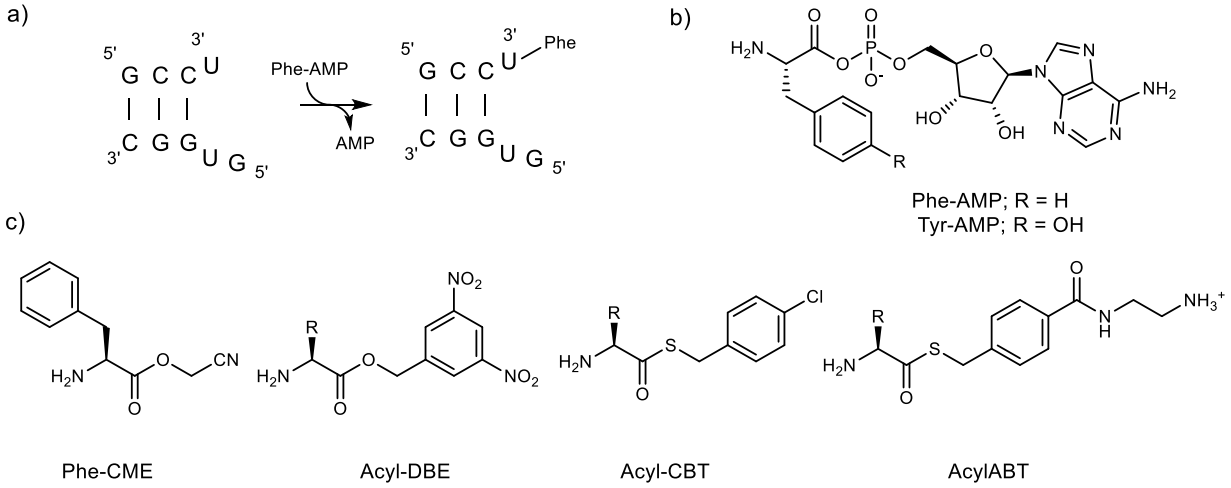


Abbildung 9. a) Selbst-aminoacylierendes Ribozym nach *Yarus* et al. b) Aminosäure-Adenylate, welche als Substrat für das selbst-aminoacylierende Ribozym verwendet wurden c) Aktivierte Substrate für diverse Flexizyme.

Die Beschränkung auf Substrate mit der Sequenz 3′-CCU und die Verwendung ausschließlich aromatischer Aminosäuren, wie z.B. Phenylalanin-Adenosinmonophosphat (Phe-AMP) stellt ein Hindernis für die Anwendung im Bereich der Erweiterung des genetischen Codes dar. Mit dem Ziel das 3′-CCA-Ende der tRNA zu modifizieren und dabei verschiedene, auch nicht proteinogene Aminosäuren verwenden zu können, entwickelte die Forschungsgruppe um *Hiroaki Suga* durch schrittweise *in vitro* Evolution eine neue Klasse aminoacylierender Ribozyme: Die Flexizyme, deren Name sich von der Flexibilität der Substratspezifität ableitet.[170] Dabei wurden zu Beginn bevorzugt aktivierte aromatische Aminosäuren, d.h. Phenylalanin, verknüpft und selektiv an die 3′-OH Gruppe des terminalen Adenosins gekoppelt.[224] Dazu wurde die tRNA mit einer randomisierten Bibliothek von RNA-Sequenzen am 5′-Ende verlängert und nach der Fähigkeit das 3′-Ende zu aminoacylieren selektiert. Nach weiterer Verfeinerung des Designs wurde eine höhere Toleranz gegenüber der Verwendung unterschiedlicher tRNAs erreicht, da die Interaktionen zwischen Ribozym und tRNA lediglich auf drei Basenpaarungen mit dem 3′-CCA-Ende der tRNA beruhten, wodurch auch andere Sequenzen mit dem CCA-Motiv aminoacyliert werden konnten.[225] Als aktivierte Aminosäuren wurden neben Adenylaten[218][220][226][221] auch Cyanomethylester (CME)[227][228] oder Thioester (CBT, ABT)[229] des Phenylalanins verwendet (**Abbildung 9b und c**). Kristallographische Untersuchungen am Flexizyme Fx3 zeigten, dass das Ribozym eine Erkennungsstelle für den

aromatischen Ring der Aminosäure besitzt.[230] Durch weitere Mutationen konnte die Spezifität für die aromatische Seitenkette umgangen und das aromatische Erkennungsmotiv in die Abgangsgruppe der aktivierten Aminosäure verschoben werden.[231] So konnte die Bandbreite an Substraten um nicht natürlich vorkommende Aminosäuren, d.h. D-Aminosäuren,[232] *N*-Alkylaminosäuren[233] und α-Hydroxysäuren[234] oder sogar Oligopeptide[235] erweitert werden. Trotz der beeindruckenden Verbesserung der aminoacylierenden Eigenschaften der Flexizyme und der Erweiterung um mögliche Substrate und um das CCA-Motiv bleibt die Bedeutung aminoacylierender Ribozyme als Vorläufer von aaRS fraglich, da bisher keine derartigen Ribozyme in der Natur identifiziert werden konnten und die Erkennung der Substrate über aromatische Motive, wie Dinitrobenzylester (DBE), präbiotisch fraglich erscheint.[174][236] Nichtsdestotrotz könnten aminoacylierende Ribozyme das Bindeglied zwischen einer reinen RNA-Welt hin zu einer RNA-Peptid-Welt gewesen sein, deren Aufgabe im Verlauf der Evolution durch effizientere Proteine übernommen wurde. Insbesondere dann, wenn präbiotisch plausible Aminosäure-Adenylate als Substrate für eine Aminoacylierung dienten.

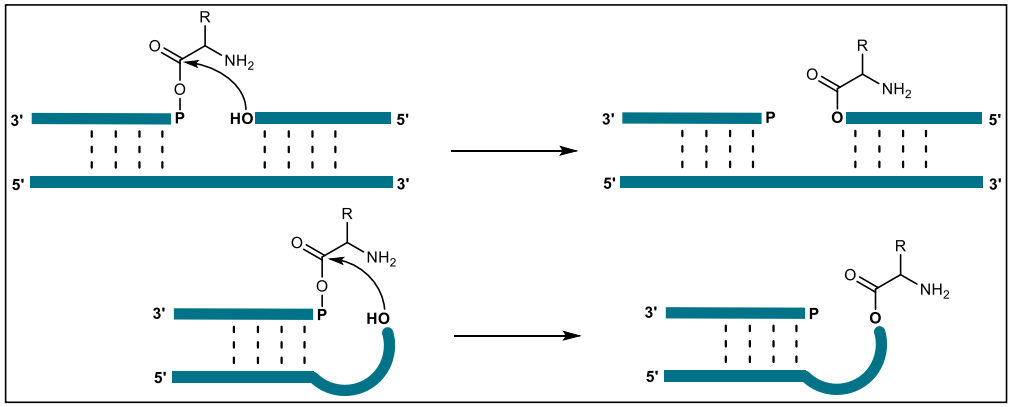
1.4 RNA-basierter Aminosäuretransfer

Der genetische Code besteht aus 20 proteinogenen Aminosäuren, die entweder vom Organismus aufgenommen (essentielle Aminosäuren) oder über biochemische Stoffwechselwege in der Zelle synthetisiert werden. Es ist jedoch davon auszugehen, dass die ersten Aminosäuren auf abiotischem Wege entstanden sind.[237][238] Im Gasentladungsexperiment nach *Miller*[38][239] konnten 10 der derzeit 20 proteinogenen Aminosäuren identifiziert werden.[240] Eine ähnliche Anzahl von Aminosäuren wurde auch in Meteoriten nachgewiesen.[241][242] Dies lässt auf einen abiotischen Ursprung schließen und diese sowohl experimentell synthetisierten als auch aus extraterrestrischem Material isolierten Aminosäuren können neben zahlreicher nicht-proteinogener Aminosäuren als präbiotisch plausibel angesehen werden. Wie jedoch die jeweiligen Aminosäuren von einer nicht-codierten Peptidsynthese[128][206][207] hin zu einer sequenz-codierten Synthese zugeordnet wurden, ist nach wie vor eine der großen offenen Fragen in der Erforschung des Ursprungs des Lebens. Der Schlüssel zum Verständnis dieses Ursprungs ist der codierte, RNA-basierte Aminosäuretransfer. Zahlreiche Forscher haben sich dementsprechend mit verschiedenen RNA-Aminosäurekonjugaten in Form von Aminoacyl-Phosphaten, Phosphoramidaten und Estern befasst.[243]

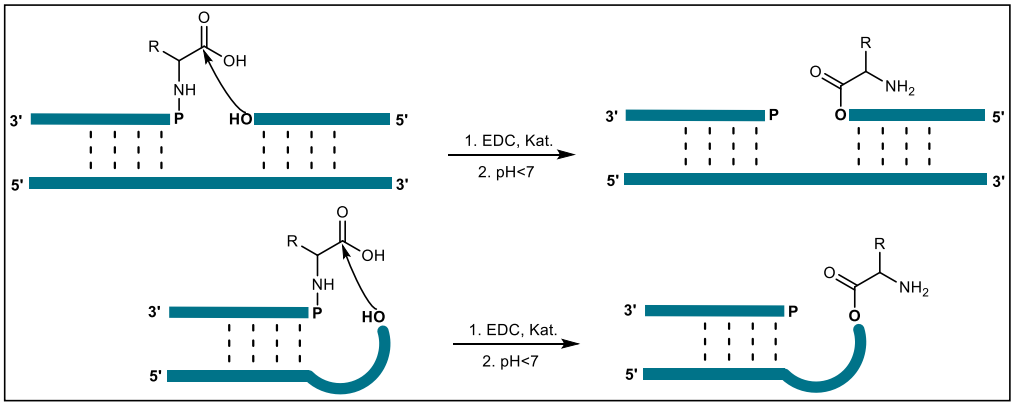
Ähnlich zu der Beladung von tRNAs mit Aminosäuren durch aaRS und der ribozymatischen Beladung von RNA durch Aminosäure-Adenylate nach *Yarus* et al., fanden *Tamura* und

Mitarbeiter heraus, dass Aminosäuren, die als gemischtes Anhydrid an das 5′-Phosphat einer Donorsequenz gebunden sind, mit dem 3′-Ende eines benachbarten Oligonukleotids reagieren können. In diesem Fall waren Donor- und Akzeptorsequenz über einen komplementären Strang miteinander verbunden (**Schema *6*a**).[175]

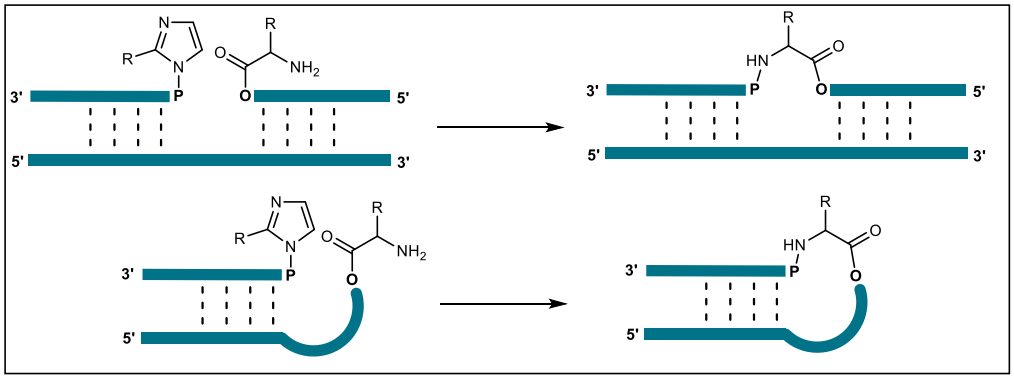
a) **5'-Aminoacyl-Phosphat und 3'-Hydroxygruppe**



b) **5'-Phosphoramidat und 3'-Hydroxygruppe**



c) **5'-Phosphorimidazolid und 3'-Aminoacylester**



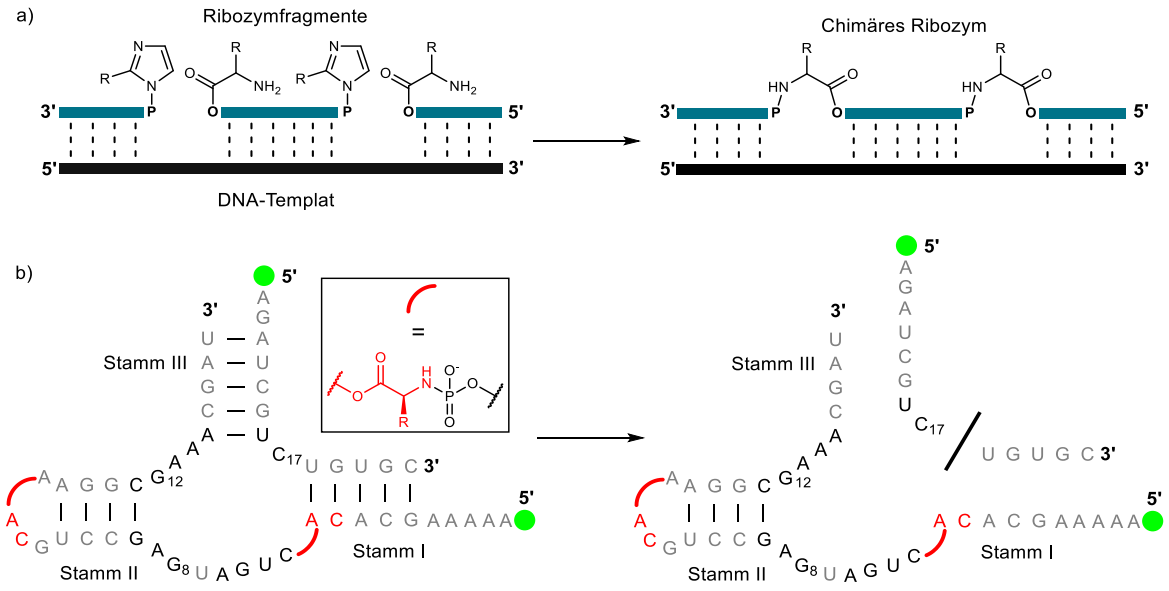
Schema 6. Präbiotische Untersuchungen zum RNA-basierten Aminosäure-Transfer und Ligation von RNA-Fragmenten durch Phosphoramidate und Ester aus Aminosäuren; entweder zwischen zweien Sequenzen, die über eine Brückensequenz verknüpft sind oder zwischen zweien komplementären Sequenzen mit 3′-Überhang. a) Aminosäure-Transfer über 5′-Aminoacyl-Phosphate. b) Aminosäure-Transfer über 5′-Phosphoramidate. c) Verknüpfung zweier RNA-Fragmente über 5′-Phosphorimidazolide und 3′-Aminoacylester.[243]

In ähnlicher Weise konnten *Sutherland* und Mitarbeiter zeigen, dass der Transfer auch zwischen komplementären Strängen möglich ist. Anstelle der verbrückenden Sequenz wurde ein Überhang von 4-5 Nukleotiden verwendet und in Abhängigkeit von der Nukleotidsequenz konnte ein Aminosäuretransfer in Ausbeuten von bis zu 55% beobachtet werden (**Schema *6*a**).[176] Es gibt auch erste Hinweise darauf, dass die Chiralität der Aminosäure einen Einfluss auf den Transfer zur RNA hat, da die Übertragung von L- gegenüber D-Aminosäuren bevorzugt wurde.[176][244][245] Dies deutet darauf hin, dass die ebenfalls chirale D-Ribose der Nukleoside bzw. Nukleotide einen Einfluss auf die Entstehung der in der heutigen Biologie vorzufindenden Homochiralität der Aminosäuren gehabt haben könnte.[246]

Neben den gemischten Anhydriden, bei denen die Carbonsäure der Aminosäure an das Phosphat gebunden ist, konnten unter präbiotischen Bedingungen auch Phosphoramidate erhalten werden, in denen das Amin der Aminosäure mit dem Phosphat reagiert (siehe **Kapitel 1.1.3**). Nach der *in situ* Aktivierung der Carbonsäure mit wasserlöslichen Aktivatoren, wie z.B. 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimid (EDC) und einem Organokatalysator, z.B. Ethylimidazol[122], können anschließend Nukleophile, u.a. 3′-Hydroxy-[247] oder 3′-Aminogruppen,[248] mit der Carbonsäure der als Phosphoramidat gebundenen Aminosäure reagieren. Dadurch wird ein chimäres Produkt erhalten, in dem zwei RNA-Fragmente über eine Aminosäure miteinander verbunden sind. In einem zweiten Schritt kann die Aminosäure, nach der Hydrolyse des Phosphoramidats unter sauren Bedingungen, auf die 2′-3′-Hydroxygruppe des benachbarten RNA-Strangs übertragen werden (**Schema *6*b**). Ebenso konnte über die Phosphoramidatchemie eine nicht enzymatische Translation rekonstruiert werden, bei der anstelle eines Tripletts, ein einzelnes Nukleotid über Watson-Crick-Basenpaarung und π-π-Wechselwirkungen als Erkennungsmerkmal verwendet wird.[249]

In ähnlicher Weise kann die Verknüpfung zweier RNA-Fragmente auch durch die Transamidierung eines aktivierten 5′-Phosphorimidazolids mit dem Amin der Aminosäure erfolgen.[250] In diesem Fall entsteht ein ebenfalls ein chimäres Molekül aus Nukleotid und Aminosäure, in dem die Aminosäure als Ester und als Phosphoramidat kovalent an beide RNA-Stränge gebunden ist (**Schema *6*c**).[251] Solche chimären Moleküle können ein Beispiel für eine Co-Evolution von RNA und Peptiden sein, bei der die Wechselwirkungen beider Moleküle zu einer höheren Komplexität geführt haben und ein Motor der Evolution gewesen sein könnten.[207] In diesem Zusammenhang haben Studien zur chemischen Primer-Verlängerung aminoacylierter RNA gezeigt, dass 2′-3′-aminoacylierte RNA eine höhere Reaktionsrate mit einem aktivierten 5′-Phosphorimidazolid aufweist, als unmodifizierte 2′-3′-Hydroxy-RNA. Ebenso konnte gezeigt werden, dass diese RNA-Peptid-Chimären auch als Template für nicht-enzymatische Replikationen dienen können.[251] Neben dem Einfluss bei der Verknüpfung polymerer Strukturen wurde auch der Aufbau katalytisch aktiver RNA-Sequenzen untersucht. Dabei diente unter anderem das Hammerhead Ribozym als

Modellsystem, um die Aktivität des chimären Ribozyms mit der des kanonischen Ribozyms als Referenz zu vergleichen (**Schema 7a** und **b**).[252]



Schema 7. a) Chemische Synthese eines chimären I/III Hammerhead Ribozyms aus RNA-Fragmenten mit Hilfe eines DNA-Templats. b) Spaltungsreaktion des chimären I/III Hammerhead Ribozyms. Die konservierte Nukleotidsequenz ist in schwarz dargestellt und das aminoacylierte 3′-CA ist rot hervorgehoben. Grüner Kreis = FAM-Fluorophor.[252]

Die 2′-3′-aminoacylierten RNA-Fragmente wurden präparativ durch Flexizym-abhängige Aminoacylierung und die 5′-Phosphorimidazolid-modifizierten RNA-Stränge aus dem entsprechenden 5′-Phosphat und 2-Methylimidazol hergestellt.[253][254] Die anschließende Ligation der Fragmente setzte einen Templat-Strang voraus und ergab zwar das zusammengesetzte Produkt, das jedoch aufgrund der Hybridisierung zwischen Templat und Produkt nicht katalytisch aktiv war. Durch die Verwendung von DNA als Templat-Sequenz, welche anschließend enzymatisch verdaut wurde, oder alternativ mehrerer kürzerer RNA-Splints, konnte die Inhibierung umgangen werden. In Abhängigkeit von der verknüpfenden Aminosäure wurde eine 20- bis 80-fach reduzierte Aktivität des chimären Ribozyms im Vergleich zum unmodifizierten Ribozym beobachtet. Gly wurde als polare, Lys als kationische, Leu als unpolare und Asp als anionische Aminosäure untersucht, wobei Lys die höchste Aktivität zeigte. Eine äquimolare Mischung der nicht zusammengesetzten Fragmente hingegen wies keine messbare Aktivität auf.[252]

RNA-Aminosäure-Konjugate könnten also nicht nur Vorläufermoleküle für die Translation gewesen sein, sondern auch in einem frühen Stadium der Evolution als strukturgebende Motive gedient haben, die RNA-Fragmente verknüpften und Eigenschaften von Ribozymen verbesserten.

1.5 Selbstspaltende RNA – Das Hammerhead Ribozym

Selbstspaltende Ribozyme sind aufgrund ihrer einfachen Struktur von ca. 50 Nukleotiden und der guten Quantifizierbarkeit ihrer katalytischen Aktivität ein ausgezeichnetes Modell für die Untersuchung katalytischer Prozesse im präbiotischen Kontext.[252][255][256] Zu der Klasse der selbstspaltenden Ribozyme gehören das Hammerhead Ribozym, das Hepatitis Delta Virus (HDV) Ribozym, das *Hairpin* Ribozym, das *Neurospora Varkud Satellite* Ribozym, das bakterielle Glucosamin-6-Phosphat-Synthase (*GlmS*) Motiv Ribozym, das *Pistol* Ribozym, das *Twister* Ribozym, das *Twister Sister* Ribozym und das *Hatchet* Ribozym.[257] Selbstspaltende Ribozyme sind in der Natur weit verbreitet und spielen eine Vielzahl von Rollen[258][259] und zudem wurden zahlreiche Methoden entwickelt, um neue katalytisch aktive RNA-Motive zu identifizieren.[260] Das Hammerhead-Ribozym wurde ursprünglich in Pflanzenviren, Viroiden und Virusoiden entdeckt, wo es der Prozessierung von RNA-Transkripten dient.[261][262] Darüber hinaus wurde es in allen drei Domänen des Lebens, Bacteria, Archaea und Eucarya, nachgewiesen und gehört zu einer der am weitesten verbreiteten Klassen von Ribozymen.[263][264][265] Dieses vielseitige Vorkommen könnte darauf hindeuten, dass es bereits früh in der Evolution entstanden ist und sich sogar mehrfach unabhängig *de novo* entwickelt hat.[266]

Die Sekundärstruktur des Hammerhead Ribozyms besteht aus drei Helices, die durch konservierte Nukleotidsequenzen verbunden sind und welche für die katalytische Aktivität essentiell sind (**Abbildung *10*a**).

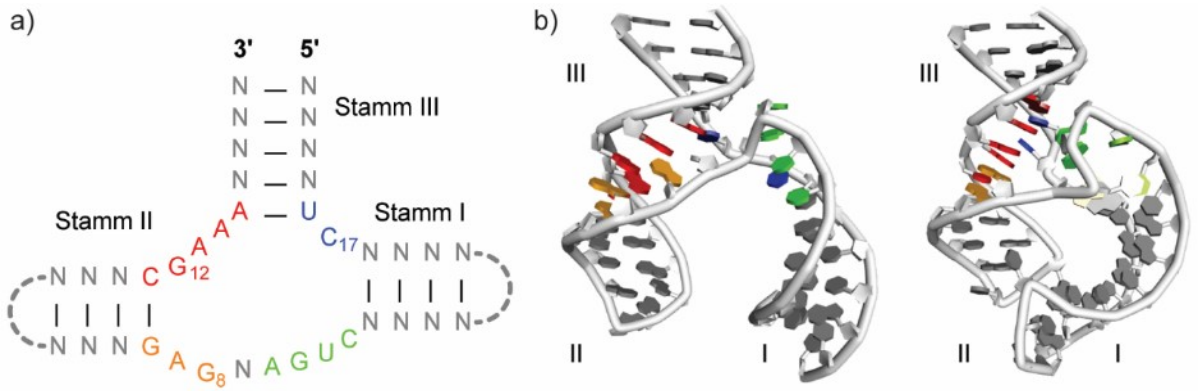
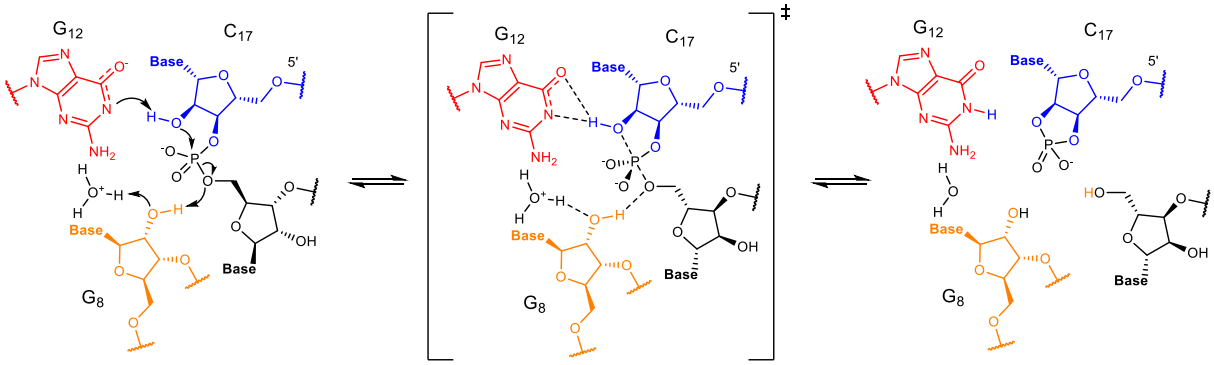


Abbildung 10. Generelles Motiv des III Hammerhead Ribozyms und Konformationen des katalytischen Zentrums. a) Schema des minimalen Hammerhead Ribozym Motivs[266] mit Hervorhebung konservierter Nukleotid Sequenzen (rot, blau, orange, grün) und der Nukleotide G_8 und G_{12}, welche die Hydrolyse der Phosphordiester Bindung des Nukleotides C_{17} katalysieren. b) Zwei Konformationen des Hammerhead Ribozyms. Links (F0), minimales Motiv ohne tertiäre Wechselwirkungen zwischen Stamm I und II. Rechts (F1), verlängertes Ribozym mit tertiären Wechselwirkungen zwischen Stamm I und II und resultierender Konformationsänderung des aktiven Zentrums. Die Kristallstrukturen wurden den PDB-Einträgen 1HMH[267] (links) und 2QUS[268] (rechts) entnommen und mit *Discovery Studio Visualizer v21.1.0* bearbeitet.

Die Hammerhead Ribozyme werden in drei Topologien eingeteilt. Typ I, II oder III, je nachdem in welchem Stamm des Ribozyms sich das 3′- und das 5′-Ende der Sequenz befinden.[257] Als Mechanismus der Spaltung im Hammerhead Ribozym findet eine Transesterifizierung durch Säure-Base-Katalyse statt.[269] Im aktiven Zentrum des Hammerhead Ribozyms befinden sich die Nukleotide G_8, G_{12} und C_{17}, wobei G_8 als Säure und G_{12} als Base in der Transesterifizierung fungieren. Die N^1-Position von G_{12} muss dabei deprotoniert sein, um die 2′-Position des Substrat Nukleotids C_{17} deprotonieren zu können. Dabei tragen monovalente Kationen, wie z.B. Na^+, zur Stabilisierung bei.[270][271] Die 2′-Hydroxygruppe von G_8 ist so positioniert, dass sie als Protonendonor für den 5′-Sauerstoff der Abgangsgruppe in der Spaltungsreaktion dienen kann. Die Reaktion verläuft über einen trigonal-bipyramidalen Überganszustand des Phosphats und ergibt ein 2′-3′-zyklisches Phosphat an C_{17} und ein 5′-Hydroxy-Terminus an der Abgangsgruppe (**Schema *8***).[268][269] Durch die Strukturanalyse mittels Einkristall-Röntgendiffraktometrie und diversen Methoden wie atomare Mutagenese konnten diese Mechanismen des Hammerhead Ribozyms aufgeklärt werden.[272] Es wurden Kristallstrukturen des präkatalytischen Zustands des minimalen Motivs des Hammerhead Ribozyms erhalten, indem Substrate verwendet wurden, die an Position C_{17} keine 2′-Hydroxygruppe aufweisen, d.h. DNA- oder 2′-Methoxy-Nukleotide.[267][273] Auch konnte eine Kristallstruktur eines Hammerhead Ribozyms erhalten werden, das ein nicht-modifiziertes Ribonukleotid an der C_{17} besaß. Hierzu wurde die Struktur in Abwesenheit von bivalenten Kationen, wie z.B. Mg^{2+} oder Mn^{2+}, aufgelöst. Diese sind beim minimalen Motiv des Ribozyms nötig, um eine Konformationsänderung in die aktive Form zu gewährleisten.[274]



Schema 8. Mechanismus der Säure-Base-Katalyse in der Transesterifizierung eines Hammerhead Ribozyms.[268]

Ebenso konnte die katalytische Aktivität des Ribozym reduziert werden, indem G_{12} durch A_{12} substituiert wurde und so das Substrat und auch die Konformation erhalten blieben.[268] Diese Studien haben gezeigt, dass, abgesehen von der konservierten Nukleotidsequenz, tertiäre Wechselwirkungen[275] zwischen Stamm I und Stamm II stattfinden müssen, um ein unter physiologischen Konzentrationen von Mg^{2+} aktives Hammerhead Ribozym zu erhalten (**Abbildung *10*b**).[268][276][277] Die tertiären Wechselwirkungen zwischen Stamm I und Stamm II

führen zu einer Konformationsänderung im aktiven Zentrum des Ribozyms vom präkatalytischen Zustand F0 hin zum katalytisch aktiven Zustand F1.

Zusammenfassend lässt sich sagen, dass das Hammerhead Ribozym ein nützliches Modell für die Untersuchung sowohl sekundärer als auch tertiärer Wechselwirkungen von Nukleinsäuren darstellt. Neben kanonischen Formen des Hammerhead Ribozyms, kann auch der Einfluss neuartiger kovalenter RNA-Aminosäurekonjugate als strukturbildendes Merkmal anhand eines chimären Hammerhead Ribozyms untersucht werden (siehe Kapitel **1.4**).[252][255][256] Besonders im Szenario einer RNA-Peptid-Welt ist es nicht nur von Interesse eine mögliche RNA-basierte Peptidsynthese zu rekonstruieren, sondern auch mögliche Einflüsse von Aminosäuren auf die katalytische Aktivität von Ribozymen nachzuvollziehen.[145][206][207]

2. Zielsetzung

RNA-Modifikationen sind ein weit verbreiteter Bestandteil in tRNAs und rRNAs. Sie werden über enzymatische Prozesse post-transkriptionell eingefügt, wobei sie der korrekten Ausprägung von Sekundärstrukturen dienen und eine wichtige Rolle in Codon-Anticodon-Wechselwirkungen spielen. Wann ursprünglich solche Modifikationen auftraten und ob diese bereits in einem frühen Stadium der chemischen Evolution existierten verbleibt eine offene Frage in der Forschung zur Entstehung des Lebens. In der tRNA befindet sich die Aminosäure, aus der die Peptidkette wächst an der am 3′-Ende der Ribose und ist dort kovalent in Form eines Esters gebunden. Entsprechend befassen sich zahlreiche Studien zum Ursprung der Translation mit RNA-Aminosäure-Konjugaten, die diese Ester-Verknüpfung nachstellen. Diese Verknüpfungen erweisen sich jedoch als hydrolytisch instabil, besonders in Umgebungen, die stark von physiologischen Bedingungen abweichen. Daraus resultiert eine geringere Halbwertszeit in wässriger Lösung, was RNA-Aminosäure-Konjugate, wie sie in der Translation Verwendung finden, zu einem frühen Zeitpunkt der Evolution als unwahrscheinlich erscheinen lässt. Jedoch ist das 3′-Ende nicht die einzige Position in der tRNA, die mit Aminosäuren modifiziert sein kann. Ebenso befinden sich in manchen tRNAs in der Nähe der Anticodonschleife nicht-kanonische Nukleotide, die Aminosäuren über eine Harnstoffbrücke an der Nukleobase tragen können.

Das Ziel dieser Arbeit war es den Einfluss nicht-kanonischer Nukleotide, in denen Aminosäuren kovalent mit der Nukleobase verknüpft sind, in einer möglichen RNA-Peptid-Welt zu untersuchen. Hierzu sollten präbiotisch plausible Syntheserouten entwickelt werden durch welche sowohl Aminosäure-modifizierte *N*-Carbamoyl-Nukleoside als auch -Nukleotide erhalten werden können. Neben der präbiotischen Synthese, sollte auch eine chemische Synthese durchgeführt werden, die es ermöglicht über RNA-Phosphoramidit-Chemie kurze Oligomere dieser Modifikationen herstellen zu können. Diese Oligomere sollten dann genutzt werden, um eine alternative Form der RNA-basierten Peptidsynthese zu etablieren, in der Aminosäuren, im Gegensatz zu der in der Translation verwendeten Ester-Chemie, zwischen den Nukleobasen übertragen werden können. Damit sollte gezeigt werden, dass RNA-Modifikationen, die in heutigen tRNAs gefunden werden können, die Eigenschaft besitzen eine primitive Form der Peptidsynthese mit hydrolytisch stabileren Konjugaten auszubilden. Neben der Synthese kurzer Peptide, war es ebenso ein Ziel den Einfluss der strukturgebenden Eigenschaften des Haarnadel-Motivs, in dem zwei RNA-Moleküle über eine Peptid-Bindung zwischen den Nukleobasen verbunden sind, zu untersuchen.

3. Veröffentlichte Arbeiten

3.1 Aminosäure-modifizierte RNA-Basen als Bausteine einer RNA-Peptid-Welt auf der frühen Erde

"Amino Acid Modified RNA Bases as Building Blocks of Early Earth RNA-peptide world"

Autoren

Milda Nainytė, Felix Müller, Giacomo Ganazzoli, Chun-Yin Chan, Antony Crisp, Daniel Globisch, Thomas Carell*

Chem. Eur. J., **2020**, *26*, 14856-14860. ‡ ‡ Für *Supplementary Information* siehe Anhang I

DOI: https://doi.org/10.1002/chem.202002929

Prolog

Die RNA ist reich an post-transkriptionellen Modifikationen, von denen einige als molekulare Fossilien betrachtet werden können. Eine davon ist das hypermodifizierte Nukleosid t^6A, eine N^6-Carbamoyl-Modifikation von Adenosin, die sich an Position 37 einiger tRNAs befindet. Dort dient es der Anticodonerkennung und trägt zur korrekten Ausprägung der Sekundärstruktur der Anticodonschleife bei, ohne selbst eine Basenpaarung einzugehen. Neben der enzymatischen Synthese dieser Modifikation existieren auch präbiotische Synthesewege, über die die Modifikationen t^6A und g^6A erhalten werden können. Diese konnten jedoch nur in Form von Nukleosiden und nur in geringer Ausbeute erhalten werden. In der vorliegenden Studie wurden neue chemische Synthesewege zu Phosphoramidit-Derivaten des Aminosäure-modifizierten *N*-Carbamoyl-Adenosins entwickelt, um die Eigenschaften der Modifikationen in Oligonukleotiden analysieren zu können. Dabei wurde das Repertoire der an die N^6-Position des Adenosins gebundenen Aminosäuren um unpolare, aromatische und ionisierbare Aminosäuren erweitert. Anschließend konnte gezeigt werden, dass die in Oligonukleotiden eingebauten modifizierten Nukleotide keine Basenpaarungen bilden, vergleichbar mit der Eigenschaft in heutiger tRNA. Die hierbei gewonnen Erkenntnisse dienen als wichtige Grundlage für die Aufklärung der Rolle Aminosäure-modifizierter Nukleotide in einer möglichen RNA-Peptid-Welt.

Autorenbeitrag

Es wurden die Verbindungen **28**, **35**, **42** und **49** hergestellt und charakterisiert. Zudem wurde das modifizierte Oligonukleotid **ON2** hergestellt und der Einfluss des Aminosäure-modifizierten Adenosins auf die Schmelztemperatur eines RNA-Doppelstrangs untersucht.

Chemistry–A European Journal
Communication
doi.org/10.1002/chem.202002929
Chemistry Europe
European Chemical Societies Publishing

Oligonucleotides

Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World

Milda Nainytė,[a] Felix Müller,[a] Giacomo Ganazzoli,[a] Chun-Yin Chan,[a] Antony Crisp,[a] Daniel Globisch,[b] and Thomas Carell*[a]

Abstract: Fossils of extinct species allow us to reconstruct the process of Darwinian evolution that led to the species diversity we see on Earth today. The origin of the first functional molecules able to undergo molecular evolution and thus eventually able to create life, are largely unknown. The most prominent idea in the field posits that biology was preceded by an era of molecular evolution, in which RNA molecules encoded information and catalysed their own replication. This RNA world concept stands against other hypotheses, that argue for example that life may have begun with catalytic peptides and primitive metabolic cycles. The question whether RNA or peptides were first is addressed by the RNA-peptide world concept, which postulates a parallel existence of both molecular species. A plausible experimental model of how such an RNA-peptide world may have looked like, however, is absent. Here we report the synthesis and physicochemical evaluation of amino acid containing adenosine bases, which are closely related to molecules that are found today in the anticodon stem-loop of tRNAs from all three kingdoms of life. We show that these adenosines lose their base pairing properties, which allow them to equip RNA with amino acids independent of the sequence context. As such we may consider them to be living molecular fossils of an extinct molecular RNA-peptide world.

The RNA-peptide co-evolution hypothesis describes the emergence of self-replicating molecules that contained amino acids and RNA.[1] At the macromolecular level, this tight coexistence of peptides and RNA is established in the ribosome, where encoding and catalytic RNA is supported by proteins.[2] Although we cannot delineate how such an early RNA-peptide world may have looked like, it seems not too implausible to assume that some of the molecular components may have survived until today as vestiges of this extinct world.[3] tRNAs derived from all three kingdoms of life contain a large number of modified bases,[4] and some of them are indeed modified with amino acids.[3] The most wide spread amino acid modified bases are adenosine nucleosides, in which the amino acid is linked via urea connector to the N^6-amino group of the heterocycle as depicted in Figure 1 a. Particularly ubiquitous are adenosine modifications containing the amino acids threonine (t^6A)[5–7] and glycine (g^6A),[8] together with hn^6A.[9,10] Based upon recent phylogenetic analyses and the fact that t^6A is found in all three kingdoms of life, it has been suggested that such amino acid modified bases were already present in the last universal common ancestor (LUCA), from which all life

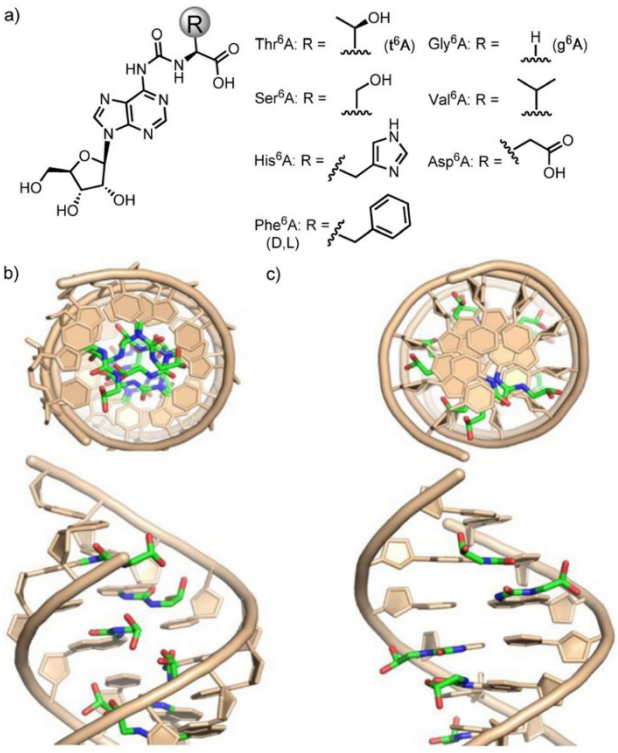


Figure 1. (a) Depiction of the amino acid modified A-bases (aa^6A) together with computer visualizations that show how such bases may reside in an (b) A-form RNA duplex and a (c) B-form DNA duplex. The sequence used for the visualization is: 5′-CAUAUAUAUAUG-3′ with A = g^6A.

[a] M. Nainytė, F. Müller, G. Ganazzoli, C.-Y. Chan, A. Crisp, Prof. Dr. T. Carell
Department of Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de

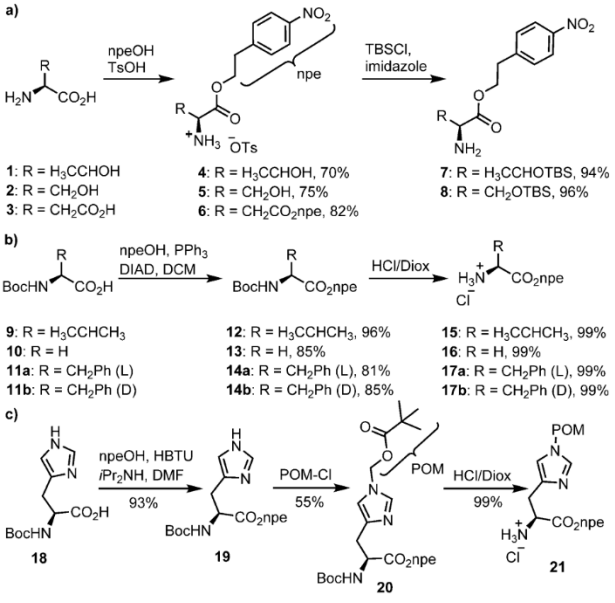
[b] Dr. D. Globisch
Department of Medicinal Chemistry, Uppsala University
Husargatan 3, 75123 Uppsala (Sweden)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/chem.202002929.

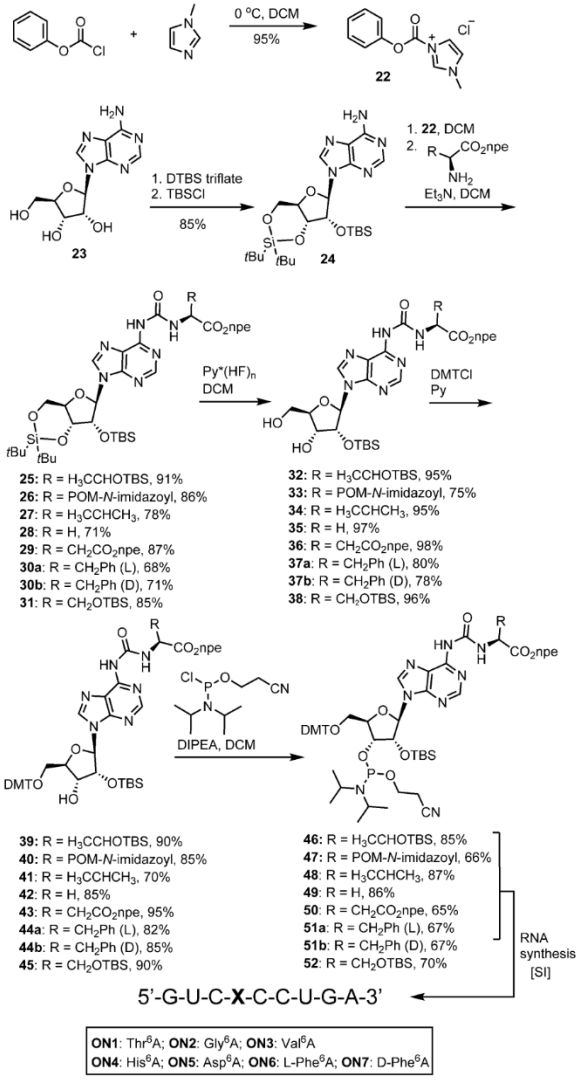
forms descended.[11–14] t^{6}A is for example today found in nearly all ANN decoding tRNAs.[15] We recently reported a plausible prebiotic route to some of these amino acid modified A-bases, which strengthens the idea that they could indeed be living chemical fossils of the extinct RNA-peptide world.[16] Despite the interesting philosophical genotype-phenotype dualism that characterizes these structures and their contemporary importance for the faithful decoding of genetic information, a general synthesis of aa^{6}A modified bases (Figure 1 a) and a systematic study of their properties is lacking.

Here we report the synthesis of a variety of aa^{6}A nucleosides with canonical amino acids (aa = Asp, Gly, His, Phe, Thr,[17] Ser, Val), their incorporation into DNA and RNA and an investigation of how they influence the physicochemical properties of oligonucleotides. We were particularly interested to study how they might affect the stability of RNA and DNA. The computer visualization shows that in A-form RNA (Figure 1 b), the amino acid part of the aa^{6}A base would need to reside inside the helix, shielded from the outside. In the B-form DNA one could imagine a decoration of the major groove with the amino acid side chains as depicted in Figure 1 c.

In the Schemes 1 and 2 we show the synthesis of the different urea linked amino acid A-derivatives (aa^{6}A). We first prepared the amino acid components for the coupling to the A-nucleoside (Scheme 1). Our starting points for Thr^{6}A, Ser^{6}A and Asp^{6}A were the free amino acids **1**–**3**, in which we first transformed all carboxylic acids into the *p*-nitrophenylethyl esters (npe, **4**–**6**).[17] The hydroxy groups of the Thr and Ser compounds were finally protected as TBS-ethers to give the final products **7** and **8** (Scheme 1 a). For Val, Gly and Phe we started with the Boc-protected amino acids **9**–**11**, which we also converted into the npe-esters **12**–**14** using Mitsunobu type chemistry[18] followed by acidic (4 M HCl in dioxane) Boc-deprotection to give the amino acid products **15**–**17** (Scheme 1 b).[19] For His^{6}A, we again started with the Boc-protected amino acid **18** (Scheme 1 c) and used HBTU activation to generate the npe ester **19**. Protection of the imidazole N^τ with POM-chloride followed again by Boc-deprotection furnished the ready to couple amino acid **21**.



Scheme 1. Synthesis of the amino acid building blocks as needed for the coupling to the nucleoside A to give Thr^{6}A, Ser^{6}A, Asp^{6}A, Val^{6}A, Gly^{6}A, Phe^{6}A and His^{6}A.

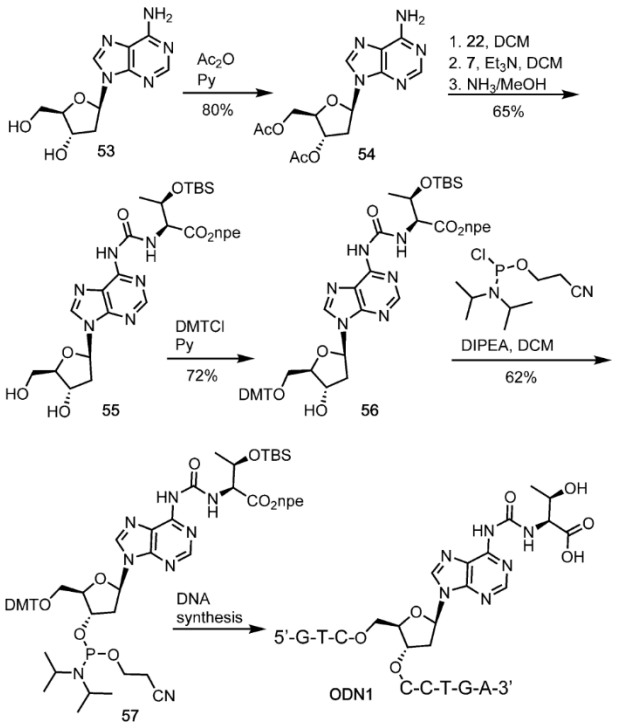


Scheme 2. Synthesis of phosphoramidite building blocks of Thr^{6}A, Ser^{6}A, Asp^{6}A, Val^{6}A, Gly^{6}A, Phe^{6}A and His^{6}A and their incorporation into RNA.

The connection of the amino acid with the A-nucleoside via the urea moiety was next carried out as depicted in Scheme 2. We first treated phenyl chloroformate with *N*-methylimidazole to obtain the 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**22**).[20] Adenosine was converted in parallel into the cyclic 3′,5′-silyl protected nucleoside, followed by conversion of the 2′-OH group into the TBS-ether.[21] The reaction of compound **24** with the activated carbonate and the corresponding amino acid, provided in all cases the amino acid coupled products **25**–**31** in good to excellent yields. Subsequent cleavage of the

cyclic silylether with HF·pyridine complex,[22,23] protection of the 5′-OH group with dimethoxytritylchloride (DMTCl)[24] allowed the final conversion of the compounds into the corresponding phosphoramidites **46**–**52**. Standard solid phase RNA chemistry[25–31] was subsequently employed to prepare RNA strands containing the individual aa^{6}A nucleosides stably embedded. The standard RNA synthesis protocol did not require any adjustment. In all cases we observed fair coupling of the aa^{6}A phosphoramidites and no decomposition during deprotection. Deprotection required three steps. First, with DBU in THF at r.t. for 2 h we cleaved the npe-protecting group. Second, we deprotected the bases and cleaved from the solid support with aqueous NH_3/$MeNH_2$. Finally, we removed the 2′-silyl group with HF in NEt_3.

In order to investigate how aa^{6}A bases would affect the stability of DNA duplexes we also prepared as a representative molecule t^{6}dA as depicted in Scheme 3. To this end we first acetyl-protected dA **53**,[32] performed the coupling of the protected threonine with the activated carbonate **22**, cleaved the acetyl groups and converted the nucleoside subsequently into the 5′-DMT protected phosphoramidite **57**. The purification of compound **57** was quite difficult due to its high polarity. We needed to use rather polar mixture of EtOAc/Hex (2/1) as the mobile phase for the chromatographic separation. This provided the phosphoramidite **57**, however the material had a lower purity in comparison to the RNA phosphoramidites. Nevertheless, solid phase DNA synthesis and deprotection of the DNA strand **ODN1** proceeded again smoothly and in high yields.



Scheme 3. Synthesis of t^{6}dA phosphoramidite and its incorporation into DNA.

Figure 2 a shows as an example the raw HPL-chromatograms of **ON1** (RNA strand with embedded t^{6}A) and the corresponding chromatogram after purification (inset) together with the obtained MALDI-TOF mass spectrum (Figure 2 b). The chromatograms of the raw material show a good quality of the obtained RNA material. The analytical chromatogram after purification and the MALDI-TOF data prove the purity of the finally obtained RNA oligonucleotide and the integrity of the t^{6}A-containing RNA strand.

Figure 2 c and 2 d show the same data set for the t^{6}dA containing DNA oligonucleotide (**ODN1**), proving again the successful synthesis of t^{6}dA containing oligonucleotide. The aa^6(d)A nucleosides can exist in two different conformations.[33] The first, s-*trans*, maintains the Watson–Crick hydrogen bonding capabilities with the urea amino acid oriented towards the imidazole ring system (Figure 3 a). This allows formation of a Hoogsteen type 7-membered ring H-bond with the N^7. In the corresponding s-*cis*-conformation, the urea amino acid orients towards the Watson–Crick side thereby establishing a typically strong intramolecular 6-membered H-bond with N^1 (Figure 3 b). In order to investigate if the embedding of the amino acid would enforce s-*trans*-conformation and hence Watson–Crick H-bonding, we measured melting points of all aa^{6}A containing RNA strands and of the t^{6}dA containing DNA strand hybridized to the corresponding counter strands (Figure 3). In the RNA:RNA situation we noted for all aa^{6}A strands that we investigated, a single clear melting point, showing that only one conformer of the aa^{6}A base likely exists in the RNA:RNA duplexes. In situation where the aa^{6}A base exists in two different stable conformations, one would expect a more complex melting behaviour. In all cases we saw that the melting point is strongly reduced by 10–15 °C. When we embedded two aa^{6}A building blocks into a short RNA strand no duplex formation

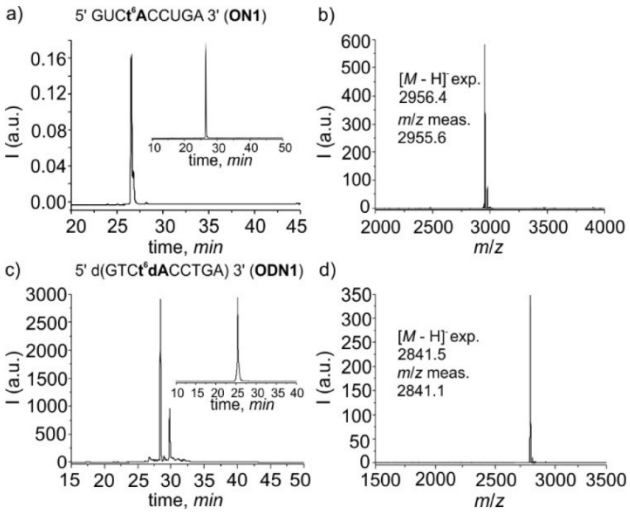
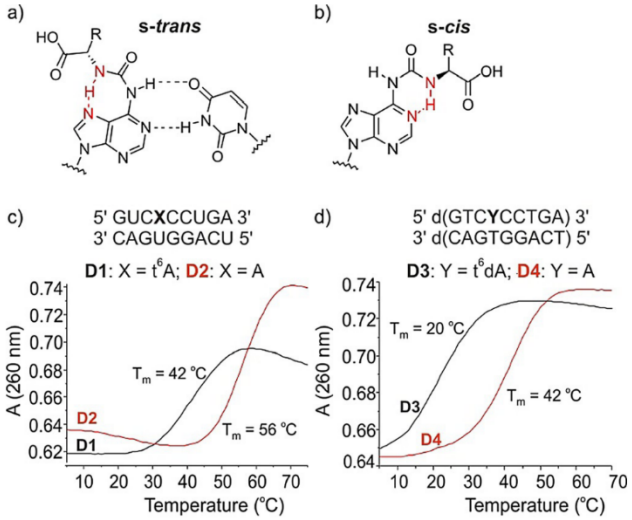


Figure 2. (a) Raw-HPL chromatogram of **ON1**, with the inset showing the HPL-chromatogram of purified **ON1**; (b) MALDI-TOF mass spectrum of **ON1** after purification; (c) raw-HPL chromatogram of **ODN1**, with the inset showing the HPL chromatogram of purified **ODN1**; (d) MALDI-TOF mass spectrum of **ODN1** after purification.



e)

	T_m, °C		T_m, °C		T_m, °C
A:U	56	Val^{6}A:U	44	L-Phe^{6}A:U	41
t^{6}A:U	42	His^{6}A:U	45	D-Phe^{6}A:U	41
g^{6}A:U	45	Asp^{6}A:U	40		

Figure 3. (a, b) Possible conformation, base pairing and intramolecular H-bond of aa^{6}A; (c, d) melting curves measured for t^{6}A containing RNA:RNA duplexes and of a t^{6}dA containing DNA:DNA duplex in comparison with the duplexes containing canonical (d)A:(d)T base pairs; (e) table of the determined melting points.

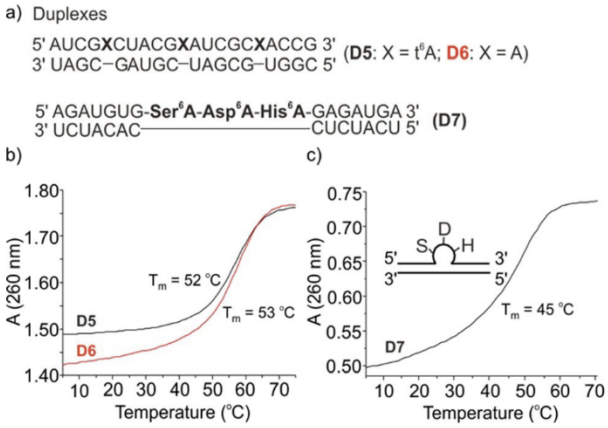


Figure 4. (a) Depiction of the RNA structures containing aa^{6}A nucleobases in extrahelical positions forming either three little bulges or assembling a Ser-Asp-His triad known as the catalytic triad in serine proteases; (b, c) depiction of melting curves of duplexes **D5**, **D6**, **D7**; S: serine, D: aspartate, H: histidine.

was obtained. Even stronger reduction of the melting point was observed for the DNA duplex containing one t^{6}dA. Here, we also saw just one sharp melting point and a reduction of the T_m by over 20 °C. These data show that the aa^{6}A bases and among them t^{6}A and g^{6}A are unable to base pair. Although we have no direct proof of the structure the data argue for a preferred *s-cis*-conformation (Figure 3b) in agreement with the literature.[34]

This conclusion is also supported by the observation that irrespective of the chirality of the attached amino acid (L- versus D-Phe^{6}A), we measured the same melting temperature. This would not be expected if the *s-trans*-conformation and base pairing would be possible. These data suggest that aa^{6}A nucleosides within RNA position a given amino acid outside the A-form helix in an unpaired situation and hence independent from the counterbase. As such, multiple aa^{6}A containing RNA strands would be structures in which the RNA part is decorated by the amino acid side chains. In order to show that RNA-structures containing multiple amino acids as representatives of an RNA-peptide world can stably form, we prepared two RNA duplexes (Figure 4). In the first (**D5**), we placed three t^{6}A bases as extra bases in an otherwise undisturbed RNA duplex. Indeed, now the stability of this duplex was indistinguishable from the same construct containing just canonical bases (**D6**). Finally, we prepared an RNA duplex **D7**, in which we placed the amino acids Ser-Asp-His directly next to each other to simulate what is known in the peptide world as the catalytic triad present in serine peptidases.[35] Again in this case a stable duplex structure forms with the three aa^{6}A bases creating a loop. Although we do not show any catalytic activity here, we believe that it is easily imaginable that if these amino acids are properly positioned in a stably folded RNA the structure could gain catalytic properties.

The melting data show, that aa^{6}A bases alone are unable to establish base pairing, which hinder them to encode sequence information. On the other side, these bases allow the incorporation of amino acids into RNA structures irrespective of the counterbase. Because RNAs are mostly stably folded structures in which many bases are not involved in any base pairing or establish no Watson–Crick interactions the amino acid adenosine nucleosides allow the stable incorporation of amino acid functionality into RNA.

In summary, here we investigated the synthesis and properties of aa^{6}A nucleoside-amino acid conjugates, some of which (t^{6}A, g^{6}A, hn^{6}A) are today found as key components in the tRNAs of many species. In these tRNAs the aa^{6}A nucleosides reside at the general purine position 37 adjacent to the anticodon loop. They are not involved in base pairing but fine tune the codon-anticodon interaction to enable faithful translation of information into a peptide sequence.[36] Here we show that these bases are indeed unable to base pair. They have to be placed outside the pairing regime that is needed for RNA folding. As such they function as anchors that allow the connection of amino acid to RNA structures independent of the counterbase. The side chains are then available to equip RNA with additional functions that might have been beneficial in an early RNA-peptide world. The fact that aa^{6}A nucleosides are stable structures and until today broadly found in today's RNA make them prime candidates to develop idea about the chemical constitution of the vanished RNA-peptide world.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1309 (325871075) and SPP1784 (255344185). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° EPiR 741912) and through a H2020 Marie Skłodowska-Curie Action (LightDyNAmics, 765866). We thank Dr. Tynchtyk Amatov for the initial synthesis of the t^6A-phosphoramidite. We thank Dr. Markus Müller for helpful discussions and preparing Figure 1. Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Keywords: amino acid nucleosides · origin of life · prebiotic chemistry · RNA world · RNA-peptide world

[1] P. T. S. van der Gulik, D. Speijer, *Life* **2015**, *5*, 230 – 246.
[2] D. L. J. Lafontaine, D. Tollervey, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 514 – 520.
[3] H. Grosjean, E. Westhof, *Nucleic Acids Res.* **2016**, *44*, 8020 – 8040.
[4] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Howkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262 – D267.
[5] G. B. Chheda, R. H. Hall, J. Mozejko, D. I. Magrath, M. P. Schweizer, L. Stasiuk, P. R. Taylor, *Biochemistry* **1969**, *8*, 3278 – 3282.
[6] S. Takemura, M. Murakami, M. Miyazaki, *J. Biochem.* **1969**, *65*, 489 – 491.
[7] D. M. Powers, A. Peterkofsky, *Biochem. Biophys. Res. Commun.* **1972**, *46*, 831 – 838.
[8] M. P. Schweizer, K. McGrath, L. Bacznyskyj, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1046 – 1052.
[9] H. Grosjean in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (Ed.: H. Grosjean), Landes Bioscience, Austin, Texas, pp. 1 – 18.
[10] D. M. Reddy, P. F. Crain, C. G. Edmonds, R. Gupta, T. Hashizume, K. O. Stetter, F. Widdel, J. A. McCloskey, *Nucleic Acids Res.* **1992**, *20*, 5607 – 5615.
[11] M. Di Giulio, *J. Theor. Biol.* **1998**, *191*, 191 – 196.
[12] H. Grosjean, V. de Crécy-Lagard, G. R. Björk, *Trends Biochem. Sci.* **2004**, *29*, 519 – 522.
[13] E. Szathmáry, *Trends Genet.* **1999**, *15*, 223 – 229.
[14] L. Perrochia, E. Crozat, A. Hecker, W. Zhang, J. Bareille, B. Collinet, H. van Tilbeurgh, P. Forterre, T. Basta, *Nucleic Acids Res.* **2013**, *41*, 1953 – 1964.
[15] P. C. Thiaville, D. Iwata-Reuyl, V. de Crécy-Lagard, *RNA Biol.* **2014**, *11*, 1529 – 1539.
[16] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943 – 5946; *Angew. Chem.* **2018**, *130*, 6050 – 6054.
[17] V. Boudou, J. Langridge, A. van Aerschot, C. Hendrix, A. Millar, P. Weiss, P. Herdewijn, *Helv. Chim. Acta* **2000**, *83*, 152 – 161.
[18] O. Mitsunobu, M. Yamada, *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2380 – 2382.
[19] G. Leszczynska, P. Leonczak, A. Dziergowska, A. Malkiewicz, *Nucleosides Nucleotides Nucleic Acids* **2013**, *32*, 599 – 616.
[20] F. Himmelsbach, B. S. Schultz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59 – 72.
[21] V. Serebryany, L. Beigelman, *Tetrahedron Lett.* **2002**, *43*, 1983 – 1985.
[22] B. M. Trost, C. G. Caldwell, *Tetrahedron Lett.* **1981**, *22*, 4999 – 5002.
[23] M. Sekine, S. Limura, K. Furusawa, *J. Org. Chem.* **1993**, *58*, 3204 – 3208.
[24] H. Schaller, G. Weinmann, B. Lerch, H. G. Khorana, *J. Am. Chem. Soc.* **1963**, *85*, 3821 – 3827.
[25] H. Seliger, D. Zeh, G. Azuru, J. B. Chattopadhyaya, *Chem. Scr.* **1983**, *22*, 95 – 102.
[26] M. H. Caruthers, D. Dellinger, K. Prosser, A. D. Brone, J. W. Dubendorff, R. Kierzek, M. Rosendahl, *Chem. Scr.* **1986**, *26*, 25 – 30.
[27] R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, S. M. Freier, *Biochemistry* **1986**, *25*, 7840 – 7846.
[28] I. Hirao, M. Ishikawa, K. Miura, *Nucleic Acids Symp. Ser.* **1985**, *16*, 173 – 176.
[29] H. Tanimura, T. Fukuzawa, M. Sekine, T. Hata, J. W. Efcavitch, G. Zon, *Tetrahedron Lett.* **1988**, *29*, 577 – 578.
[30] H. Tanimura, M. Sekine, T. Hata, *Nucleosides Nucleotides* **1986**, *5*, 363 – 383.
[31] H. Tanimura, T. Imada, *Chem. Lett.* **1990**, *19*, 1715 – 1718.
[32] E. M. van der Wenden, J. K. von Frijtag Drabbe Künzel, R. A. A. Mathot, M. Danhof, A. P. IJzerman, W. Soudijn, *J. Med. Chem.* **1995**, *38*, 4000 – 4006.
[33] R. Parthasarathy, J. M. Ohrt, G. B. Chheda, *Biochem. Biophys. Res. Commun.* **1974**, *60*, 211 – 218.
[34] F. V. Murphy IV, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Str. Mol. Biol.* **2004**, *11*, 1186 – 1191.
[35] L. Polgár, *Cell. Mol. Life Sci.* **2005**, *62*, 2161 – 2172.
[36] P. C. Thiaville, R. Legendre, D. Rojas-Benítez, A. Baudin-Baillieu, I. Hatin, G. Chalancon, A. Glavic, O. Namy, V. de Crécy-Lagard, *Microb. Cell.* **2016**, *3*, 29 – 45.

Manuscript received: June 18, 2020
Accepted manuscript online: June 23, 2020
Version of record online: October 14, 2020

3.2 Ein präbiotisch plausibles Szenario einer RNA-Peptid-Welt

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

Autoren

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

† Beitrag der Autoren zu gleichen Teilen.

Nature, **2022**, *605*, 279-284. ‡

‡ Für *Supplementary Information* siehe Anhang II

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

Prolog

Die RNA-Welt-Theorie besagt, dass RNA, die sowohl genetische Informationen speichern als auch Reaktionen katalysieren kann, am Anfang der chemischen Evolution stand. Diese Hypothese wird durch die wesentliche Rolle der RNA bei der Translation gestützt. Das Ribosom, dessen aktives Zentrum aus RNA besteht, katalysiert die Peptidyltransferreaktion, bei der die in der mRNA enthaltene Information in eine Peptid-Sequenz übersetzt wird, die an das 3'-Ende der tRNA gebunden ist. Die Entstehung dieser komplexen Maschinerie ist eine der zentralen ungelösten Fragen zur Entstehung des Lebens. Es wird vermutet, dass statt einer reinen RNA-Welt eine Co-Evolution zwischen RNA und Peptiden stattgefunden hat. Dabei könnte zunächst eine zufällige Peptidsynthese stattgefunden haben, die zu einer Erweiterung der katalytischen Eigenschaften der RNA durch Peptide geführt haben könnte. In der vorliegenden Studie konnte gezeigt werden, dass RNA mit Hilfe von nicht-kanonischen Nukleotiden, die heute als konservierte Modifikationen in tRNAs wiedergefunden werden können, eine primitive Form der Peptidsynthese ausbilden kann, bei der die Aminosäuren kovalent an die Nukleobasen gebunden sind.

Autorenbeitrag

(Siehe Publikation)

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

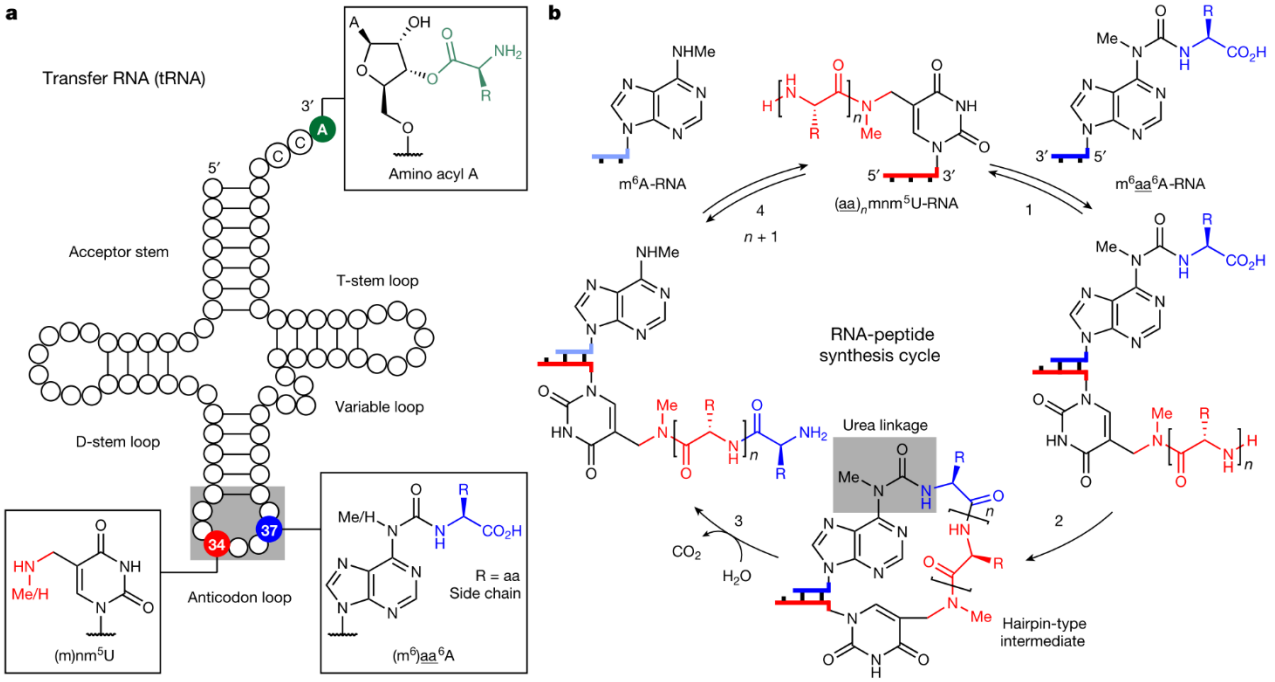


Fig. 1 | Concept of how nucleoside relics of the RNA world enable RNA-based peptide synthesis. a, tRNA structure showing selected ribose and nucleobase modifications. The 3′-amino acid-acylated adenosine is located at the CCA 3′ end in contemporary tRNAs. 5-Methylaminomethyl uridine, mnm^{5}U, is found 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′-ymnm^{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%.

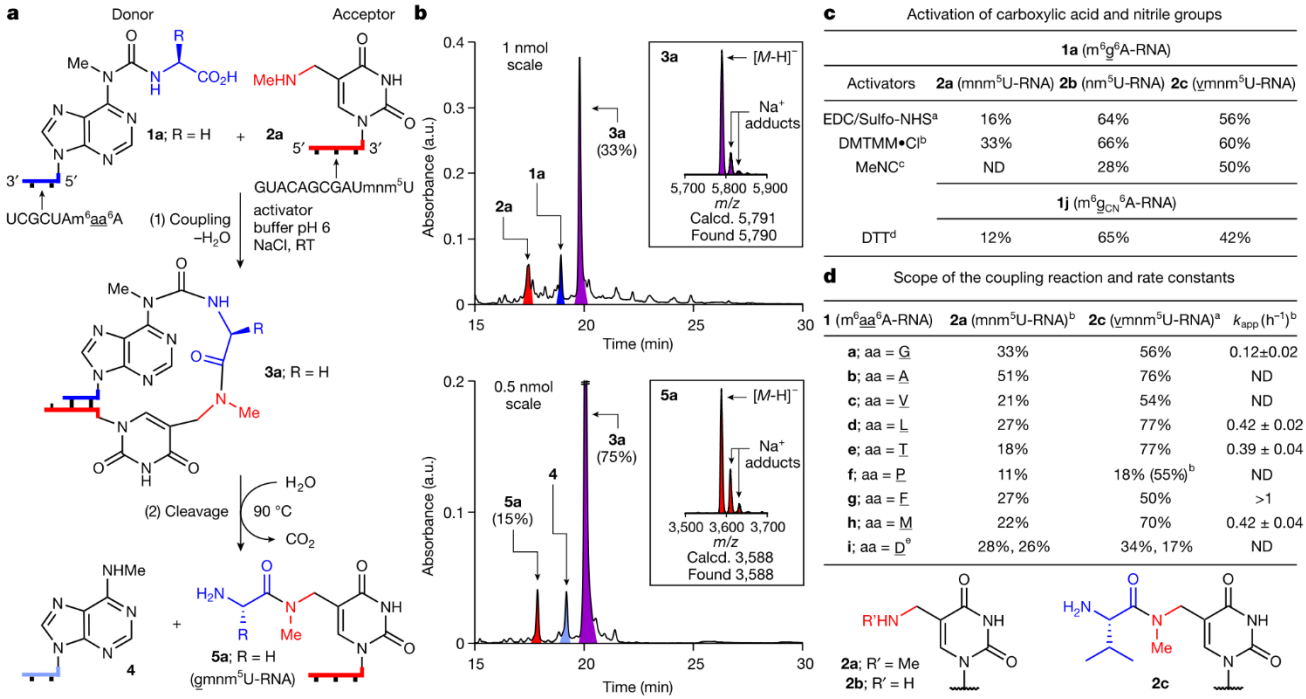
Longer peptide structures on RNA

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



Activators	1a (m6g6A-RNA) 2a (mnm5U-RNA)	2b (nm5U-RNA)	2c (vmnm5U-RNA)
EDC/Sulfo-NHS[a]	16%	64%	56%
DMTMM•Cl[b]	33%	66%	60%
MeNC[c]	ND	28%	50%
	1j (m6g$_{CN}$6A-RNA)		
DTT[d]	12%	65%	42%

1 (m6aa6A-RNA)	2a (mnm5U-RNA)[b]	2c (vmnm5U-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′-gagggagggvmnm^{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

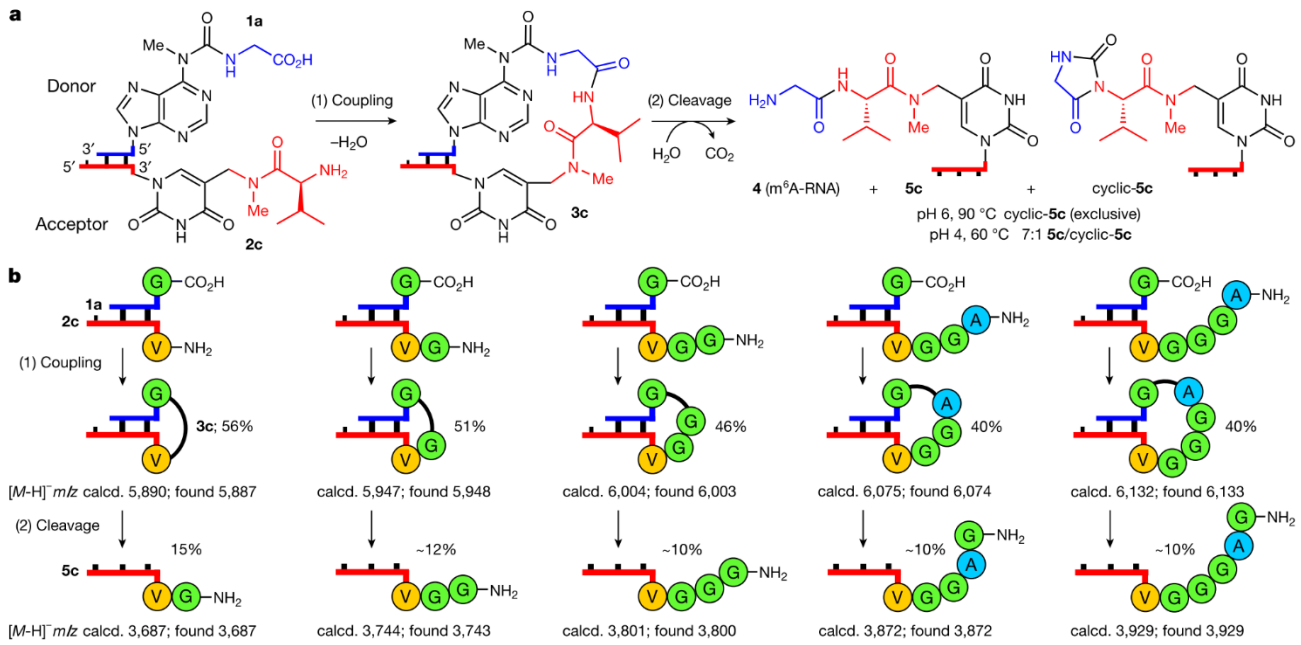


Fig. 3 | Growth of longer peptide structures on RNA. a, Scheme for the reaction of **1a** (5′-m^{6}g^{6}A-RNA-3′) with **2c** (3′-ymnm^{5}U-RNA-5′) including coupling (1) and cleavage (2). **b**, Coupling reactions between **1a** and RNA–peptide acceptor strands using EDC/Sulfo-NHS (see reaction condition 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

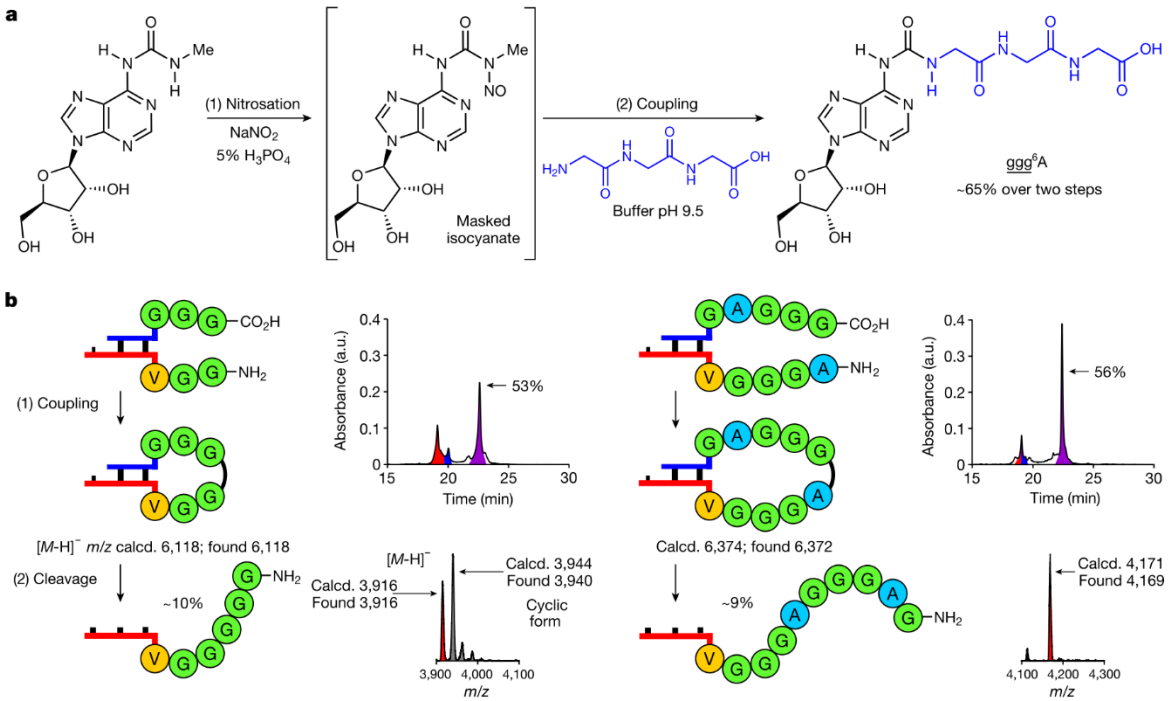


Fig. 4 | Capture of peptides by nitrosated N^6-methylurea adenosine for fragment condensation. a, Prebiotically plausible formation of peptide^{6}A structures, such as ggg^{6}A. **b**, Coupling reactions between RNA-peptide conjugates using EDC/Sulfo-NHS (see reaction condition a in Fig. 2) and cleavage 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.

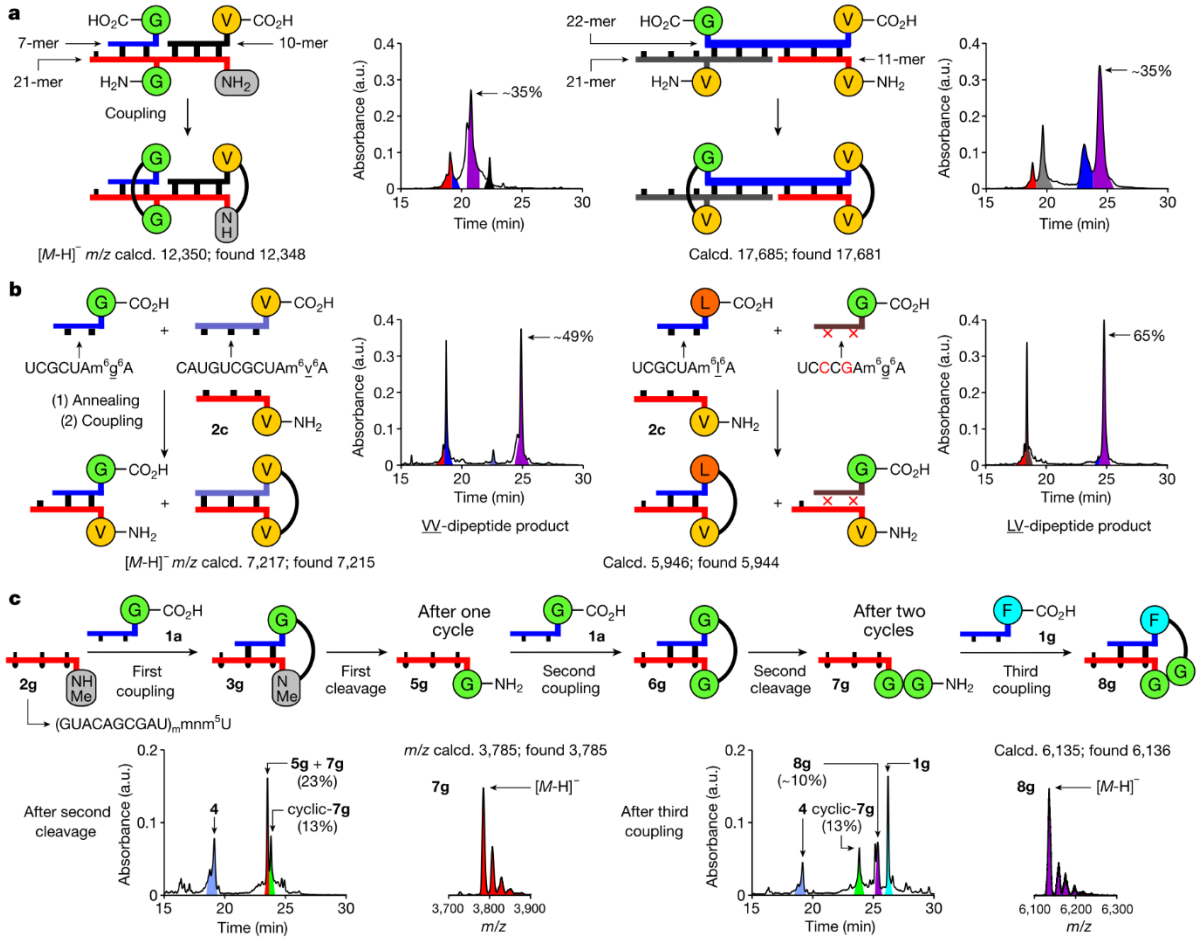


Fig. 5 | Parallel growth of peptides at various positions on RNA, effect of base pairing and RNA–peptide synthesis cycles. a, Coupling of oligonucleotides containing multiple donor or acceptor units (EDC/Sulfo-NHS, see reaction condition a in Fig. 2). **b**, Annealing followed by coupling (EDC/Sulfo-NHS, 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}I^{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

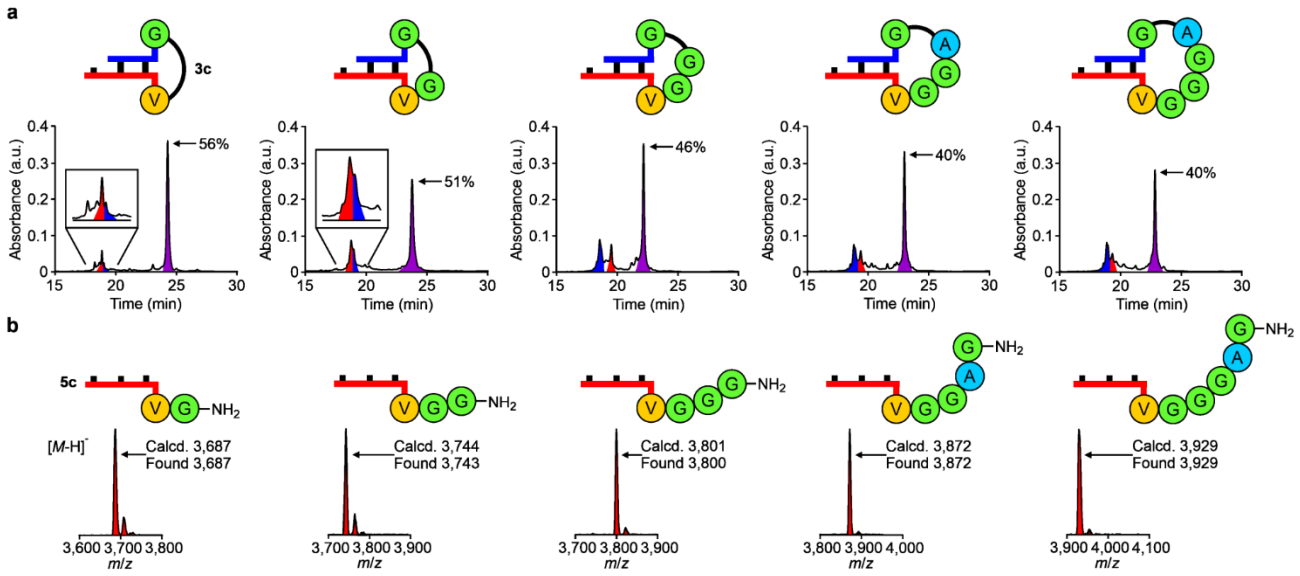
The data that support the findings of this study are available within the paper and its Supplementary Information.

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).

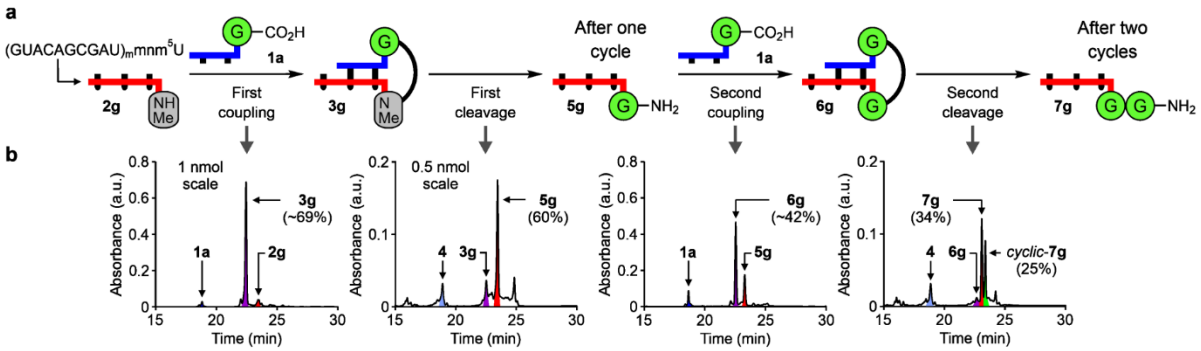
Author contributions F.M., L.E., F.X. and E.W. synthesized the modified phosphoramidites and RNA strands and performed the peptide coupling and urea cleavage experiments. M.N. synthesized RNA donor strands and performed preliminary experiments. T.A. refined and developed mechanistic concepts and performed 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.



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.

3.3 Beladung von RNA mit Aminosäuren in einer putativen RNA-Peptid-Welt

"Loading of Amino Acids onto RNA in a Putative RNA-Peptide World"

Autoren

Johannes N. Singer†, Felix M. Müller†, Ewa Węgrzyn, Christina Hölzl, Hans Hurmiz, Chuyi Liu, Luis Escobar* und Thomas Carell*

† Beitrag der Autoren zu gleichen Teilen.

Angew. Chem. Int. Ed., **2023**, *62*, e202302360. ‡

‡ Für *Supplementary Information* siehe Anhang III

Prolog

RNA besitzt die Eigenschaft, sowohl Informationen zu speichern als auch Reaktionen katalysieren zu können. Dieser Dualismus rückt die RNA in den Mittelpunkt der Theorien über die Entstehung des Lebens. Es bleibt jedoch unklar, ob die Proto-RNA nur aus den vier kanonischen Basen bestand oder ob es eine größere chemische Vielfalt gab. In den heutigen Lebensformen wurde eine Vielfalt an Modifikationen der RNA nachgewiesen und auch in präbiotischen Synthesen wurden weitaus vielseitigere Verbindungen als die vier kanonischen Nukleotide A, U, C, und G erhalten. In dieser Arbeit wurde gezeigt, dass Nukleoside und RNA unter präbiotischen Bedingungen als erster Schritt einer RNA-basierten Peptidsynthese an ihren exozyklischen Aminen mit Aminosäuren beladen werden können. Dabei wurden *N*-Methylcarbamoyl-Nukleotide in der Gegenwart von Nitriten unter sauren Bedingungen in die entsprechenden *N*-Isocyanate umgewandelt, die mit ungeschützten Aminosäuren weiter reagieren konnten. Interessanterweise zeigte sich, dass die Eigenschaft Aminosäuren in Form von *N*-Carbamoyl Nukleosiden zu tragen nicht allein auf Adenosin beschränkt ist, sondern auch Guanosin (g^2G) und Cytidin (g^4C) mit Aminosäuren dekoriert werden können. Aminosäure-modifizierte Nukleotide könnten somit eine präbiotisch plausible Modifikation der RNA gewesen sein, die eine Co-Evolution von RNA und Peptiden ermöglichte.

Autorenbeitrag

Es wurden die Synthese für die Verbindungen **3**-**6** entwickelt und die Synthesen von **1h**, **4**, sowie der Oligonukleotide **ON3** und **ON4** durchgeführt. Ebenso wurden die präbiotischen Reaktionsbedingungen für die Herstellung der Aminosäure-modifizierten Oligonukleotide **ON1a-h** untersucht. Ebenso wurde die konsekutive Abfolge der Beladungs-, Kopplungs- und Spaltungsreaktionen, die in einem *One-Pot*-Experiment resultierten, etabliert.

Prebiotic Chemistry Hot Paper

How to cite:
International Edition: doi.org/10.1002/anie.202302360
German Edition: doi.org/10.1002/ange.202302360

Loading of Amino Acids onto RNA in a Putative RNA-Peptide World

Johannes N. Singer[+], Felix M. Müller[+], Ewa Węgrzyn, Christina Hölzl, Hans Hurmiz, Chuyi Liu, Luis Escobar, and Thomas Carell**

Abstract: RNA is a molecule that can both store genetic information and perform catalytic reactions. This observed dualism places RNA into the limelight of concepts about the origin of life. The RNA world concept argues that life started from self-replicating RNA molecules, which evolved toward increasingly complex structures. Recently, we demonstrated that RNA, with the help of conserved non-canonical nucleosides, which are also putative relics of an early RNA world, had the ability to grow peptides covalently connected to RNA nucleobases, creating RNA-peptide chimeras. It is conceivable that such molecules, which combined the information-coding properties of RNA with the catalytic potential of amino acid side chains, were once the structures from which life emerged. Herein, we report prebiotic chemistry that enabled the loading of both nucleosides and RNAs with amino acids as the first step toward RNA-based peptide synthesis in a putative RNA-peptide world.

Introduction

The essential biological process of translation converts a genotype into a phenotype. It is based on information-encoding messenger RNA (mRNA), the ribosome as the catalytic entity, and transfer RNAs (tRNAs) that carry specific amino acids and bind to the information-encoding units on the mRNA.[1,2] The attachment of an specific amino acid to its cognate tRNA (aminoacylation) requires the activation of the amino acid as an adenylate[3] and a dedicated aminoacyl tRNA synthetase,[4–6] which transfers the amino acid to the 3′-CCA end of the tRNA by catalyzing the formation of an ester bond. Although tRNAs are considered to be highly conserved structures,[7–10] the emergence of this complex process is still one of the main unsolved questions in origin of life research.

The main problem for chemical evolution is that complex chemical structures, such as mRNAs, ribosomes and amino acid loaded tRNAs, are required to generate a machinery that can efficiently form peptide bonds. But how could such complex systems have evolved in the absence of efficient peptide synthesis machines? To solve this chicken-and-egg conundrum, we recently postulated that RNA-amino acid conjugates could have been a starting point for evolving more complex systems.[11] We learned that RNA was capable of self-decorating with peptides with the help of non-canonical nucleosides,[12] that is, N^6-methylated derivatives of glycine- and threonine-modified N^6-carbamoyl adenosine, **1a** (g^6A)[13] and **1b** (t^6A).[14] In addition, **1a–b** are found in contemporary tRNAs in all three kingdoms of life[15,16] as potential relics of an early RNA world.[17,18] In aqueous solution, these compounds exhibited a good stability over a wide pH range at moderate temperatures due to the urea bond.[19,20] In the presence of a second non-canonical nucleoside, namely, 5-methylaminomethyl uridine (mnm^5U),[21,22] and when both nucleosides were placed in close proximity by RNA hybridization,[23] we observed the formation of peptides. This RNA-templated reaction established a primitive peptide synthesis cycle that afforded RNAs decorated with peptides. Although our system is far away from a translational machinery, it showed that RNA-peptide chimeras, having, in principle, information-encoding properties and an expanded repertoire of functional groups, could have existed in a prebiotic world.

We would like to emphasize that the obtained RNA-peptide chimeras are not considered by us to be the direct precursors of the contemporary ribosome. In this regard, seminal works by other research groups showed that amino acids could be connected to 5′-phosphate groups as acyl phosphate mixed anhydrides[24] or phosphoramidates,[25–31] and transferred to the ribose 2′/3′-hydroxy group[32] of terminal nucleotides in RNAs.[33–40]

The first step of the described peptide growth on RNA, however, required the efficient loading of amino acids onto N^6-carbamoyl adenosine nucleotides. Previously, we reported that amino acids, namely, Gly and Thr, could react

[*] J. N. Singer,[+] F. M. Müller,[+] E. Węgrzyn, Dr. C. Hölzl, H. Hurmiz, C. Liu, Dr. L. Escobar, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians-Universität (LMU) München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: luisescobar1992@hotmail.es
thomas.carell@lmu.de
Dr. C. Hölzl
Present address: AbbVie Deutschland GmbH & Co. KG
Knollstrasse 50, 67061 Ludwigshafen am Rhein (Germany)
[+] These authors contributed equally to this work.

with methylisocyanate (MIC) affording *N*-methylcarbamoyl amino acids in quantitative yields (Scheme 1a).[41] After nitrosation of the *N*-methylcarbamoyl amino acids, the obtained *N*-nitroso derivatives decomposed at pH 8 and 70 °C to give *N*-isocyanates, which reacted with adenosine to provide the required amino acid modified N^6-carbamoyl (**1a** and **1b**) and *O*-carbamoyl nucleosides. While this chemistry rested on prebiotically plausible starting materials, it had the drawback that the yields were sensitive to the exact reaction conditions and fairly low.

Herein, we show that a prebiotically more plausible loading of amino acids is possible with N^6-methylcarbamoyl adenosine **2** (Scheme 1b). The loading reaction was so efficient that it was also possible when **2** was embedded in RNA strands. The reported chemistry could be expanded to N^2-methylcarbamoyl guanosine **3** and N^4-methylcarbamoyl cytidine **4**. The obtained results suggest that the loading of amino acids onto RNAs constitutes an additional property, along with information-encoding, which needs to be considered in the context of prebiotic chemistry.

Results and Discussion

Prebiotically Plausible Synthesis of N^6-Methylcarbamoyl Adenosine

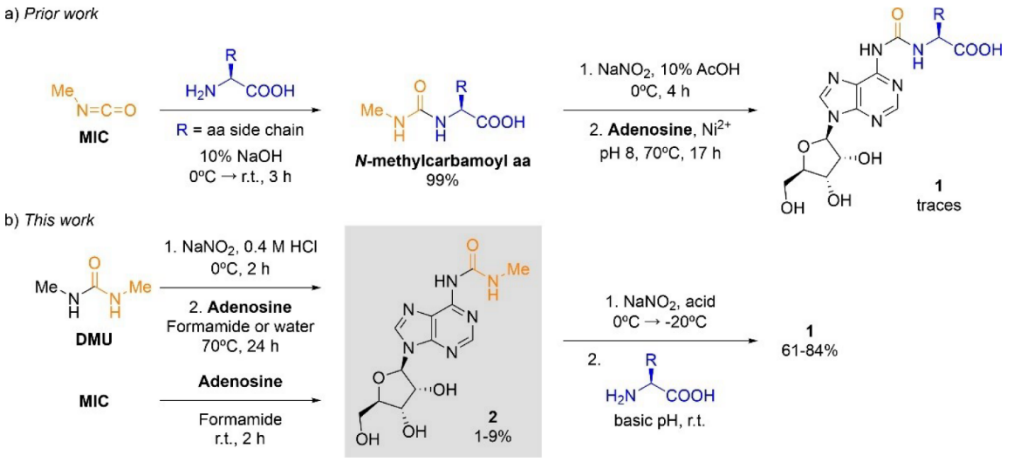
Initially, we investigated the formation of the urea-modified adenosine **2** from either 1,3-dimethylurea (DMU) or methylisocyanate (MIC) under prebiotically plausible reaction conditions (Scheme 1b). Both DMU and MIC are prebiotically plausible starting materials. MIC could be generated from CH_4 and HNCO by vacuum-UV light irradiation at 20 K.[42] Moreover, MIC was detected in the comet 67P/Churyumov-Gerasimenko[43] and in the protostar IRAS 16293-2422.[42,44] In turn, DMU was formed by continuous flow plasma discharge experiments from gas mixtures (N_2/CO/CO_2/H_2).[45] It is also possible that DMU was produced by the combination of MIC with methylamine, which was also present on the above-mentioned comet.[43]

In a 0.4 M HCl aqueous solution at 0 °C, the reaction of DMU with 1 equiv of $NaNO_2$[46–48] gave its *N*-nitroso derivative in 98 % yield.[49] Subsequently, the addition of adenosine to a formamide[43,50] or an aqueous solution, containing 50 equiv of the *N*-nitroso derivative of DMU, followed by heating at 70 °C afforded the N^6-carbamoylated adenosine **2** in up to 2 % yield. The high-performance liquid chromatography—mass spectrometry (HPLC-MS) analyses of the crude reaction mixtures showed the formation of **2**, as well as other adenosine derivatives having either methyl or *N*-methylcarbamoyl substituents, or both (Figure S1a–b). Most likely, the isomeric products of **2** had the *N*-methylcarbamoyl substituent at one of the ribose hydroxy groups, that is, *O*-methylcarbamoyl adenosine derivatives. In formamide solution at room temperature (r.t.), the reaction of adenosine with 5.5 equiv of MIC yielded **2** in a maximum amount of 9 % (Figure S2). Taken together, these results showed that two prebiotically plausible synthetic pathways could provide the urea-modified adenosine **2**.

Next, we probed the stability of the N^6-methylcarbamoyl adenosine **2** at 70 °C in aqueous solution at different pH values within the range 6–9.5. Under all these conditions, the HPLC-MS analyses, after 24 h, indicated that **2** was stable in aqueous solution. Thus, heating at 70 °C the above obtained crude reaction mixtures for 24 h in 50 mM borate buffer pH 9.5 led to the hydrolysis of some of the *O*-methylcarbamoyl adenosine derivatives, whereas **2** persisted in the aqueous solution (Figure S3).

Prebiotic Loading of Amino Acids onto Nucleosides

Based on our previous study on the prebiotic synthesis of amino acid modified N^6-carbamoyl adenosine **1** from *N*-methylcarbamoyl amino acids (Scheme 1a),[41] we envisaged that the analogous nitrosation of the urea-modified adenosine **2**, followed by its conversion into the corresponding N^6-



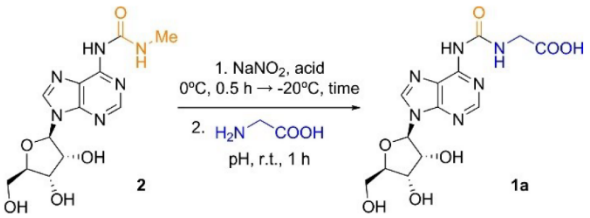
Scheme 1. Prebiotic synthesis of amino acid modified N^6-carbamoyl adenosine nucleosides (**1**) from: a) *N*-methylcarbamoyl amino acids and b) N^6-methylcarbamoyl adenosine (**2**). aa = amino acid.

isocyanate could also lead to the formation of **1** (Scheme 1b). First, we optimized the conditions for the reaction of **1** with Gly using HPLC-MS analyses (Table 1 and Tables S1–S4). In the first reaction step, we performed the nitrosation of **2** with 12.5 equiv of $NaNO_2$ in an acidic aqueous solution. Among the acids tested, 5 % H_3PO_4[51] in water gave the best results, indicating that the formation of the *N*-nitroso derivative of **2** required rather harsh acidic conditions (pH < 1.5). Next, we investigated the reaction under freezing conditions, and we found that the yield strongly improved at −20 °C for 22 h. Most likely, under these conditions, the reactants were excluded from the growing ice crystals and concentrated in the interstitial (eutectic) phase.[52–55] In the second reaction step, the thawed aqueous solution, containing initially **2** and $NaNO_2$, was adjusted to near neutral or basic pH. At r.t. and in the presence of 10 equiv of Gly, the loading of the amino acid onto **2** was very efficient at pH 9.5,[56] affording **1a** (g^6A) in a remarkable 84 % yield. When the amino acid was added in the first step, the nitrosation reaction, the amino acid modified N^6-carbamoyl adenosine was also obtained, yet in lower yield (Figure S6 and Table S5). Note that the prebiotic formation of **1a** required dramatic changes in the pH of the aqueous solution. In this respect, it is worth mentioning that the second reaction step, coupling of the amino acid, also occurred under slightly acidic conditions in 5 % yield (Table 1). We also attempted the loading of Gly using the N^6-methylated version of **2**. However, under the optimized reaction conditions, the yield of the expected nucleoside product (m^6g^6A) was reduced to 8 % (Figure S9).

To evaluate the scope of the prebiotic synthesis of amino acid modified N^6-carbamoyl adenosine nucleosides **1**, we carried out the reaction with a series of L-amino acids, having aliphatic (Ala, Pro), aromatic (Phe), neutral polar (Thr) and ionizable (Asp, Lys) side chains (Table 2).[57] Under the reaction conditions optimized above for **1a** (g^6A), we observed, in all cases, the formation of the nucleoside products **1b–g** in yields between 61–78 % (Figures S4–S5). We also detected, in all cases, the presence of ca. 10 % adenosine and the remaining starting material **2** (Figure 1a). Interestingly, the reaction of **2** with Lys, with its α- and ε-amino groups, provided a selectivity of 80 : 20 for the product ***α*-1f** (k^6A) versus ***ε*-1f**.

Next, we studied the loading of an amino nitrile onto N^6-methylcarbamoyl adenosine **2**. Amino nitriles were suggested as prebiotically plausible precursors for

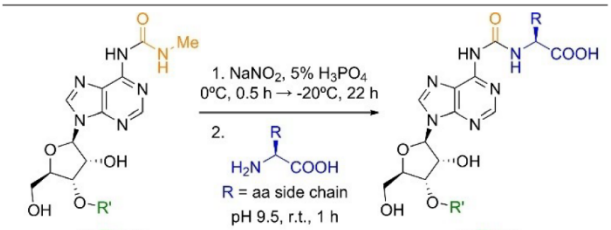
Table 1: Optimization of reaction conditions using N^6-methylcarbamoyl adenosine (**2**) and Gly.



Reaction step 1		Reaction step 2	Yield [%][a]
Acid	Freezing time [h]	pH	
neat acetic acid			n.d.
neat formic acid			18
1 M HCl	22	9.5	78
1 M H_2SO_4			64
5% H_3PO_4			**84**
	0		4
5% H_3PO_4	3	9.5	42
	22		84
		6.2	5
		7.4	34
5% H_3PO_4	22		
		8.6	60
		9.5	84

[a] Yields determined by HPLC analysis using the calibration curve of **1a** (Figure S10). n.d. = not detected.

Table 2: Amino acid scope using N^6-methylcarbamoyl adenosine (**2**) and RNA oligonucleotide (**ON2**).



Amino acid	Nucleoside product	Yield [%][a]	RNA product	Yield [%][a,b]
Gly	**1a** (g^6A)	84	**ON1a**	18
Thr	**1b** (t^6A)	61	**ON1b**	18
Phe	**1c** (f^6A)	78	**ON1c**	23
Pro	**1d** (p^6A)	74	**ON1d**	15
Asp	**1e** (d^6A)	64	**ON1e**	10
Lys	***α*-1f** (k^6A)	70	***α*-ON1f**	10
	***ε*-1f**	16	***ε*-ON1f**	n.d.
Ala	**1g** (a^6A)	75	**ON1g**	16
Gly_{CN}[c]	**1h** ($g_{CN}{}^6A$)	26	**ON1h**	n.d.

[a] Yields determined by HPLC analysis using the calibration curves of reference compounds (Figures S10 and S19). [b] Reactions performed in the presence of GdmCl. [c] The structure of the Gly derivative contained a nitrile group instead of a carboxylic acid. A = adenosine. n.d. = not detected.

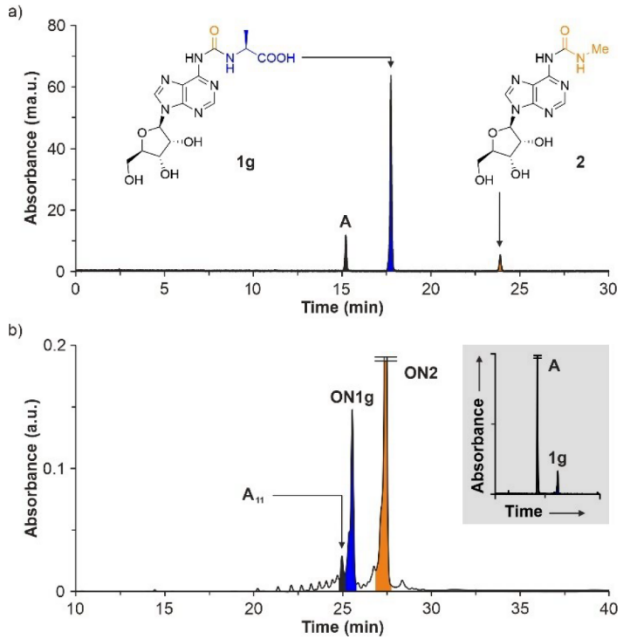
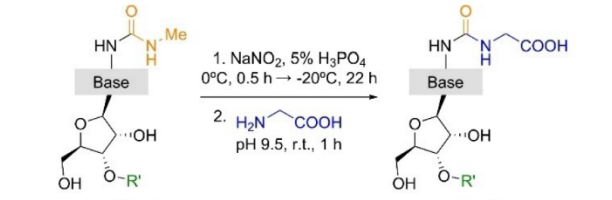


Figure 1. HPLC chromatograms of the crude reaction mixtures of Ala with: a) **2** and b) **ON2**. In b, inset shows enzymatic digestion of **ON1g**. A = adenosine and $\mathbf{A_{11}}$ = 5′-AAAAAAAAAAA-3′.

peptides.[11,58–61] Indeed, the reaction of **2** with the nitrile derivative of Gly, that is, Gly_{CN}, generated the nucleoside product **1h** ($g_{CN}{}^6A$) in 26 % yield (Table 2).

Encouraged by the reaction of **2** with amino acids, we explored the loading of Gly with other urea-modified nucleosides: N^2-methylcarbamoyl guanosine **3** and N^4-methylcarbamoyl cytidine **4** (Table 3). We found that the addition

Table 3: Base scope using different nucleosides and RNA oligonucleotides.



Base	Nucleoside	Product	Yield [%][a]	RNA	Product	Yield [%][a,b]
A	**2**	**1a** (g^6A)	84	**ON2'**	**ON1a'**	18
G	**3**	**5** (g^2G)	78	**ON3**	**ON5**	35
C	**4**	**6** (g^4C)	58	**ON4**	**ON6**	5

[a] Yields determined by HPLC analysis using the calibration curves of reference compounds (Figures S10, S11 and S19). [b] Reactions performed in the presence of GdmCl. A = adenosine; C = cytidine, G = guanosine, U = uridine and X = xanthosine.

of Gly to an aqueous solution of **3** at pH 9.5, after its nitrosation, afforded **5** (g^2G) in 78 % yield (Figure 2a). Similarly, the combination of **4** with Gly gave **6** (g^4C) in 58 % yield (Figure S8). These results suggested that the putative RNA-peptide synthesis might not be limited to RNAs containing the amino acid modified N^6-carbamoyl adenosine **1**, although the non-canonical nucleosides **5** (g^2G) and **6** (g^4C) were so far not found in contemporary RNAs.[62]

All together, these data showed that *N*-methylcarbamoyl nucleosides, **2**–**4**, could be efficiently loaded with amino acids and amino nitriles in aqueous solution under prebiotically plausible conditions to form amino acid modified *N*-carbamoyl nucleosides.

Prebiotic Loading of Amino Acids onto RNA

We investigated whether the loading of amino acids was possible when N^6-methylcarbamoyl adenosine **2** was incorporated into RNA. For these experiments, and the subsequent RNA-based amino acid transfer, we synthesized the phosphoramidite derivative of **2** in four synthetic steps with a 35 % overall yield (Scheme S19). Then, we used solid-phase RNA synthesis to incorporate the urea-modified nucleoside at the 5′-end of the 11-mer homo-A RNA strand **ON2** (Table 2).

We started with the homo-A RNA strand **ON2** and analyzed the loading of a series of amino acids. For example, when we treated **ON2** with 500 equiv of $NaNO_2$ in 5 % H_3PO_4 at −20 °C and then, we added, at r.t., an excess of Ala, followed by basification to pH 9.5, the HPLC chroma-

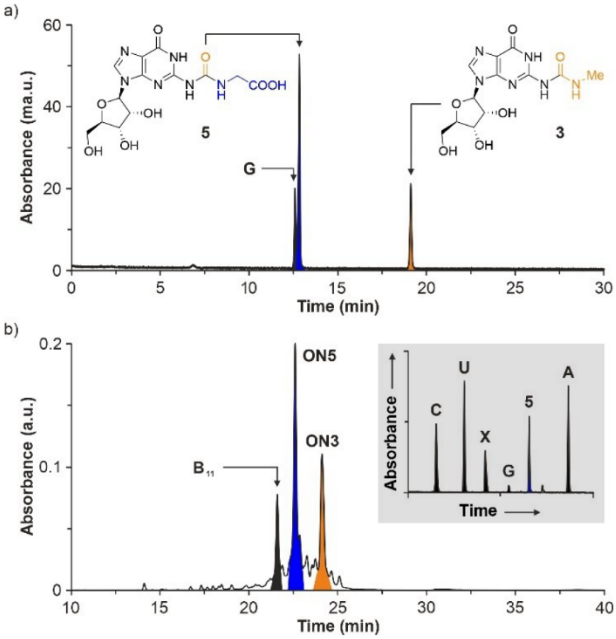


Figure 2. HPLC chromatograms of the crude reaction mixtures of Gly with: a) **3** and b) **ON3**. In b, inset shows enzymatic digestion of **ON5**. A = adenosine; C = cytidine; G = guanosine; U = uridine; X = xanthosine and $\mathbf{B_{11}}$ = 5′-GAUCXCUXUAC-3′.

togram of the crude reaction mixture, as well as the matrix-assisted laser desorption/ionization—time-of-flight (MALDI-TOF) mass spectrum of the purified compound revealed the formation of the RNA strand **ON1g** with a terminal a^6A nucleotide (Figure S14). Next to the remaining starting material, we noted the formation of three major side products. Using MALDI-TOF MS, we assigned the degradation products to 10-mer and 11-mer RNA strands, $\mathbf{A_{10}}$ and $\mathbf{A_{11}}$, that had lost the terminal urea-modified nucleoside and the N^6-methylcarbamoyl substituent, respectively. We also detected a 10-mer RNA strand with a terminal a^6A nucleotide, $\mathbf{A_9\text{-}a^6A}$. It seems that the nitrosation reaction provoked the partial degradation of the RNA strands. However, the addition of 100 mM of guanidinium chloride (GdmCl) salt reduced the side reactions dramatically (Figure 1b), potentially by formation of salt-bridges between the (nitrosated)[63,64] guanidinium cation and the phosphodiester backbone.[65] Other salts, such as NaCl and $NaClO_4$, had little effect on the RNA degradation (Figure S14).

In order to further prove the loading of Ala onto the RNA strand **ON2**, we performed an enzymatic digestion of the isolated RNA product **ON1g** into the nucleosides, and analyzed the digest with the help of the corresponding nucleoside reference compounds by HPLC-MS.[66] We treated **ON1g** with a nucleoside digestion mixture at 37 °C for 2 h. Analysis of the obtained nucleoside mixture showed two peaks corresponding to **1g** (a^6A) and adenosine in a ca. 1:10 ratio (Figure 1b), which was in agreement with the nucleotide composition of the digested RNA strand.

In the presence of GdmCl and for all the amino acids studied, we observed the formation of the RNA strands **ON1a–g** in 10–23 % yield (Table 2 and Figures S15–S16). It is worth mentioning that the reaction of Lys with the RNA strand **ON2** yielded predominantly the α-regioisomer, ***α*-ON1f**, of the k^6A nucleotide. For the loading with the amino nitrile Gly_{CN}, we did not detect the expected RNA product **ON1h**.

In analogy to the loading of amino acids onto different nucleosides, we evaluated the loading of Gly with three 11-mer RNA strands, **ON2'–ON4** (Table 3), having all four canonical nucleotides, but bearing distinct *N*-methylcarbamoyl nucleotides at the 5′-end. The RNA strands were synthesized from the corresponding phosphoramidite derivatives of **2–4** (Schemes S19–S21). In all cases, the reactions of the three RNA strands with Gly, under the conditions described above with GdmCl, gave the respective products, **ON1a'**, **ON5–ON6**, in 5–35 % yield (Table 3, Figure 2b and Figure S12). MALDI-TOF MS analyses of the obtained RNA products showed the successful loading of Gly onto the parent RNA strands. In addition, we detected, after the enzymatic digestion of the RNA products, the partial conversion of guanosine to xanthosine by nitrosative deamination.[67,68]

Moreover, we investigated the loading of Phe onto an RNA strand **ON2"** bearing two N^6-methylcarbamoyl adenosine nucleotides, placed at the 5′-end and an internal position of the sequence (Scheme S8). Using similar reaction conditions to those described above, we observed the formation of the double-loaded RNA strand **ON1c'** as the major species in the crude reaction mixture (Figure S13).

Collectively, these data showed that RNA strands containing *N*-methylcarbamoyl nucleotides at either terminal or internal positions, or both, could be loaded with amino acids in a sufficient extent to perform RNA-based amino acid transfer (see below).

RNA-Based Amino Acid Transfer Following the Loading Reaction

Finally, we wanted to demonstrate that the prebiotic loading of an amino acid onto an RNA strand, functionalized with a urea-modified adenosine nucleotide at the 5′-end, could undergo RNA-based amino acid transfer (Figure 3a). The loading reaction provides a terminal amino acid modified N^6-carbamoyl adenosine (donor) as the basis for the synthesis reaction. For the amino acid transfer, we hybridize the obtained donor RNA strand with a complementary acceptor RNA strand, containing a 5-aminomethyl uridine at the 3′-end, forming a duplex. Activation of the carboxylic acid in the donor RNA strand leads to the formation of a hairpin intermediate, which is opened at elevated temperatures by urea cleavage, transferring the loaded amino acid to the acceptor RNA strand. The obtained acceptor RNA strand can be involved in subsequent reactions, enabling peptide growth on RNA.[11] Since the nitrosation reaction converts guanosine to xanthosine, we employed a binary code of adenosine and uridine nucleotides in the initial RNA strand **ON7**. We also selected Phe for the experiment because of its high reaction efficiency (Table 2).[11]

In this experiment, the individual reaction steps were monitored by HPLC (Figure 3b). In the first step, the prebiotic loading of Phe onto the RNA strand **ON7**, under the reaction conditions described above with GdmCl, gave the donor RNA strand **ON8** in ca. 38 % yield. In the second step, we combined the isolated donor RNA strand **ON8** with an equimolar amount of the acceptor RNA strand **ON9** in 100 mM 2-(*N*-morpholino)ethanesulfonate (MES) buffer pH 6 and 1 M NaCl. The RNA strand **ON9** contained 2′-methoxy nucleotides, which are found in contemporary ribosomal RNA (rRNA)[69] and minimize degradation. The addition of 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM·Cl) to the solution mixture of **ON8** and **ON9** at 15 °C led to the formation of the hairpin intermediate **ON10** in ca. 40 % yield. Since the melting temperature of the reactive duplex **ON8·ON9** was ca. 35 °C in the buffered aqueous solution (Figure S20a–b), it was assembled almost quantitatively in the reaction mixture. As reported previously, the activation of the carboxylic acid could also be performed prebiotically with methylisonitrile (MeNC) and 4,5-dicyanoimidazole (DCI).[11,70] In the third step, we warmed the aqueous solution of the hairpin intermediate **ON10** to 90 °C at pH 6 to cleave the urea bond,[71,72] which furnished the RNA-peptide chimera **ON11** in ca. 48 % yield. In contrast to our previous experiments with donor RNA strands containing amino acid modified N^6-methyl-N^6-carbamoyl adenosine at

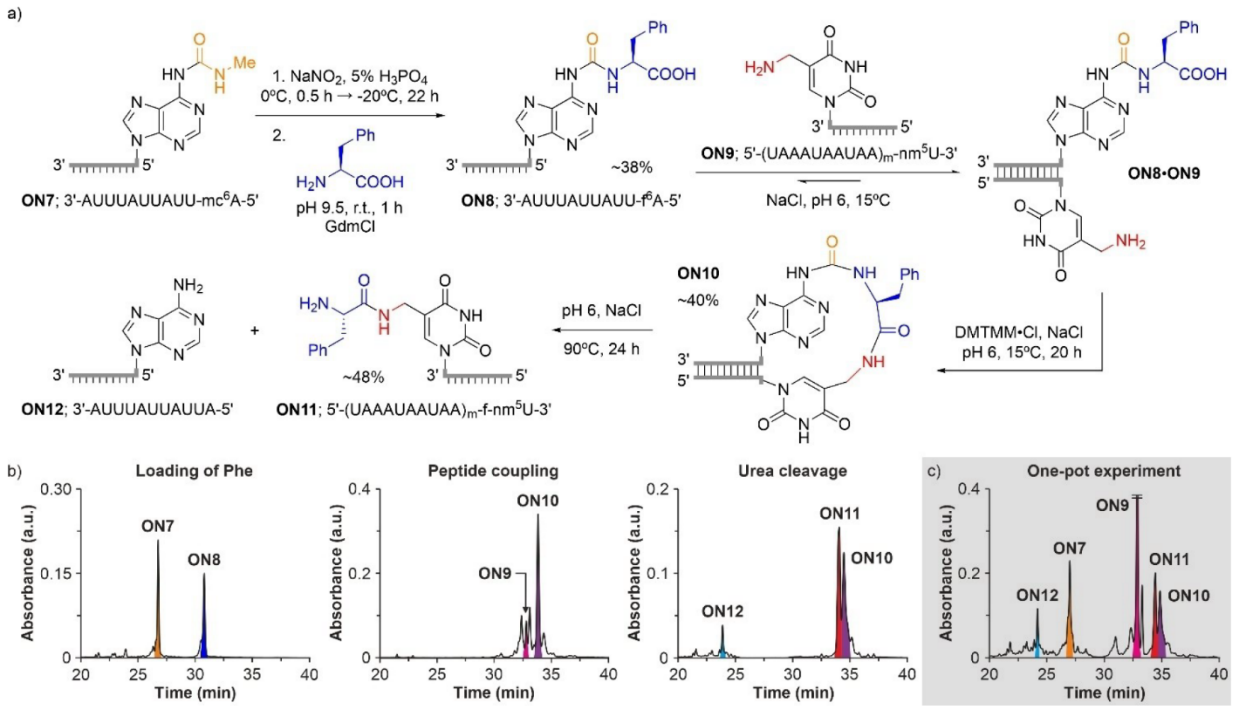


Figure 3. a) Reaction scheme for RNA-based amino acid transfer: loading of Phe, peptide coupling and urea cleavage. HPLC chromatograms of the crude reaction mixtures of: b) the individual steps and c) the one-pot experiment. mc^6A = N^6-methylcarbamoyl adenosine; nm^5U = 5-aminomethyl uridine; A = adenosine and U = uridine. m (subscript) indicates that the RNA strands contained 2′-methoxy nucleotides.

the 5′-end,[11] the absence of the N^6-methyl substituent in **ON8** prevented the formation of a hydantoin side-product in the urea cleavage step.[73]

Finally, we performed the loading of Phe and the RNA-based amino acid transfer in a one-pot experiment (Figure 3c). The crude reaction mixtures were filtered after each step to remove the excess of salts and low-molecular weight compounds but retaining the RNA strands. Indeed, analysis of the one-pot experiment by HPLC showed the formation of the RNA-peptide chimera **ON11** in ca. 11 % overall yield (Figure 3c), which was in line with the result obtained for the three individual reaction steps (ca. 7 %).

Conclusion

Recently, we reported that RNA with the help of two non-canonical nucleosides, that is, amino acid modified N^6-carbamoyl adenosine and 5-methylaminomethyl uridine, gained the ability to carry amino acids at certain nucleobase positions. The subsequent transfer of amino acids from one nucleobase to another enabled the synthesis of small peptides on RNA, affording RNA-peptide chimeras. The ability of RNA to decorate itself with the help of non-canonical nucleosides constitutes an additional function, which places RNA in a prime position for enabling the chemical evolution of life.

In this work, we described a new method to load RNA strands with amino acids. While this loading reaction was so far only possible with adenosine in low yields, we could now improve the loading efficiencies to yields up to 84 %. Most importantly, the loading reaction was also possible in RNA strands with acceptable efficiencies between 10–23 %. These results suggested that RNA-peptide chimeras are molecules that could have formed under early Earth conditions. In addition, we expanded the loading reaction to *N*-methylcarbamoyl guanosine and cytidine, and to their corresponding RNA strands. A caveat of the reported process is that we detected deamination of guanosine under the loading conditions. In contrast, the loading with Phe and the subsequent RNA-based amino acid transfer was found to be surprisingly efficient. This observation suggests that loading preferences, for yet unknown reasons, might exist. Thus, we believe that different amino acids might have distinct preferences to be loaded onto specific sequences. This is the prerequisite for RNA-encoded peptide synthesis. Our study also showed that the absence of a N^6-methyl substituent in N^6-methylcarbamoyl adenosine enhanced the loading efficiency and prevented the formation of hydantoin side-products.

The reported chemistry also required big changes in the pH of the aqueous solution. While the nitrosation of the N^6-methylcarbamoyl nucleoside took place under acidic conditions (5 % H_3PO_4), the reaction with the amino acid was performed at the optimal pH 9.5. Therefore, we need to assume that the first reaction step occurred in an acidic environment, such as an acidic pond, from which a mixture of the nitrosated nucleoside and the amino acids flow out,

e.g. along a carbonate containing river bed or into a carbonate dominated second basin. Repetitive loading and RNA-based peptide synthesis cycles would require continuous fluctuations of the pH between acidic and moderately basic in a dynamic process. However, it is worth mentioning that a low amino acid loading efficiency is not necessarily a disadvantage for an RNA-peptide world, since a high turnover of canonical nucleotides into *N*-carbamoyl derivatives would lead to the loss of the information-encoding property of RNA, hindering replication. We are just at the beginning of learning about the possibilities that are offered by RNA-peptide conjugates and how they could be integrated into coding schemes, e.g. via Hoogsteen-type base-pairing.

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 program under grant agreement No. 741912 (EPiR). L. E. thanks the Alexander von Humboldt Foundation for a postdoctoral fellowship (ESP 1214218 HFST-P). We also thank M.Sc. Johann de Graaff for helpful discussions. 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 Supporting Information of this article.

Keywords: Amino Acids · Nucleosides · Origin of Life · Prebiotic Chemistry · RNA-Peptide World

[1] V. Ramakrishnan, *Cell* **2002**, *108*, 557–572.
[2] T. A. Steitz, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253.
[3] M. B. Hoagland, E. B. Keller, P. C. Zamecnik, *J. Biol. Chem.* **1956**, *218*, 345–358.
[4] M. Delarue, *Curr. Opin. Struct. Biol.* **1995**, *5*, 48–55.
[5] M. Ibba, D. Söll, *Annu. Rev. Biochem.* **2000**, *69*, 617–650.
[6] R. B. Ganesh, S. J. Maerkl, *Front. Bioeng. Biotechnol.* **2022**, *10*, 918659.
[7] M. Di Giulio, *J. Mol. Evol.* **1997**, *45*, 571–578.
[8] K. Tamura, *J. Biosci.* **2011**, *36*, 921–928.
[9] A. Harish, G. Caetano-Anollés, *PLoS One* **2012**, *7*, e32776.
[10] J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, L. D. Williams, *Chem. Rev.* **2020**, *120*, 4848–4878.
[11] F. Müller, L. Escobar, F. Xu, E. Węgrzyn, M. Nainytė, T. Amatov, C. Y. Chan, A. Pichler, T. Carell, *Nature* **2022**, *605*, 279–284.
[12] T. Carell, C. Brandmayr, A. Hienzsch, M. Muller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110–7131.
[13] M. P. Schweizer, K. McGrath, L. Baczynskyj, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1046–1052.
[14] L. Perrochia, E. Crozat, A. Hecker, W. Zhang, J. Bareille, B. Collinet, H. van Tilbeurgh, P. Forterre, T. Basta, *Nucleic Acids Res.* **2013**, *41*, 1953–1964.
[15] H. Grosjean, E. Westhof, *Nucleic Acids Res.* **2016**, *44*, 8020–8040.
[16] J. Beenstock, F. Sicheri, *Nucleic Acids Res.* **2021**, *49*, 10818–10834.
[17] W. Gilbert, *Nature* **1986**, *319*, 618.
[18] G. F. Joyce, *Nature* **2002**, *418*, 214–221.
[19] K. R. Lynn, *J. Phys. Chem.* **1965**, *69*, 687–689.
[20] J. Akester, J. Cui, G. Fraenkel, *J. Org. Chem.* **1997**, *62*, 431–434.
[21] F. V. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186–1191.
[22] A. Kitamura, M. Nishimoto, T. Sengoku, R. Shibata, G. Jäger, G. R. Björk, H. Grosjean, S. Yokoyama, Y. Bessho, *J. Biol. Chem.* **2012**, *287*, 43950–43960.
[23] M. Nainytė, F. Müller, G. Ganazzoli, C. Y. Chan, A. Crisp, D. Globisch, T. Carell, *Chem. Eur. J.* **2020**, *26*, 14856–14860.
[24] P. Berg, *J. Biol. Chem.* **1958**, *233*, 608–611.
[25] R. Lohrmann, L. E. Orgel, *Nature* **1973**, *244*, 418–420.
[26] J. L. Shim, R. Lohrmann, L. E. Orgel, *J. Am. Chem. Soc.* **1974**, *96*, 5283–5284.
[27] M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed.* **2015**, *54*, 14564–14569.
[28] H. Griesser, M. Bechthold, P. Tremmel, E. Kervio, C. Richert, *Angew. Chem. Int. Ed.* **2017**, *56*, 1224–1228.
[29] H. Griesser, P. Tremmel, E. Kervio, C. Pfeffer, U. E. Steiner, C. Richert, *Angew. Chem. Int. Ed.* **2017**, *56*, 1219–1223.
[30] M. Räuchle, G. Leveau, C. Richert, *Eur. J. Org. Chem.* **2020**, 6966–6975.
[31] B. Jash, P. Tremmel, D. Jovanovic, C. Richert, *Nat. Chem.* **2021**, *13*, 751–757.
[32] N. S. M. D. Wickramasinghe, M. P. Staves, J. C. Lacey, Jr., *Biochemistry* **1991**, *30*, 2768–2772.
[33] M. Illangasekare, G. Sanchez, T. Nickles, M. Yarus, *Science* **1995**, *267*, 643–647.
[34] K. Tamura, P. Schimmel, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8666–8669.
[35] K. Tamura, P. Schimmel, *Science* **2004**, *305*, 1253.
[36] K. Tamura, P. R. Schimmel, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13750–13752.
[37] N. V. Chumachenko, Y. Novikov, M. Yarus, *J. Am. Chem. Soc.* **2009**, *131*, 5257–5263.
[38] R. M. Turk, N. V. Chumachenko, M. Yarus, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4585–4589.
[39] R. M. Turk, M. Illangasekare, M. Yarus, *J. Am. Chem. Soc.* **2011**, *133*, 6044–6050.
[40] L. F. Wu, M. Su, Z. W. Liu, S. J. Bjork, J. D. Sutherland, *J. Am. Chem. Soc.* **2021**, *143*, 11836–11842.
[41] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946.
[42] 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.
[43] F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Krüger, L. Le Roy, A. J.

MacDermott, S. McKenna-Lawlor, U. J. Meierhenrich, G. M. M. Caro, F. Raulin, R. Roll, A. Steele, H. Steininger, R. Sternberg, C. Szopa, W. Thiemann, S. Ulamec, *Science* **2015**, *349*, aab0689.
[44] R. Martín-Doménech, V. M. Rivilla, I. Jiménez-Serra, D. Quénard, L. Testi, J. Martín-Pintado, *Mon. Not. R. Astron. Soc.* **2017**, *469*, 2230–2234.
[45] M. N. Heinrich, W. R. Thompson, C. Sagan, *Bull. Am. Astron. Soc.* **1991**, *23*, 1211.
[46] Y. L. Yung, M. B. McElroy, *Science* **1979**, *203*, 1002–1004.
[47] R. L. Mancinelli, C. P. McKay, *Origins Life Evol. Biospheres* **1988**, *18*, 311–325.
[48] S. Ranjan, Z. R. Todd, P. B. Rimmer, D. D. Sasselov, A. R. Babbin, *Geochem. Geophys. Geosyst.* **2019**, *20*, 2021–2039.
[49] G. W. Breton, M. Turlington, *Tetrahedron Lett.* **2014**, *55*, 4661–4663.
[50] R. Saladino, C. Crestini, S. Pino, G. Costanzo, E. Di Mauro, *Phys. Life Rev.* **2012**, *9*, 84–104.
[51] D. J. Ritson, S. J. Mojzsis, J. D. Sutherland, *Nat. Geosci.* **2020**, *13*, 344–348.
[52] T. Vajda, *Cell. Mol. Life Sci.* **1999**, *56*, 398–414.
[53] C. Menor-Salván, M. R. Marín-Yaseli, *Chem. Soc. Rev.* **2012**, *41*, 5404–5415.
[54] K. Kitada, Y. Suda, N. Takenaka, *J. Phys. Chem. A* **2017**, *121*, 5383–5388.
[55] N. Kitadai, S. Maruyama, *Geosci. Front.* **2018**, *9*, 1117–1153.
[56] B. T. Golding, C. Bleasdale, J. McGinnis, S. Müller, H. T. Rees, N. H. Rees, P. B. Farmer, W. P. Watson, *Tetrahedron* **1997**, *53*, 4063–4082.
[57] M. Kimura, S. Akanuma, *J. Mol. Evol.* **2020**, *88*, 372–381.
[58] S. Islam, M. W. Powner, *Chem* **2017**, *2*, 470–501.
[59] P. Canavelli, S. Islam, M. W. Powner, *Nature* **2019**, *571*, 546–549.
[60] C. S. Foden, S. Islam, C. Fernandez-Garcia, L. Maugeri, T. D. Sheppard, M. W. Powner, *Science* **2020**, *370*, 865–869.
[61] J. Singh, D. Whitaker, B. Thoma, S. Islam, C. S. Foden, A. E. Aliev, T. D. Sheppard, M. W. Powner, *J. Am. Chem. Soc.* **2022**, *144*, 10151–10155.
[62] S. P. Dutta, C. I. Hong, G. P. Murphy, A. Mittelman, G. B. Chheda, *Biochemistry* **1975**, *14*, 3144–3151.
[63] A. F. McKay, *Chem. Rev.* **1952**, *51*, 301–346.
[64] I. Fernández, P. Hervés, M. Parajó, *J. Phys. Org. Chem.* **2008**, *21*, 713–717.
[65] J. Vušurović, E.-M. Schneeberger, K. Breuker, *ChemistryOpen* **2017**, *6*, 739–750.
[66] 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.
[67] J. Xu, N. J. Green, D. A. Russell, Z. Liu, J. D. Sutherland, *J. Am. Chem. Soc.* **2021**, *143*, 14482–14486.
[68] S. Mair, K. Erharter, E. Renard, K. Brillet, M. Brunner, A. Lusser, C. Kreutz, E. Ennifar, R. Micura, *Nucleic Acids Res.* **2022**, *50*, 6038–6051.
[69] W. A. Decatur, M. J. Fournier, *Trends Biochem. Sci.* **2002**, *27*, 344–351.
[70] Z. W. Liu, L. F. Wu, J. F. Xu, C. Bonfio, D. A. Russell, J. D. Sutherland, *Nat. Chem.* **2020**, *12*, 1023–1028.
[71] M. Hutchby, C. E. Houlden, J. G. Ford, S. N. G. Tyler, M. R. Gagné, G. C. Lloyd-Jones, K. I. Booker-Milburn, *Angew. Chem. Int. Ed.* **2009**, *48*, 8721–8724.
[72] A. Ohkubo, R. Kasuya, K. Miyata, H. Tsunoda, K. Seio, M. Sekine, *Org. Biomol. Chem.* **2009**, *7*, 687–694.
[73] G. Danger, R. Plasson, R. Pascal, *Chem. Soc. Rev.* **2012**, *41*, 5416–5429.

Manuscript received: February 15, 2023
Accepted manuscript online: March 7, 2023
Version of record online: ■■, ■■

4. Unveröffentlichte Ergebnisse

Nicht-kanonische RNA-Modifikationen sind ein weit verbreiteter Bestandteil von tRNAs und rRNAs.[186] Sie werden über enzymatische Prozesse post-transkriptionell eingefügt, dienen der korrekten Ausprägung von Sekundärstrukturen und spielen eine wichtige Rolle bei Codon-Anticodon-Wechselwirkungen. Es wird vermutet, dass eine nicht-codierte Peptidsynthese der Translation vorausging und die Co-Existenz von RNA und Peptiden einen evolutionären Vorteil generiert haben könnte.[128][207][212] Studien zur *in vitro* Evolution aktiver Ribozyme beziehen sich aufgrund biochemischer Limitationen hauptsächlich auf kanonische Nukleotide. Die Substitution bestimmter Nukleotide in bekannten Ribozymen bietet jedoch die Möglichkeit, den Einfluss modifizierter Nukleotide zu untersuchen. In den hier beschriebenen publizierten Arbeiten konnte gezeigt werden, dass Aminosäure-modifizierte Nukleotide unter präbiotisch plausiblen Bedingungen hergestellt werden können und anschließend eine RNA-basierte Peptidsynthese stattfinden kann, bei der die Aminosäure zwischen den Nukleobasen zweier komplementärer RNA-Stränge übertragen wird. Dieser Prozess erfolgt in zwei Schritten. Zunächst wird die Carbonsäure aktiviert, wodurch sie mit einem Amin zu einem Amid reagiert und anschließend kann die Harnstoffbindung gespalten werden (**siehe Kapitel 3.2 und 3.3**). Im ersten Schritt entsteht ein Zwischenprodukt, in dem zwei RNA-Stränge über eine Peptidbrücke zwischen den Nukleobasen verknüpft sind. Interessanterweise erhöht sich durch diese Verknüpfung die Schmelztemperatur des RNA-Duplex und führt damit zu einer höheren Stabilität, analog zu einem Haarnadel-Motiv, bei dem ein RNA-Strang selbstkomplementär ist. Es ist daher denkbar, dass Aminosäure-modifizierte RNA nicht nur der Ursprung einer Peptidsynthese gewesen sein könnte, sondern auch ein Motor der Evolution durch strukturgebende Eigenschaften gewesen sein könnte. Ziel dieser unveröffentlichten Studien war es, neben der Verknüpfung über Aminosäure-modifiziertes N^6-Carbamoyladenosin auch die Verknüpfung zweier Oligonukleotide über N^4-Carbamoylcytidin- und N^2-Carbamoylguanosin-Derivate zu zeigen und anschließend den Einfluss des Haarnadel-Motivs auf die katalytische Aktivität anhand des Hammerhead Ribozyms zu untersuchen.

4.1 Nicht-kanonische Nukleobasenmodfikationen als Haarnadelmotiv

4.1.1 Ergebnisse und Diskussion

Zuvor konnte gezeigt werden, dass neben Aminosäure-modifiziertem N^6-Carbamoyladenosin (g^6A) auch die Derivate von N^4-Carbamoylcytidin (g^4C) und N^2-Carbamoylguanosin (g^2G) unter präbiotischen Bedingungen entstehen können. Die Beladung mit Glycin war sowohl auf die

Nukleoside als auch auf das auch auf Oligonukleotide in Ausbeuten von 5-35% möglich. Um die Ausbildung eines Haarnadel-Motivs der jeweiligen Modifikationen zu untersuchen, wurden die Oligonukleotide **ON1**-**ON3** mit den entsprechenden Nukleoside g^6A, g^4C und g^2G am 5′-Ende synthetisiert. Dazu wurden verschiedene synthetische Ansätze gewählt. Zum einen können die modifizierten Oligonukleotide über Phosphoramidit-Chemie oder alternativ über die präbiotische Beladung von RNA mit Aminosäuren ausgehend von den *N*-Methylcarbamoyl-Nukleotiden erhalten werden. Zuvor wurde berichtet, dass nitrosierende Bedingungen zu einer Desaminierung von Guanosin hin zu Xanthosin führen, wodurch die Eigenschaft eine Basenpaarung mit Cytosin ausbilden zu können, verloren geht. Aus diesem Grund wurde für diese Untersuchungen eine 8-mer-Sequenz gewählt, die aus A, C, U und dem entsprechenden *N*-Carbamoyl-Nukleotid auf der Seite des Donorstrangs besteht. **ON1** wurde durch Inkorporation des g^6A-Phoshoramidit mittels Festphasen-RNA-Synthese erhalten. Aufgrund der vergleichsweisen hohen Ausbeute von bis zu 35% bei der Beladung von N^2-Methylcarbamoylguanosin wurde das entsprechende mit N^2-Glycinylcarbamoylguanosin modifizierte **ON2** ausgehend von **ON4** unter nitrosierenden Bedingungen und anschließender Kopplung mit Glycin erhalten (**Abbildung 11**).

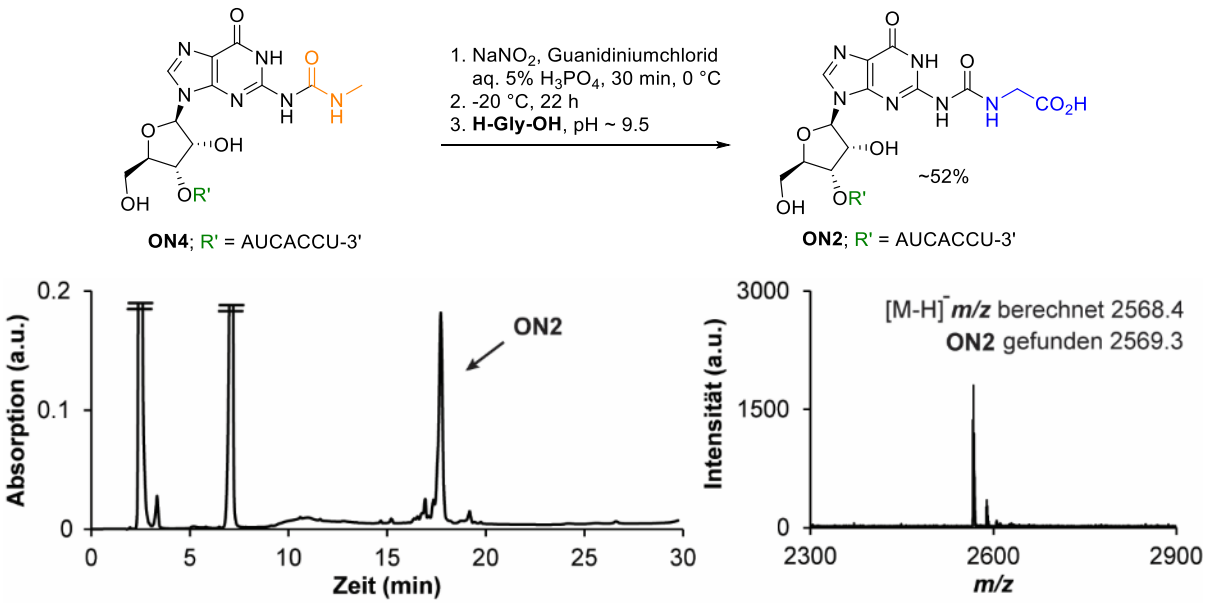
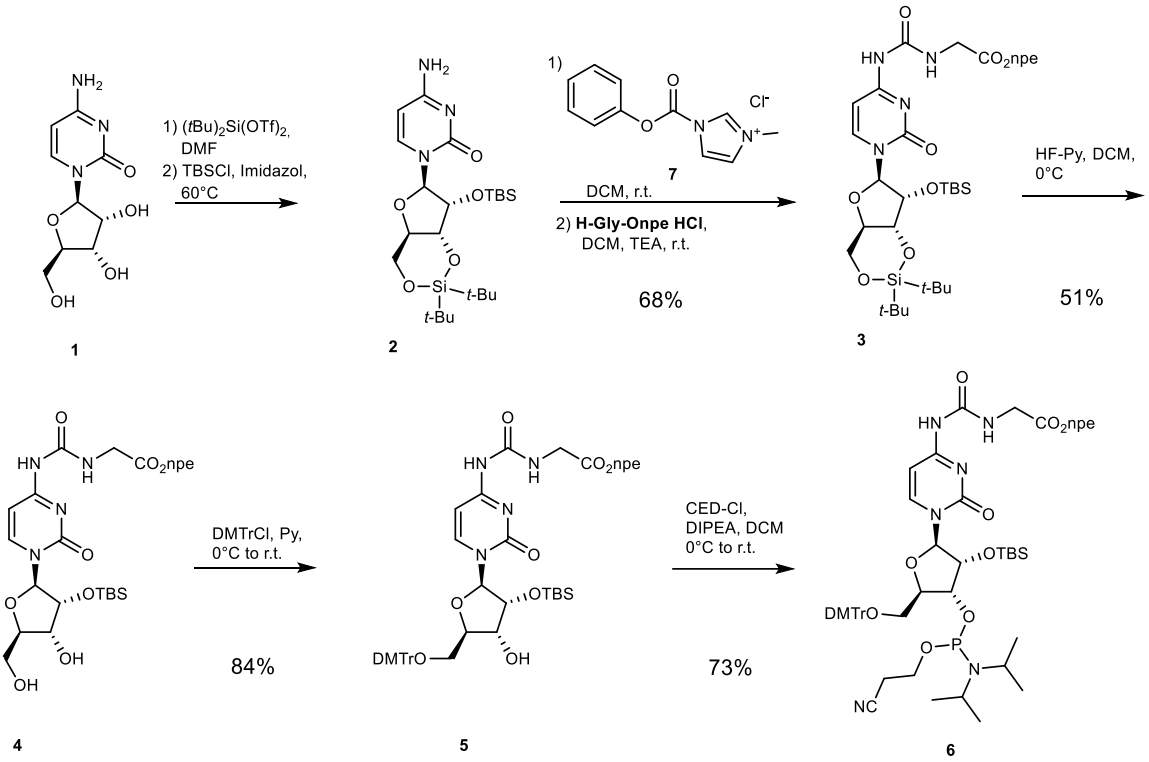


Abbildung 11: Reaktionsschema der Beladungsreaktion des modifizierten Oligonukleotids **ON4** mit Glycin zu **ON2** und das dazugehörige HPLC-Chromatogramm und Maldi-TOF Spektrum.

ON4 wurde mit 500 Äquivalenten. $NaNO_2$ in einer 5%igen wässrigen H_3PO_4-Lösung bei -20°C umgesetzt. Anschließend wurde ein Überschuss an Glycin zugegeben und der pH-Wert auf 9.5 eingestellt. Durch Zugabe von Guanidiniumchlorid (GdmCl) zur Nitrosierungsreaktion konnte die Ausbeute der Reaktion auf ca. 52% gesteigert werden. Dies kann auf die Stabilisierung des Phosphatrückgrads durch die Bildung von Salzbrücken mit dem nitrosierten Gdm-Kation[278][279][280] oder auf die Vermeidung von Desaminierungsprodukten, die zuvor die Aufreinigung erschwerten, zurückzuführen sein. Das mittels HPLC isolierte und über MALDI-

TOF-Massenspektrometrie identifizierte Produkt **ON2** wurde in den folgenden Untersuchungen zur Ausbildung eines Haarnadelmotivs verwendet.

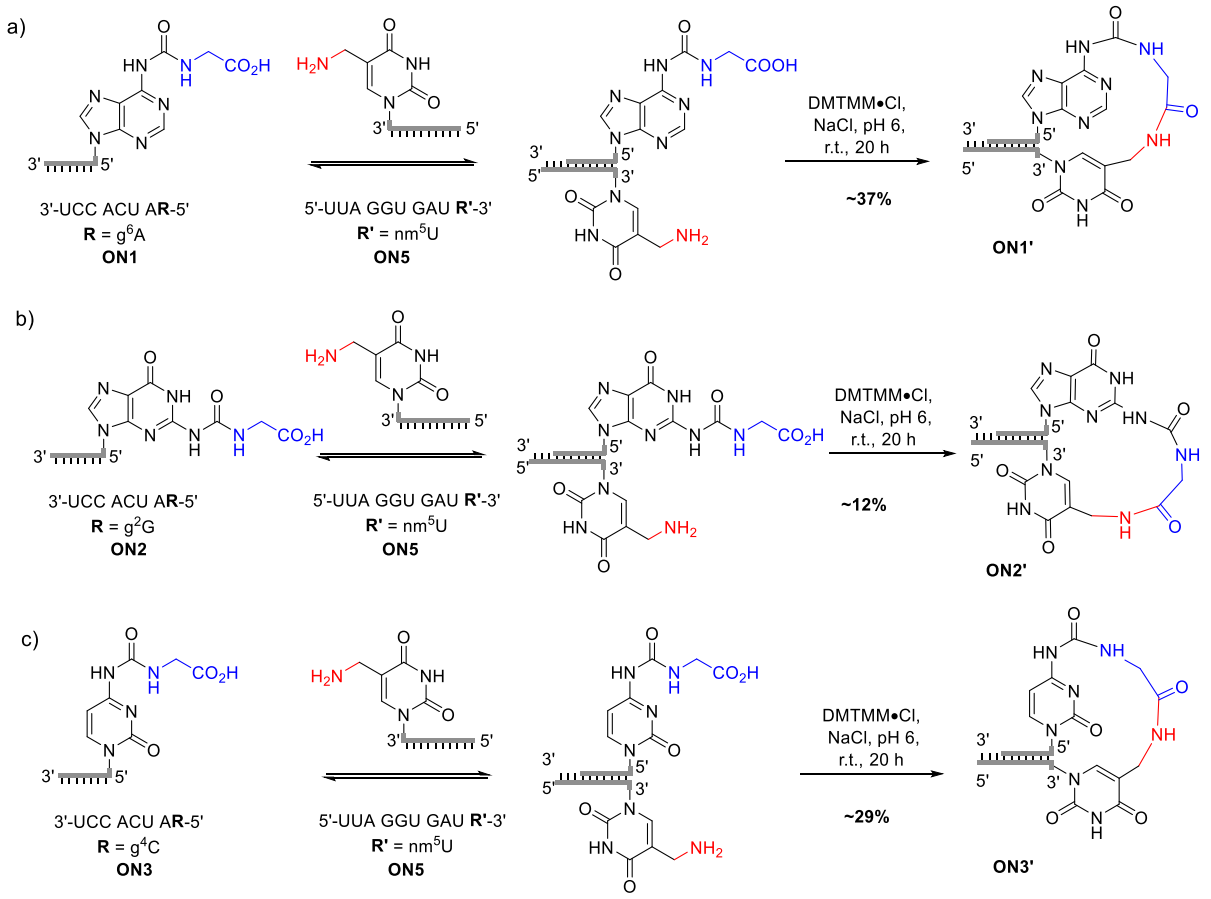
Für die Synthese von **ON3**, das N^4-Carbamoylcytidin als Modifikation enthält, wurde von der präbiotischen Synthese über Nitrosierung abgesehen, da zuvor gezeigt werden konnte, dass die Beladung des Oligonukleotids mit Glycin in geringen Ausbeuten von 5% erfolgt. Stattdessen wurde für das g^4C-Phosphoramidit **6** ein neuer Syntheseweg analog zum g^6A-Phosphoramidit entwickelt (**Schema 9**).



Schema 9. Syntheseroute zum g^4C-Phosphoramidit **6** ausgehend von kommerziell erhältlichem Cytidin **1**.

Ausgehend vom kommerziell erhältlichen Cytidin **1** wurde das silyl-geschützte Cytidin **2** in zwei Schritten durch Reaktion des 1,3-Diols an der 3′- und 5′-Position mit Di-*tert*-butylsilyl-bis(trifluoromethansulfonat) und anschließender Umsetzung mit *tert*-Butyldimethylsilylchlorid erhalten.[281] Als nächstes wurde das geschützte Nukleosid **2** mit 1-*N*-Methyl-3-phenoxycarbonyl-imidazoliumchlorid[282] **7** umgesetzt, gefolgt von der Reaktion mit npe-geschütztem Glycin,[283] wodurch das Aminosäurekonjugat **3** in einer Ausbeute von 68% erhalten wurde. Die selektive Abspaltung der zyklischen Schutzgruppe an der 3′-5′-Position mit HF-Pyridin bei 0 °C und anschließender Schützung der 5′-Hydroxygruppe mit DMTrCl ergab das Nukleosid **5**. Das finale Phosphoramidit **6** konnte so in 4 Schritten in 27% Ausbeute erhalten werden. Dieses wurde darauf mittels RNA-Festphasen-Synthese erfolgreich in **ON3** eingebaut.

Die synthetisierten Aminosäure-modifizierten *N*-Carbamoylnukleotide **ON1-3** wurden auf ihre Eigenschaft untersucht, mit einem komplementären Oligonukleotid Amide zu bilden und somit zwei RNA-Fragmente miteinander zu verknüpfen. Als Reaktionspartner wurde dazu das 11-mere Oligonukleotid **ON5** synthetisiert, das am 3′-Ende ein nm^5U-Nukleotid enthält. In drei separaten Ansätzen wurde jeweils ein Oligonukleotid **ON1-3** mit dem Akzeptorstrang **ON5** hybridisiert und nach Zugabe eines Kopplungsreagenzes (DMTMM) die Bildung der verknüpften RNA-Fragmente **ON1′-ON3′** beobachtet (**Schema 10** und **Kapitel 5.3**).



Schema 10. Hybridisierung und Amidkopplung der Aminosäure-modifizierten *N*-Carbamoylnukleotide **ON1-3** mit dem komplementären Oligonukleotid **ON5**.

Die Oligonukleotide **ON1-3** und **ON5** wurden in äquimolaren Mengen in einer Lösung aus 100 mM 2-(*N*-Morpholino)ethansulfonat (MES) Puffer pH 6 und 100 mM NaCl gelöst. Durch Zugabe von DMTMM wurden die hybridisierten RNA-Doppelstränge in die kovalent verbundenen Oligonukleotide **ON1′-ON3′** umgewandelt. Dabei wurde beobachtet, dass die verschiedenen Modifikationen, g^6A, g^4C und g^2G, einen Einfluss auf die Effizienz der Reaktion nach 20 h haben. Die Kopplungen von **ON1** bzw. **ON3** und **ON5** ergaben vergleichbare Ausbeuten von 29-37%, wohingegen die Kopplung von **ON2** und **ON5** eine geringere Ausbeute von 12% zeigte. Ein Grund dafür könnte die Position der Verknüpfung zwischen der Aminosäure und der Nukleobase über die Harnstoffbrücke sein. Bei g^6A und g^4C besteht eine intramolekulare Wasserstoffbrückenbindung zwischen der NH-Gruppe der Aminosäure und

der N^1-Position des Purins bzw. der N^3-Position des Pyrimidins. Bei g^2G hingegen existieren zwei Konformationen zwischen denen die Harnstoffbrücke rotieren kann. In der geschlossenen Konformation besteht eine Wasserstoffbrückenbindung zwischen dem Carbonyl-Sauerstoff der Harnstoffeinheit und dem N^1-Proton der Nukleobase. In der offenen Konformation hingegen interagiert die NH-Gruppe der Aminosäure mit der N^3-Position der Nukleobase, wodurch interessanterweise eine Watson-Crick-Basenpaarung mit einem komplementären C gebildet werden kann.[284] Jedoch resultieren daraus vermutlich geringere Ausbeuten in der Kopplungsreaktion mit einem hybridisierten nm^5U-Oligonukleotid.

Zusammenfassend konnte gezeigt werden, dass neben der in tRNAs gefundenen Modifikation g^6A, auch die Derivate g^4C und g^2G die Eigenschaft besitzen, RNA-Fragmente kovalent miteinander zu verknüpfen. Auch wenn die *N*-Carbamoylderivate von Cytidin und Guanosin bisher nicht in der Natur gefunden wurden,[285] können sie unter den gleichen präbiotisch plausiblen Bedingungen wie für die Synthese von *N*-Carbamoyladenosin erhalten werden und könnten somit mögliche strukturgebende Merkmale einer frühen RNA-Peptid-Welt gewesen sein. Neben der Verknüpfung zweier RNA-Fragmente könnten die Modifikationen g^4C und g^2G analog zu den Studien mit g^6A (siehe **Kapitel 3.2**) in einer RNA-basierten Peptidsynthese als Aminosäure-Donor fungieren. Hierzu kann in künftigen Studien die Spaltung der Harnstoffbindung und somit der Transfer der Aminosäure auf das entsprechende Akzeptor-Oligonukleotid untersucht werden.

4.2 Der Einfluss modifizierter RNA auf die Ribozymaktivität

4.2.1 Ergebnisse und Diskussion

Wie zuvor dargestellt, können nicht-kanonische Nukleotide dazu beitragen, dass sich RNA-Fragmente über eine Amidbindung kovalent zu größeren Strukturen zusammenfügen können. Das bei der Peptidkopplung entstehende Haarnadel-Motiv führt zum einen zu einer höheren Schmelztemperatur des RNA-Duplex (siehe **Kapitel 3.2**) und zum anderen können chimäre Aminosäure-RNA-Konjugate erhalten werden, wodurch das Repertoire an funktionellen Gruppen erweitert wird. Um den Einfluss dieser chimären Moleküle auf einen möglichen evolutionären Vorteil, der sich als gesteigerte katalytische Aktivität ausprägen kann, untersuchen zu können, bedarf es eines geeigneten Modell-Systems. Zuvor konnte von *Szostak* und Mitarbeitern gezeigt werden, dass sich chimäre Ribozyme aus RNA-Fragmenten zusammensetzen können, die über Aminosäuren in Form von Phosphoramidaten mit dem 5′-Phosphat und als Ester mit der 3′-Hydroxygruppe der Ribose verbunden sind (siehe **Kapitel 1.4**).[252]

In diesem Abschnitt wurde der Einfluss des Haarnadel-Motivs auf die katalytische Aktivität untersucht. Als Modellsystem wurde in analoger Weise zu den Studien von *Szostak* und Mitarbeitern ein I/III-Hammerhead Ribozym ausgewählt, das einen Substrat-Strang in einer Transesterifizierung am Nukleotid C_{17} spaltet.[252] Zunächst wurde die Spaltung des Substrats **ON7** durch ein kanonisches Ribozym **ON6** untersucht, das im Stamm-II eine Haarnadel-Struktur aufweist. Die beiden Oligonukleotide wurden im Verhältnis 1:100 **ON6**/**ON7** in einer wässrigen Lösung, bestehend aus 100 mM NaCl, 5 mM $MgCl_2$ und 50 mM Phosphatpuffer pH 8, bei 25 °C inkubiert. Das Reaktionsgemisch wurde mittels analytischer HPLC analysiert und die Spaltung von **ON7** über eine Kalibriergerade quantifiziert (siehe **Kapitel 5.2**). Nach einer Reaktionszeit von 2 h wurde eine Spaltung von 81% in die beiden Produkte **ON7′** und **ON7″** beobachtet (**Abbildung *12*a**). Wurde die Reaktion in Abwesenheit von Mg^{2+} durchgeführt, konnte keine Spaltung beobachtet werden. Dies bestätigt frühere Studien, in denen festgestellt wurde, dass zweiwertige Kationen, wie z.B. Mg^{2+} oder Mn^{2+} für die Ausprägung der aktiven Konformation des Hammerhead Ribozyms benötigt werden.[274] Neben dem Einfluss der Mg^{2+}-Konzentration konnte auch gezeigt werden, dass die Sekundärstruktur entscheidend für die katalytische Aktivität ist. Sobald das Haarnadel-Motiv des Ribozyms durch zwei nur um 4 Basenpaarungen komplementäre Sequenzen **ON8** und **ON9** ersetzt wurde, konnte nach 2 h Reaktionszeit kein messbarer Umsatz beobachtet werden (**Abbildung *12*b**). Dies ist auf die geringere Schmelztemperatur des Duplex zwischen **ON8** und **ON9** im Vergleich zum Haarnadel-Motiv in **ON6** zurückzuführen.

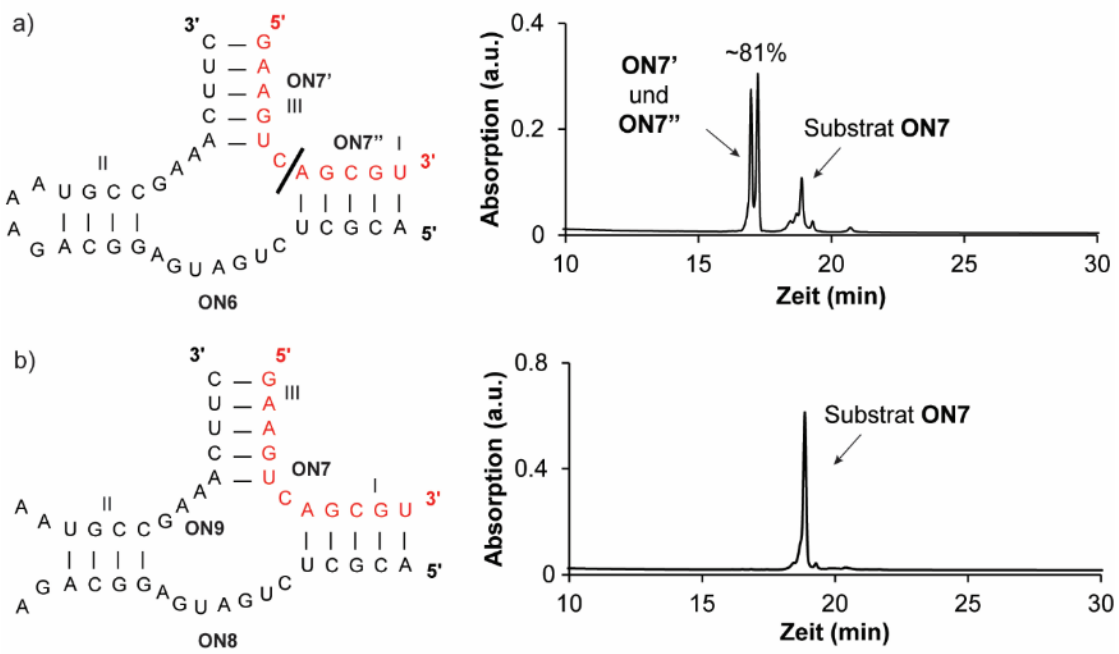


Abbildung 12. a) Spaltungsreaktion des Substrats **ON7** (rot) in die Fragmente **ON7′** und **ON7″** durch das I/III-Hammerhead Ribozym **ON6** (schwarz). b) Kontrollreaktion mit den Ribozym-Fragmenten **ON8** und **ON9**. Reaktionsbedingungen: 50 µM Substrat **ON7**; 0.5 µM Ribozym **ON6** bzw. Ribozym-Fragmente **ON8** und **ON9**; 5 mM $MgCl_2$; 100 mM NaCl; 50 mM Phosphatpuffer pH8; 25 °C; 2 h.

Ausgehend von diesen Beobachtungen wurde ein Hammerhead Ribozym entworfen, in dem die vier Nukleotide G, A, A und A im Stamm-II des kanonischen Ribozyms **ON6** durch die Aminosäure-modifizierten Nukleoside g^6A und nm^5U ersetzt sind, die dann einer Amidkopplung unterzogen werden können. Die entsprechenden Oligonukleotide **ON10** und **ON11** wurden mittels Festphasen-RNA-Synthese hergestellt. Das 17-mere **ON10** enthielt das Nukleosid nm^5U am 3′-Ende und das 13-mere **ON11** wurde mit dem Nukleosid g^6A am 5′-Ende modifiziert. Beide Oligonukleotide wurden in äquimolaren Mengen in einer Lösung aus 100 mM MES-Puffer pH 6 und 100 mM NaCl gemäß den zuvor beschriebenen Reaktionsbedingungen gelöst. Durch Zugabe von DMTMM und einer Inkubation für 20 h bei 25 °C wurde das kovalent verbundenen Oligonukleotid **ON12** in einer Ausbeute von ~3% erhalten. Die niedrige Ausbeute ist auf die geringe Anzahl von Basenpaarungen zwischen den Oligonukleotiden zurückzuführen. Durch Anpassung der Reaktionsbedingungen, um die Hybridisierung von **ON10** und **ON11** zu fördern, konnte die Ausbeute auf 20% gesteigert werden. Die NaCl-Konzentration wurde auf 1000 mM erhöht und die Reaktionstemperatur auf 15 °C gesenkt. Das modifizierte Ribozym wurde mittels HPLC isoliert und über MALDI-TOF-Massenspektrometrie charakterisiert (**Abbildung *13***).

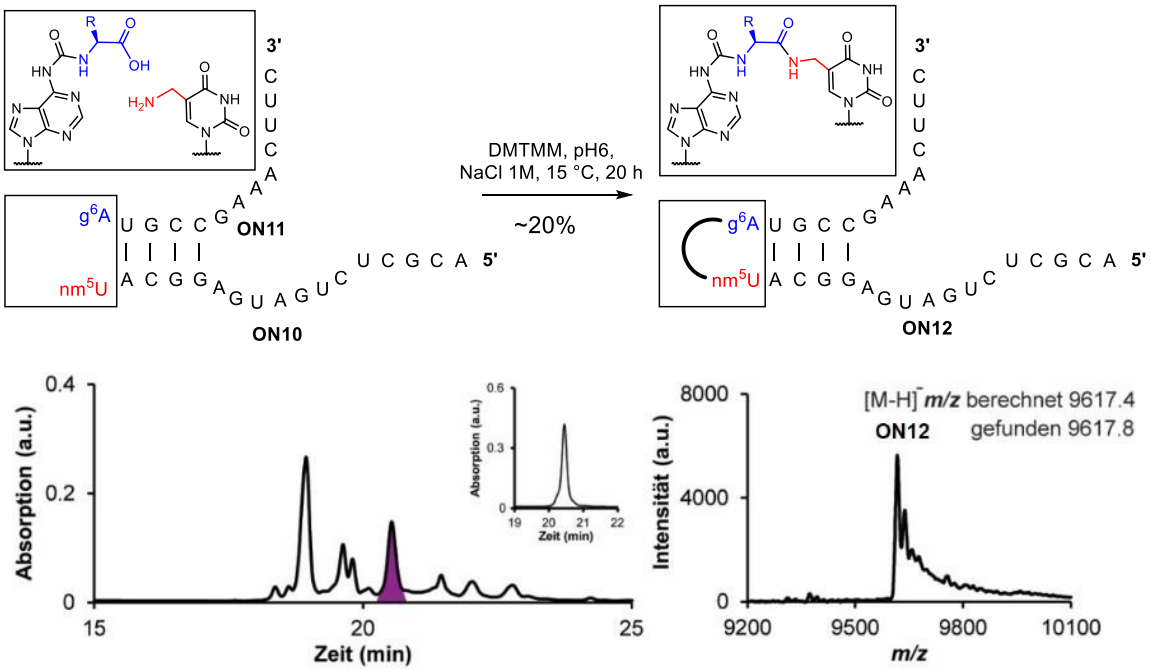


Abbildung 13: Kopplungsreaktion zwischen den nm^5U- und g^6A-modifizierten Fragmenten **ON10** und **ON11** um das vollständige Ribozym **ON12** zu erzeugen. HPLC-Chromatogramm und MALDI-TOF-Spektrum des isolierten Ribozyms **ON12**. Das Ribozym **ON12** ist in violett hervorgehoben und das Chromatogramm der isolierten Fraktion als Einschub dargestellt.

Anschließend wurde in einer Kontrollreaktion, analog zu den Experimenten mit den unmodifizierten Oligonukleotiden die Spaltung des Substrats **ON7** untersucht. Bei den nicht gekoppelten Strängen **ON10** und **ON11** wurde keine messbare Spaltung von **ON7** beobachtet,

wohingegen das gekoppelte Ribozym **ON12** das Substrat nach 2 h zu 51% zu den Produkten **ON7'** und **ON7''** umsetzte (**Abbildung *14***). Die geringere katalytische Aktivität des chimären Ribozyms **ON12** im Vergleich zum unmodifizierten Ribozym **ON6** ist vermutlich auf die unterschiedliche Anzahl von Nukleotiden im Haarnadel-Motiv zurückzuführen. Bei **ON6** sind es die vier Nukleotide G, A, A und A, während es bei **ON12** nur die zwei Nukleotide nm^{5}U und g^{6}A sind. Um eine bessere Vergleichbarkeit zu erreichen, ist es notwendig weitere Studien durchzuführen, ob eine Amidkopplung auch möglich ist, wenn sich neben nm^{5}U und g^{6}A zwei weitere Nukleotide befinden, wie z.B. A und G, die keine Watson-Crick-Basenpaarung bilden. Auf diese Weise könnte eine größere strukturelle Ähnlichkeit mit dem aus vier Nukleotiden bestehenden Haarnadel-Motiv erreicht werden.

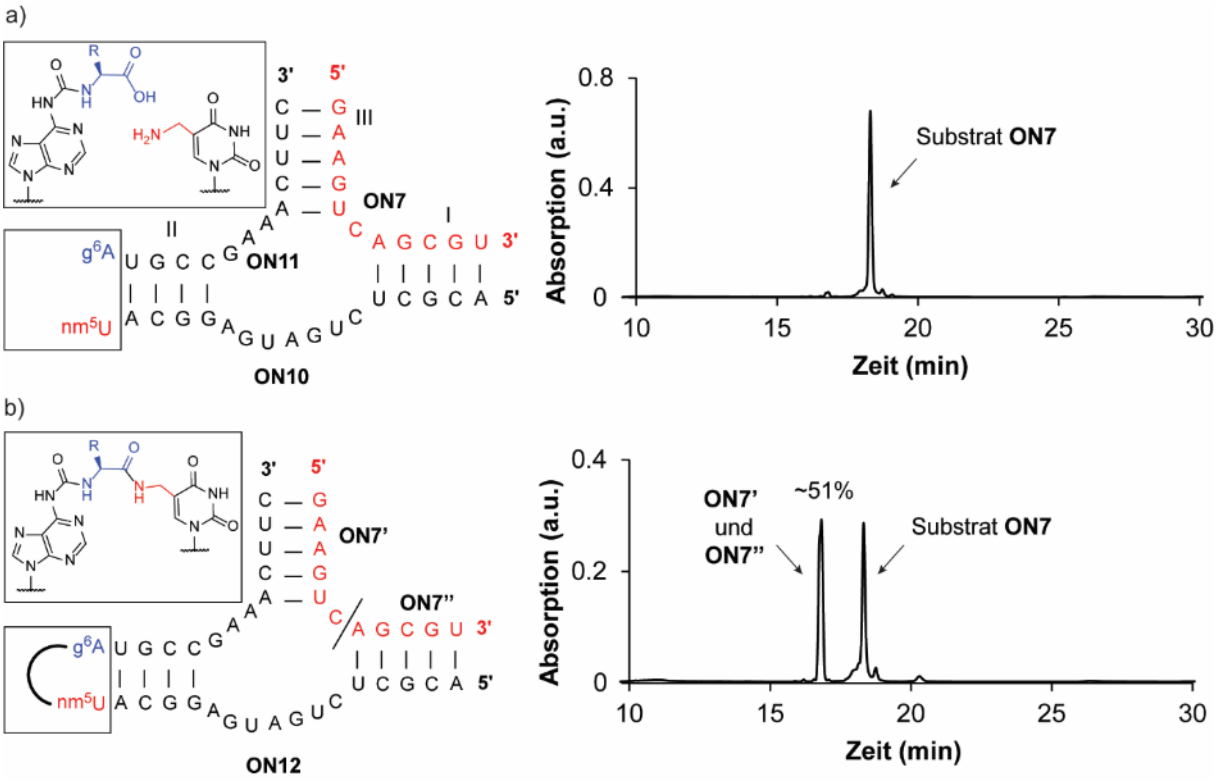


Abbildung 14. a) Kontrollreaktion mit den nm^{5}U- und g^{6}A-modifizierten Ribozym-Fragmenten **ON10** und **ON11**. b) Spaltungsreaktion des Substrats **ON7** (rot) in die Fragmente **ON7'** und **ON7''** durch das Aminosäure-modifizierte I/III-Hammerhead Ribozym **ON12** (schwarz). Reaktionsbedingungen: 50 µM Substrat **ON7**; 0.5 µM Ribozym **ON12** bzw. Ribozym-Fragmente **ON10** und **ON11**; 5 mM $MgCl_2$; 100 mM NaCl; 50 mM Phosphatpuffer pH8; 25 °C; 2 h.

Zusammenfassend konnte in diesen ersten Untersuchungen gezeigt werden, dass nicht-kanonische Nukleotide, die zwei RNA-Stränge über eine Amidbindung zwischen den Nukleobasen verknüpfen, die katalytische Aktivität eines Ribozyms im Vergleich zu den nicht verknüpften Einzelsträngen erhöhen können. Die Verknüpfung über ein Amid führt zu einem Haarnadel-Motiv, das eine analoge Sekundärstruktur zu einem natürlich vorkommenden Haarnadel-Motiv ausbilden kann. Neben der Bestimmung der Schmelztemperaturen der Ribozyme ist für zukünftige Untersuchungen auch die Bestimmung der Reaktionsraten oder

die Strukturaufklärung mittels Röntgendiffraktometrie von Interesse, um einen genaueren Vergleich zwischen dem unmodifizierten und dem mit Aminosäuren modifizierten Ribozym zu ermöglichen. Darüber hinaus kann neben Glycin auch der Einfluss unpolarer, aromatischer oder ionisierbarer Aminosäuren von Interesse sein. Insbesondere dann, wenn sich die Aminosäuren in der Nähe des aktiven Zentrums des Ribozyms befinden. In den hier beschriebenen Untersuchungen befand sich die Aminosäure jedoch außerhalb des aktiven Zentrums und diente dort ausschließlich als kovalentes Bindeglied. Um eine direkte Beteiligung von Aminosäure-modifizierten Nukleotiden an der Säure-Base-Katalyse der Spaltung der Phosphordiesterbindung zu untersuchen, ist es notwendig, die modifizierten Nukleotide in der Nähe des aktiven Zentrums einzubauen. Dort könnten sie die Funktion von monovalenten Ionen wie z.B. Na^+ ersetzen[270] oder direkt als Säure und Base wirken. Das Hammerhead Ribozym als Modellsystem bietet auch die Möglichkeit tertiäre Wechselwirkungen zwischen Stamm-I und Stamm-II zu untersuchen.[268][276][277] In diesem Fall tragen diese Wechselwirkungen zur Konformationsänderung vom präkatalytischen Zustand zum katalytisch aktiven Zustand bei. Die kovalente Verknüpfung des Stamm I und Stamm II über nicht-kanonische Nukleotide könnte zur Stabilisierung des Ribozyms in der katalytisch aktiven Konformation beitragen. Neben dem Hammerhead Ribozym können nicht-kanonische Nukleoside auch in anderen Ribozymen, z.B. aminoacylierenden Ribozymen oder Ligase Ribozymen, in Betrachtet gezogen werden.[252] Zusammenfassend zeigen die hier beschriebenen Ergebnisse, dass Nukleobasen-modifizierte Nukleotide als strukturgebende Motive ein Motor der Evolution gewesen sein könnten, der im präbiotischen Kontext einer RNA-Peptid-Welt berücksichtigt werden muss.

5. Experimenteller Teil

Generelle Informationen zur Synthese der Phosphoramidite

Die Reagenzien wurden von kommerziellen Anbietern bezogen und ohne weitere Aufreinigung verwendet, sofern nicht anders angegeben. Wasserfreie Lösungsmittel, die unter inerter Atmosphäre gelagert wurden, wurden ebenfalls gekauft. Alle Reaktionen, an denen luft- oder feuchtigkeitsempfindliche Reagenzien/Zwischenprodukte beteiligt waren, wurden unter inerter Atmosphäre mit getrockneten Glasgeräten durchgeführt. Routinemäßige ^{1}H-NMR-, ^{13}C{^{1}H} NMR- und ^{31}P{^{1}H} NMR-Spektren wurden mit einem Bruker Ascend 400 Spektrometer (400 MHz für ^{1}H NMR, 100 MHz für ^{13}C NMR und 162 MHz für ^{31}P NMR) oder einem Bruker Ascend 500 Spektrometer (500 MHz für ^{1}H NMR, 125 MHz für ^{13}C NMR und 202 MHz für ^{31}P NMR) aufgenommen. Die verwendeten deuterierten Lösungsmittel sind in der Charakterisierung angegeben und die chemischen Verschiebungen (δ) sind in ppm angegeben. Die jeweiligen Lösungsmittelpeaks wurden als Referenz verwendet. Alle NMR-*J*-Werte sind in Hz angegeben. COSY-, HSQC- und HMBC-Experimente wurden durchgeführt, um die Zuordnung der ^{1}H- und ^{13}C-Signale zu erleichtern. Die NMR-Spektren wurden mit der Software MestReNova, Version 10.0, ausgewertet. Hochauflösende Massenspektren (HRMS) wurden mit einem Thermo Finnigan LTQ-FT mit ESI als Ionisierungsmodus gemessen. IR-Spektren wurden mit einem Perkin-Elmer Spectrum BX II FT-IR-Gerät oder einem Shimadzu IRSpirit FT-IR-Gerät aufgenommen. Beide sind mit einem ATR-Zubehör ausgestattet. Für die Säulenchromatographie wurde Kieselgel mit einer Teilchengröße von 40-63 µm verwendet. Der Reaktionsfortschritt wurde durch Dünnschichtchromatographie (TLC) auf Kieselgel 60 F254 überwacht und mit *p*-Anisaldehyd angefärbt.

Generelle Informationen zur Synthese der Oligonukleotide

Die Phosphoramidite der kanonischen Ribonukleoside (Bz-A-CE, Dmf-G-CE, Ac-C-CE und U-CE) wurden von LinkTech und Sigma-Aldrich bezogen. Oligonukleotide (ON) wurden im 1 µmol-Maßstab unter Verwendung von RNA SynBase™ CPG 1000/110 und High Load Glen UnySupport™ als Festphase mit einem automatischen RNA-Synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) durch Standard-Phosphoramidit-Chemie synthetisiert. ONs wurden im DMT-OFF-Modus unter Verwendung von Dichloressigsäure (DCA) als Entschützungsreagenz in Dichloromethan, BTT oder Activator 42® als Aktivator in MeCN, Ac_2O als Capping-Reagenz in Pyridin/THF und I_2 als Oxidationsmittel in Pyridin/H_2O synthetisiert.

Entschützung von *p*-Nitrophenylethyl- (npe) und (Trimethylsilyl)ethoxycarbonyloxy- (teoc)-Gruppen

Für die Entschützung der *p*-Nitrophenylethyl (npe)-Gruppe in ONs, die Aminosäure-modifizierte Carbamoyl-Adenosin-Nukleoside enthalten, wurden das Festphasenmaterial in einer 9:1 THF/DBU-Lösungsmischung (1 mL) suspendiert und 2 h bei r.t. inkubiert. Danach wurde der Überstand entfernt und die Festphase mit THF (3×1 mL) gewaschen.

Für die Entschützung der 2-(Trimethylsilyl)ethoxycarbonyl (teoc)-Gruppe in ONs, die 5-Methylaminomethyl-Uridin-Nukleoside enthalten, wurde die Festphase in einer gesättigten Lösung von $ZnBr_2$ in 1:1 $MeNO_2$/IPA (1 mL) suspendiert und über Nacht bei r.t. inkubiert. Danach wurde der Überstand entfernt, und die Festphase wurde mit 0.1 M EDTA in Wasser (1 mL) und Wasser (1 mL) gewaschen.

Abspaltung von der Festphase, Abspaltung der TBS-Gruppen und Ausfällung der synthetisierten ON

Die Festphase wurde in einer wässrigen 1:1 Mischung (0.6 mL) aus 30% NH_4OH und 40% $MeNH_2$ suspendiert. Die Suspension wurde bei 65 ºC erhitzt (10 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 vereinigten wässrigen Lösungen wurden unter vermindertem Druck mit einem SpeedVac-Konzentrator konzentriert. Anschließend wurde das Rohprodukt in DMSO (100 µl) gelöst und Triethylamintrihydrofluorid (125 µl) hinzugefügt. Die Lösung wurde 1.5 h lang bei 65 ºC erhitzt. Schließlich wurde das ON durch Zugabe von 3 M NaOAc in Wasser (25 µL) und *n*-Butanol (1 mL) ausgefällt. Die Mischung wurde 2 h lang bei -80 ºC aufbewahrt und 1 h lang bei 4 ºC zentrifugiert. Der Überstand wurde entfernt und der weiße Niederschlag gefriergetrocknet.

Reinigung des synthetisierten ON mittels HPLC und Entsalzung

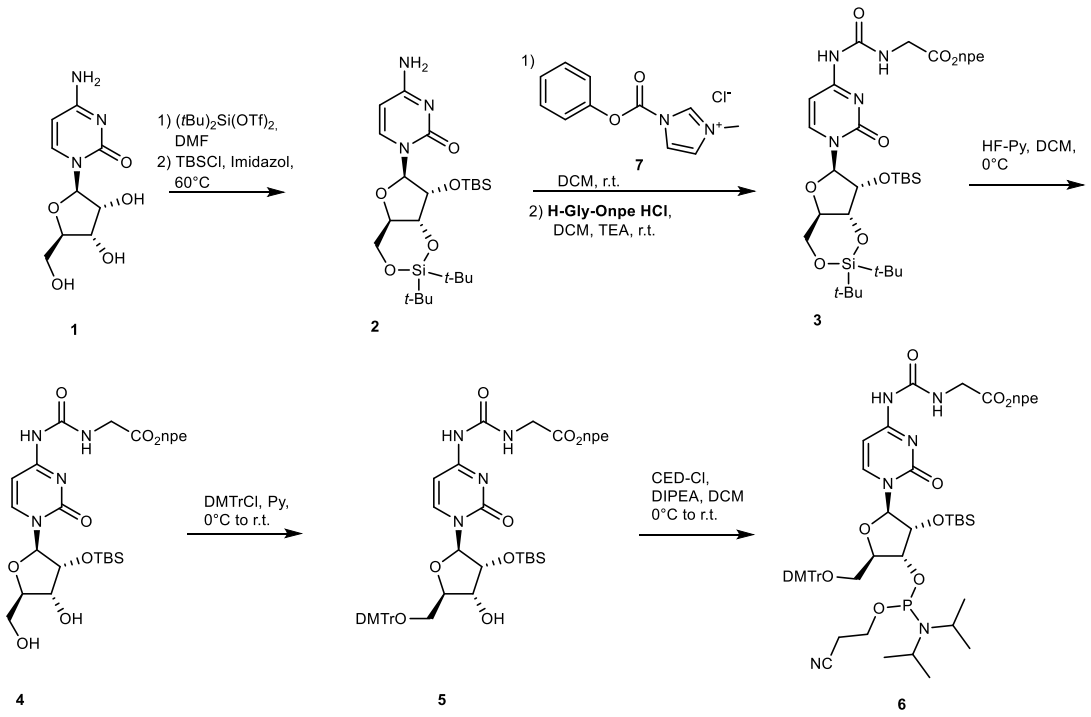
Das Rohprodukt wurde mittels semipräparativer HPLC (1260 Infinity II Manual Preparative LC System von Agilent, ausgestattet mit einem G7114A Detektor) unter Verwendung einer Reverse-Phase (RP) VP 250/10 Nucleodur 100-5 C18ec Säule von Macherey-Nagel gereinigt (Puffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O und Puffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-25% von B in 45 min; Flussrate = 5 mL/min). Das gereinigte ON wurde mittels RP-HPLC (1260 Infinity II LC System von Agilent, ausgestattet mit einem G7165A Detektor) unter Verwendung einer EC 250/4 Nucleodur 100-3 C18ec Säule von Macherey-Nagel analysiert (Gradient: 0-30% von B in 45 min; Durchflussrate = 1 mL/min). Schließlich wurde das aufgereinigte ON mit einer C18 RP-Kartusche von Waters entsalzt.

Bestimmung der Konzentration und der Masse der synthetisierten ONs

Die Absorption des synthetisierten ON 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 ONs, die nicht-kanonische Basen enthalten, wurde angenommen, dass die Extinktionskoeffizienten identisch zu den kanonischen Basen sind.

Das synthetisierte ON (2-3 µL) wurde über einen 0.025 µm VSWP-Filter (Millipore) entsalzt, in einer 3-Hydroxypicolinsäure-Matrix (HPA, 1 µL) kristallisiert und mittels MALDI-TOF-Massenspektrometrie (negativer Modus) analysiert.

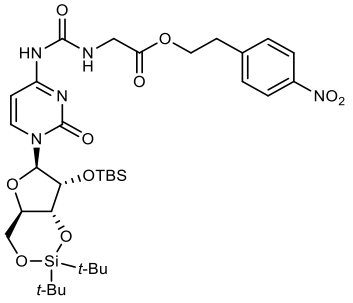
5.1 Synthese des g^{4}C Phosphoramidits



Schema 11: Syntheseweg zum N4-Glycinyl(npe)carbamoylcytidine-Phosphoramidit **6**.

Die N^4-Glycinyl(npe)carbamoylcytidin-Derivate **3**-**6** wurden in analoger Weise zu den N^6-Glycinyl(npe)carbamoyladenosin-Bausteinen in Kapitel 3.1 hergestellt.

Silyl-geschütztes *N*^{4}-Glycinyl(npe)carbamoylcytidine 3

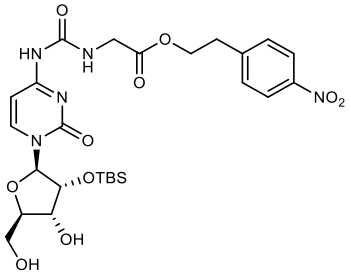


3

Das silyl-geschützte Cytidinderivat **2** (2.00 g, 4.02 mmol, 1.00 Äq.) wurde in trockenem DCM (30 mL) unter N_2-Atmosphäre gelöst. 1-*N*-Methyl-3-phenoxycarbonyl-imidazoliumchlorid **7** (1.52 g, 8.04 mmol, 2.00 Äq.) wurde dem Reaktionsgemisch zugesetzt und die resultierende Suspension wurde bei r.t. über Nacht gerührt. Anschließend wurde H-Gly-Onpe-Hydrochlorid (2.09 g, 8.04 mmol, 2.00 Äq.) zusammen mit TEA (1.1 mL, 8.0 mmol, 2.0 Äq.) als Suspension in DCM (10 mL) zugegeben. Das Reaktionsgemisch wurde über Nacht bei Raumtemperatur gerührt. Die Reaktion wurde durch die Zugabe von gesättigter wässriger $NaHCO_3$-Lösung beendet und die wässrige Phase mit DCM (3 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden über $MgSO_4$ getrocknet. Das Lösemittel wurde *in vacuo* entfernt und das Rohprodukt mittels Säulenchromatographie (Silikagel, EtOAc/Hex, 0% → 50%) aufgereinigt und das Produkt **3** (2.05 g, 2.74 mmol, 68%) in Form eines weißen Schaums erhalten.

R_f = 0.54 (1:1 *i*-Hexane/EtOAc) IR: $\tilde{\nu}$ = 3213 (w), 2932 (w), 2889 (w), 2857 (w), 1731 (m), 1719 (m), 1638 (m), 1599 (w),1565 (m), 1516 (s), 1473 (m), 1427 (w), 1386 (w), 1371 (w), 1344 (s), 1289 (w), 1246 (m), 1229 (m), 1204 (m), 1185 (m), 1162 (m), 1139 (m), 1124 (m), 1107 (m), 1081 (m), 1052 (s), 1025 (m), 991 (m), 956 (w), 937 (w), 898 (w), 854 (m), 840 (s), 825 (s), 783 (s), 750 (m), 720 (m), 686 (m) cm^{-1}; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.85 (br s, 1H), 9.76 (br s, 1H), 8.08 (d, *J* = 8.7 Hz, 2H), 7.63 (d, *J* = 7.7 Hz, 1H), 7.55 (d, *J* = 7.7 Hz, 1H), 7.38 (d, *J* = 8.7 Hz, 2H), 5.67 (s, 1H), 4.56 (dd, *J* = 9.3, 5.1 Hz, 1H), 4.46 – 4.32 (m, 3H), 4.28 (td, *J* = 10.2, 5.1 Hz, 1H), 4.03 (dd, *J* = 10.2, 9.3 Hz, 1H), 3.92 (t, *J* = 6.4 Hz, 2H), 3.82 (dd, *J* = 9.3, 5.1 Hz, 1H), 3.05 (t, *J* = 6.4 Hz, 2H), 1.03 (s, 9H), 1.02 (s, 9H), 0.92 (s, 9H), 0.17 (s, 3H), 0.14 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 170.0, 164.9, 156.3, 154.6, 146.8, 145.7, 141.7, 130.0, 123.7, 97.4, 94.0, 75.7, 75.4, 74.9, 67.8, 64.5, 42.0, 34.9, 27.6, 27.4, 27.1, 26.0, 22.9, 20.4, 19.9, 18.2, -4.2, -4.7; HRMS (ESI): berechnet für $C_{34}H_{54}N_5O_{10}Si_2$: m/z = 748.3404 $[M+H]^+$; gefunden: m/z = 748.3408.

3′-5′-entschütztes *N*^{4}-Glycinyl(npe)carbamoylcytidine 4

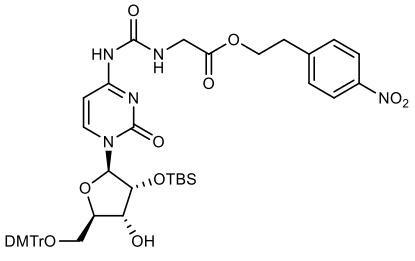


4

Das modifizierte Cytidin **3** (2.05 g, 2.74 mmol, 1.00 Äq.) wurde in Pyridin (3 mL) und DCM (15 mL) in einem Reaktionsgefäß aus Kunststoff gelöst und die Lösung in einem Eisbad abgekühlt. Es wurde HF-Pyridin (0.36 mL, 14 mmol, 5.0 Äq.) tropfenweise zugegeben und das Gemisch für 2 h bei 0 °C gerührt. Die Reaktion wurde mit gesättigter wässriger $NaHCO_3$-Lösung gequencht und mit DCM (3 x 25 mL) extrahiert. Die organischen Phasen wurden über Na_2SO_4 getrocknet und das Lösemittel *in vacuo* entfernt. Das Rohprodukt wurde mittels Säulenchromatographie (Silikagel, MeOH/DCM, 0%→10%) aufgereinigt und das Produkt **4** (843 mg, 1.39 mmol, 51%) als farbloser Schaum erhalten.

R_f = 0.32 (5:100 MeOH/DCM); ^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.74 (br s, 1H), 9.64 (br s, 1H), 8.11 - 8.05 (m, 3H), 7.57 (d, *J* = 7.6 Hz, 1H), 7.40 - 7.35 (m, 2H), 5.55 (d, *J* = 3.4 Hz, 1H), 4.61 (dd, *J* = 5.6, 3.4 Hz, 1H), 4.38 (t, *J* = 6.4 Hz, 2H), 4.23 (dd, *J* = 6.5, 5.6 Hz, 1H), 4.13 - 4.08 (m, 1H), 4.00 (s, 1H), 3.97 - 3.91 (m, 3H), 3.80 (dd, *J* = 12.3, 5.1 Hz, 1H), 3.38 (br s, 1H), 3.05 (t, *J* = 6.4 Hz, 2H), 2.58 (d, *J* = 6.5 Hz, 1H), 0.89 (s, 9H), 0.09 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$) δ: 170.1, 165.1, 156.8, 154.5, 146.8, 145.8, 130.0, 123.8, 97.7, 94.5, 85.4, 74.6, 69.8, 64.5, 61.3, 41.9, 34.9, 25.8, 18.1, -4.5, -5.2; HRMS (ESI): berechnet für $C_{26}H_{38}N_5O_{10}Si$: m/z = 608.2382 $[M+H]^+$; gefunden: m/z = 608.2383.

5′-DMTr-geschütztes *N*^{4}-Glycinyl(npe)carbamoylcytidine 5



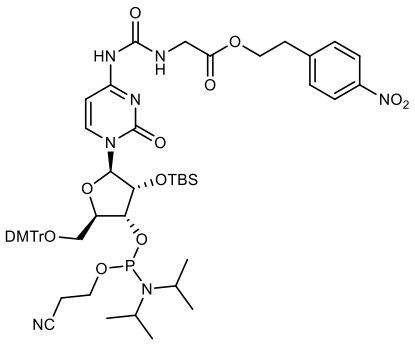
5

Das 3′-5′-entschützte Cytidin-Derivat **4** (843 mg, 1.39 mmol, 1.00 Äq.) wurde in Pyridin (5 mL) gelöst und DMTrCl (705 mg, 2.08 mmol, 1.50 Äq.) zugegeben. Die Lösung wurde über Nacht bei Raumtemperatur gerührt und anschließend wurden das Lösungsmittel unter reduziertem Druck entfernt. Das Rohprodukt wurde mittels Säulenchromatographie (Silikagel, EtOAc/Hex,

0% → 50%) unter Zusatz von 0.1% Pyridin zum Eluenten aufgereinigt, wobei das DMT-geschützte Produkt **5** (1.06 g, 1.16 mmol, 84%) als farbloser Feststoff isoliert wurde.

R_f = 0.51 (1:1 *i*-Hexane/EtOAc); IR: $\tilde{\nu}$: 2926 (w), 2853 (w), 1712 (m), 1642 (s), 1607 (m), 1574 (m), 1504 (s), 1463 (m), 1444 (w), 1385 (w), 1345 (m), 1249 (s), 1174 (m), 1116 (s), 1056 (s), 1033 (m), 1004 (m), 964 (w), 914 (w), 828 (s), 812 (s), 788 (s), 747 (m), 718 (w), 702 (m), 676 (w), 667 (w), 661 (w) cm^{-1}; ^1H NMR (400 MHz, DMSO-d_6) δ: 10.16 (br s, 1H), 8.19 (d, J = 7.4 Hz, 1H), 8.16 - 8.09 (m, 2H), 7.59 - 7.52 (m, 2H), 7.42 - 7.37 (m, 2H), 7.36 - 7.31 (m, 2H), 7.29 - 7.23 (m, 5H), 6.94 - 6.86 (m, 4H), 6.07 (br s, 1H), 5.70 (d, J = 1.6 Hz, 1H), 5.13 (d, J = 5.9 Hz, 1H), 4.34 (t, J = 6.3 Hz, 2H), 4.19 - 4.12 (m, 2H), 4.05 (dt, J = 7.6, 3.1 Hz, 1H), 4.00 (d, J = 5.9 Hz, 2H), 3.74 (s, 6H), 3.35 – 3.30 (m, 2H) 3.06 (t, J = 6.3 Hz, 2H), 0.88 (s, 9H), 0.12 (s, 3H), 0.08 (s, 3H); ^{13}C NMR (100 MHz, DMSO-d_6) δ: 169.8, 162.2, 158.2, 154.0, 144.5, 130.3, 129.8, 128.0, 127.8, 126.9, 123.4, 113.3, 94.5, 90.9, 81.5, 76.3, 68.2, 64.3, 61.8, 55.1, 41.4, 34.0, 25.8, 18.0, -4.7, -4.9; HRMS (ESI): berechnet für $C_{47}H_{56}N_5O_{12}Si$: m/z = 910.3689 $[M+H]^+$; gefunden: m/z = 910.3683.

N^4-Glycinyl(npe)carbamoylcytidine-Phosphoramidit 6



6

Die 5′-DMT-geschützte Verbindung **5** (500 mg, 549 µmol, 1.00 Äq.) wurde unter einer Argon-Atmosphäre in trockenem DCM (3 mL) gelöst und auf 0 °C abgekühlt. Nach der Zugabe von DIPEA (383 µL, 2.20 mmol, 4.00 Äq.) wurde 2-Cyanoethyl-*N*,*N*-diisopropylchlorphosphoramidit (306 µL, 1.37 mmol, 2.50 Äq.) tropfenweise hinzugefügt. Das Reaktionsgemisch wurde für 2 h bei Raumtemperatur gerührt. Die Reaktion wurde durch die Zugabe von gesättigter wässriger $NaHCO_3$-Lösung beendet und mit DCM (3 x 25 mL) extrahiert. Die organischen Phasen wurden über Na_2SO_4 getrocknet, filtriert und das Lösemittel unter reduziertem Druck entfernt. Der Rückstand wurde durch Säulenchromatographie (Silikagel, EtOAc/Hex, 0%→50%) unter Zusatz von 0.1 % Pyridin zum Eluent aufgereinigt. Das Phosphoramidit **6** (446 mg, 0.402 mmol, 73%) wurde als Mischung beider Diastereoisomere in Form eines weißen Schaums erhalten.

R_f = 0.50 (1:1 *i*-Hexane/EtOAc); **^{31}P {^1H} NMR (162 MHz, Acetone-d_6) δ**: 150.5, 148.1; **HRMS (ESI)**: berechent für $C_{56}H_{71}N_7O_{13}PSi$: m/z = 1108.4622 $[M-H]^-$; gefunden: m/z = 1108.4647.

5.2 Charakterisierung der synthetisierten Oligonukleotide

Charakterisierung der modifizierten Oligonukleotide ON1-5

ON1: 5′-**R**AU CAC CU-3′; **R** = g^6A

ON2: 5′-**R**AU CAC CU-3′; **R** = g^2G

ON3: 5′-**R**AU CAC CU-3′; **R** = g^4C

ON4: 5′-**R**AU CAC CU-3′; **R** = $_m urea^2G$

ON5: 5′-UUA GGU GAU **R**-3′; **R** = nm^5U

Tabelle 1. Charakterisierung der synthetisierten Oligonukleotide **ON1**-**5** für Studien der Ausbildung des Haarnadelmotivs mit Aminosäure-modifizierten Oligonukleotiden.

Strang	Nukleosid	Aminosäure	t_R (min)	*m*/*z* berechnet. für $[M\text{-}H]^-$	gefunden
ON1	g^6A	Gly	18.3	2552.4	2552.3
ON2	g^2G	Gly	17.7	2568.4	2569.3
ON3	g^4C	Gly	16.8	2528.4	2528.5
ON4	$_m urea^2G$	-	16.8	2524.4	2524.3
ON5	nm^5U	-	17.5	3189.4	3190.5

Charakterisierung der Oligonukleotide ON6-12 für den Einfluss von Modifikationen der Nukleobase auf die katalytische Aktivität des Hammerhead Ribozyms

ON6: 5′-ACG CUC UGA UGA GGC AGA AAU GCC GAA ACU UC-3′

ON7: 5′-GAA GUC AGC GU-3′

ON8: 5′-ACG CUC UGA UGA GCC AGA-3′

ON9: 5′-AAU GCC GAA ACU UC-3′

ON10: 5′-ACG CUC UGA UGA GGC A**R**-3′; **R** = nm^5U

ON11: 5′-**R**UG CCG AAA CUU C-3′; **R** = g^6A

ON12: 5′-ACG CUC UGA UGA GGC A**RR′** UGC CGA AAC UUC-3′; **R** = nm^5U; **R′** = g^6A

Tabelle 2. Charakterisierung der synthetisierten Oligonukleotide für die Studien eines modifizierten Hammerhead Ribozyms.

Strang	Eigenschaft	t_R (min)	*m/z* berechnet. für [*M*-H]$^-$	gefunden
ON6	I/III-Ribozym	19.7	10265.4	10269.6
ON7	I/III-Substrat	18.0	3527.5	3527.1
ON8	I/II-Stamm-Fragment	19.6	5791,8	5793.9
ON9	II/III-Stamm-Fragment	19.1	4411.6	4412.0
ON10	nm^{5}U-Stamm-Fragment	18.9	5451.8	5453.6
ON11	g^{6}A-Stamm-Fragment	19.6	4182.6	4183.4
ON12	Modifiziertes Ribozym	20.5	9617.4	9618.8

Kalibriergerade des I/II-Substrats ON7

Das kanonische Oligonukleotid **ON7** wurde für die Erstellung einer HPLC-Kalibrierungskurve verwendet. Stammlösungen von **ON7** wurden in Wasser (100 µM) hergestellt. Separate Standardlösungen mit 1.2, 1.0, 0.8, 0.6, 0.4, 0.2 und 0.1 nmol **ON7** wurden in einem Endvolumen von 20 µL vorbereitet. Die Standardlösungen wurden in eine analytische HPLC injiziert, die mit einer C18-Säule und unter Verwendung der Puffer A und B injiziert (Puffer A: 0.1M AcOH/Et_3N pH 7 in H_2O und Puffer B: 0.1M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN Gradient: 0-40% von B in 45 min; Flussrate = 1 mL/min). Die Absorption wurde bei 260 nm gemessen, und die Flächen der chromatographischen Peaks wurden durch Integration der Chromatogramme bestimmt. Die Aufzeichnung der chromatographischen Fläche (a.u.) gegen die Menge (nmol) folgte einer linearen Beziehung.

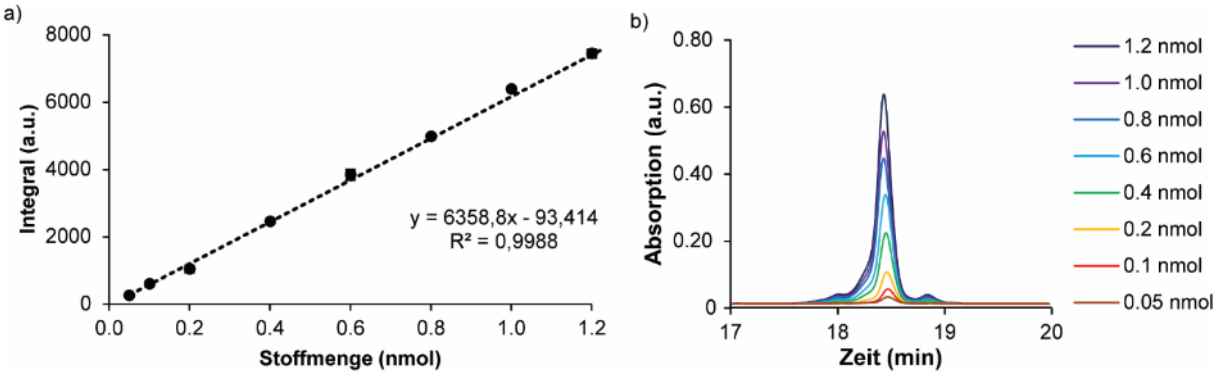


Abbildung 15: a) Kalibriergerade des Substrats **ON7** und b) die übereinandergelegten UV-Spuren bei verschiedenen Konzentrationen.

5.3 Präbiotische Reaktionsbedingungen zur Modifikation der RNA

Durchführung der Beladungsreaktion von N^2-Methylcarbamoylguanosin mit Glycin

Schritt 1: Das Oligonukleotid **ON4** (20 nmol, 1.0 Äq.) und $NaNO_2$ (10 µmol, 500 Äq.) wurden in 5%iger wässriger H_3PO_4-Lösung (400 µL) mit 100 mM Guanidiniumchlorid (GdmCl) gelöst. Die Reaktion wurde 30 Minuten lang bei 0°C gerührt. Schritt 2: Danach wurde die Lösung für 22 h bei -20°C im Gefrierschrank aufbewahrt. Schritt 3: Glycin (20 µmol, 1000 Äq.) in 30 mM Borat-gepufferter Lösung (1000 µL) wurde zu der aufgetauten Oligonukleotidlösung gegeben. Der pH-Wert wurde mit einer 4M wässrigen NaOH-Lösung (ca. 130 µL) auf ca. 9.5 eingestellt. Die Reaktion wurde 1 h gerührt und anschließend wurde die Reaktion mit 1M wässriger HCl-Lösung (35 µL) gequencht. Die gesamte Reaktionsmischung wurde mit Wasser verdünnt (bis zu 5 mL) und mittels semi-präparativer HPLC aufgereinigt (Puffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O und Puffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-30% von B in 30 min; Durchflussrate = 5 mL/min).

Peptidkopplungsreaktion zwischen den Aminosäure-modifizierten Oligonukleotiden ON1-3 und ON5

Eine äquimolare Lösungsmischung von **ON1**-**3** (50 µM) und **ON5** (50 µM), die MES-Puffer pH 6 (100 mM), NaCl (100 mM) enthielt, wurde auf 90°C für 2 Minuten erhitzt und anschließend über 20 Minuten auf r.t. heruntergekühlt. DMTMM-Cl (50 mM) wurde hinzugefügt und die Reaktionslösung wurde für 20 h bei 25 ºC inkubiert. Danach wurde ein Aliquot (20 µL) der Reaktionsmischung entnommen und mittels HPLC (Puffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O und Puffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-40% von B in 45 min; Flussrate = 1 mL/min; Injektionsvolumen = 20 µL) und MALDI-TOF-Massenspektrometrie analysiert.

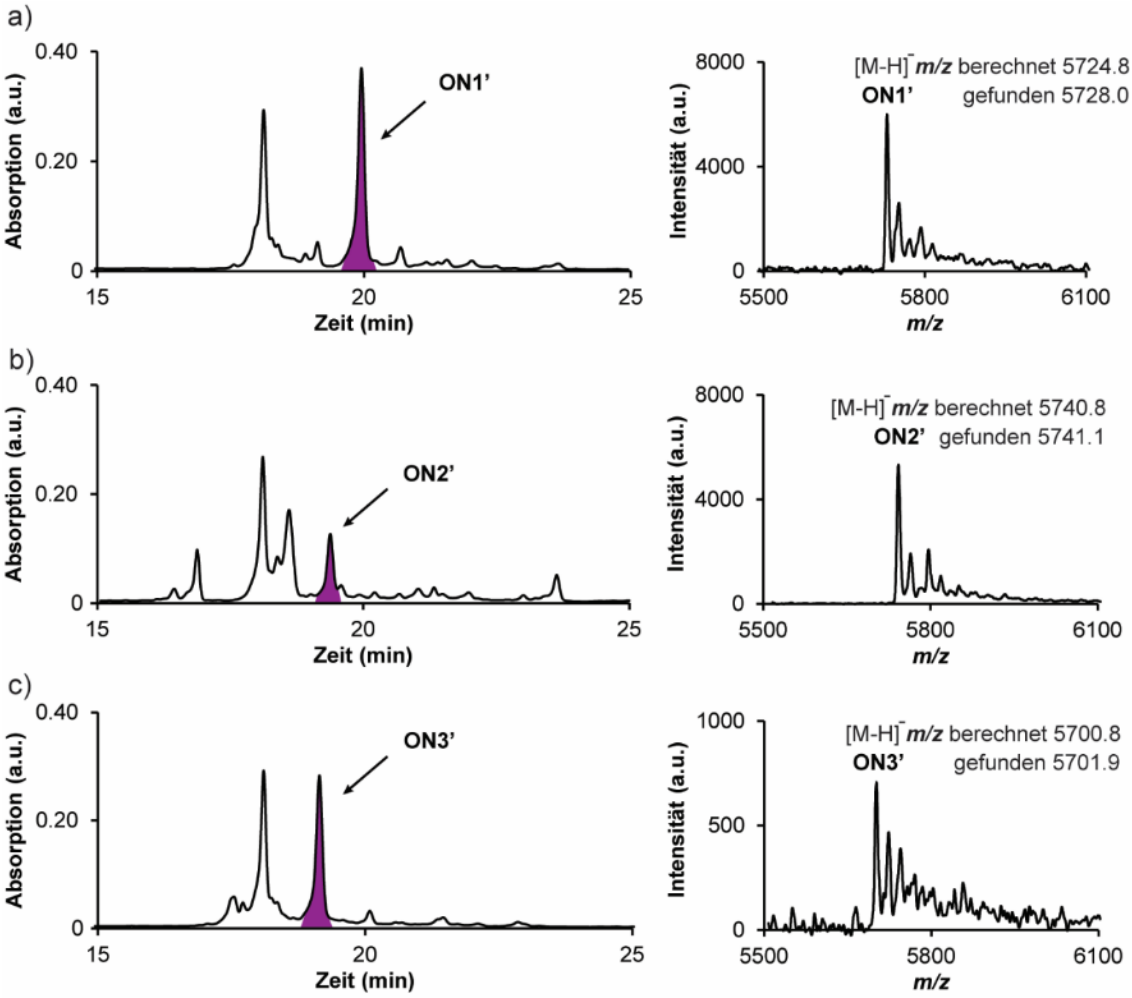


Abbildung 16: HPLC-Chromatogramme der Kopplungsreaktionen von a) **ON1**, b) **ON2** und c) **ON3** mit **ON5** und die entsprechenden MALDI-TOF Spektren der isolierten Produkte **ON1′-ON3′**.

Peptidkopplungsreaktion zwischen den Ribozymfragmenten ON10 und ON11

Eine äquimolare Lösungsmischung von **ON10** (50 µM) und **ON11** (50 µM), die MES-Puffer pH 6 (100 mM), NaCl (100 mM bzw. 1000mM) enthielt, wurde auf 90°C für 2 Minuten erhitzt und anschließend über 20 Minuten auf r.t. heruntergekühlt. DMTMM-Cl (50 mM) wurde hinzugefügt und die Reaktionslösung wurde 20 h lang bei 15 ºC inkubiert. Danach wurde ein Aliquot (20 µL) der Reaktionsmischung entnommen und mittels HPLC (Puffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O und Puffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-40% von B in 45 min; Flussrate = 1 mL/min; Injektionsvolumen = 20 µL) und MALDI-TOF-Massenspektrometrie analysiert.

Reaktionsbedingungen der katalytischen Spaltung des Substrats ON7

Zu einer Lösungsmischung aus dem Substrat **ON7** (50 µM) und dem entsprechenden Ribozym (0.5 µM), die Phosphatpuffer pH 8 (50 mM), NaCl (100 mM) enthielt, wurde $MgCl_2$ (5 mM) hinzugegeben. Die Reaktionslösung wurde für 2 h bei 25 ºC inkubiert. Danach wurde ein Aliquot (20 µL) der Reaktionsmischung entnommen und mittels HPLC (Puffer A: 0.1 M

AcOH/Et_3N pH 7 in H_2O und Puffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-40% von B in 45 min; Flussrate = 1 mL/min; Injektionsvolumen = 20 µL). Die Produktmischung aus **ON7′** und **ON7″** wurde mittels MALDI-TOF-Massenspektrometrie analysiert.

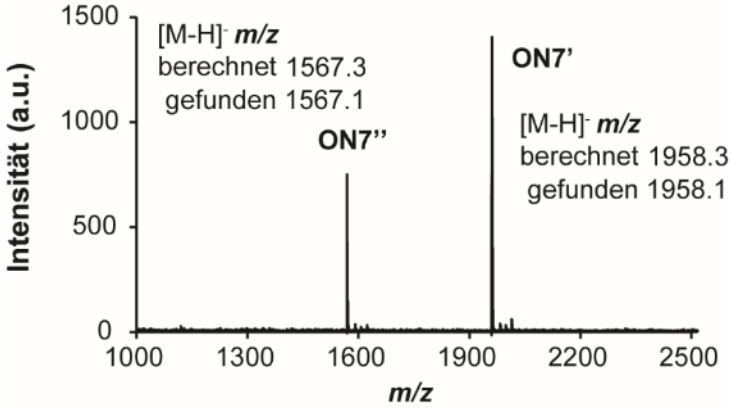


Abbildung 17. MALDI-TOF-Spektrum der Spaltungsprodukte **ON7′** und **ON7″**.

6. Abkürzungsverzeichnis

aa-AMP	Aminosäure-Adenosinmonophosphat
A	Adenin (Nukleobase), Adenosin (Nukleosid)
Ac	Acetyl
aaRS	Aminoacyl-tRNA-Synthetase
Bz	Benzoyl
BTT	5-(Benzylthio)-1*H*-tetrazol
C	Cytosin (Nukleobase, Cytidin (Nukleosid)
CE	Cyanoethyl
COSY	*correlation spectroscopy*
DBE	Dinitrobenzylester
DBU	1,8-Diazabicyclo[5.4.0]undec-7-en
DAP	Diamidophosphat
DCA	Dichloressigsäure
DCM	Dichlormethan
DIPEA	Diisopropylethylamin
DMF	*N,N*-Dimethylformamid
DMSO	Dimethylsulfoxid
DMTr	4,4′-Dimethoxytrityl
DNA	Desoxyribonukleinsäure
DCA	Dichloressigsäure
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimid
EDTA	Ethylendiamintetraacetat (Dinatrium-Salz)
ESI	*electron spray ionisation*
G	Guanin (Nukleobase), Guanosin (Nukleosid)
Gdm	Guanidinium
GlmS	Glucosamin-6-Phosphat-Synthase
HDV	Hepatitis Delta Virus
HMBC	*heteronuclear multiple bond correlation*
HPA	3-Hyroxypicolinsäure
HPLC	*high performance liquid chromatogrphy*
HRMS	*high resolution mass spectrometry*

HSQC	*heteronuclear single quantum coherence*
IPA	Isopropylalkohol
IR	Infrarotspektroskopie
KEOPS	*kinase, endopeptidase and other proteins of small size*
LC	*liquid chromatography*
LUCA	*last universal common ancestor*
MALDI-TOF	*matrix assisted laser disorption ionisation - time of flight*
MeCN	Acetonitril
MES	2-(*N*-Morpholino)-ethansulfonat
mRNA	*messenger* Ribonukleinsäure
MIC	Methylisocyanat
MS	Mssenspektrometrie
NMP	Nukleosid-Monophosphat
NMR	*nuclear magnetic resonance*
npe	*p*-Nitrophenylethyl
NTP	Nukleosid-Triphosphat
ON	Oligonukleotid
ppm	*parts per million*
PTase	Peptidyltransferase
RP	*reversed phase*
RNA	Ribonukleinsäure
rRNA	*ribosomal* Ribonukleinsäure
SSI	*self-splicing intron*
teoc	(Trimethylsilyl)ethoxycarbonyloxy
TBS	*tert*-Butyldimethylsilyl
TC-AMP	Threonylcarbamoyl-Adenosinmonophosphat
TEA	Triethylamin
THS	Tetrahydrofuran
TLC	Dünnschichtchromatographie
tRNA	*transfer* Ribonukleinsäure
U	Uracil (Nukleobase), Uridin (Nukleosid)

7. Literaturverzeichnis

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.

[2] International Human Genome Sequencing Consortium, *Nature* **2001**, *409*, 860–921.

[3] C. Darwin, *On the Origin of Species*, Harvard Univ. Press, **1859**.

[4] J. S. Wicken, *J. Theor. Biol.* **1985**, *117*, 545–561.

[5] P. M. Sharp, *Nature* **1997**, *385*, 111.

[6] G. Danger, L. L. S. d'Hendecourt, R. Pascal, *Nat. Rev. Chem.* **2020**, *4*, 102–109.

[7] C. Woese, *Proc. Natl. Acad. Sci.* **1998**, *95*, 6854–6859.

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

[9] C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci.* **1990**, *87*, 4576–4579.

[10] E. V. Koonin, *Nat. Rev. Microbiol.* **2003**, *1*, 127–136.

[11] J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, L. D. Williams, *Chem. Rev.* **2020**, *120*, 4848–4878.

[12] F. Crick, *Nature* **1970**, *227*, 561–563.

[13] C. R. Woese, *Proc. Natl. Acad. Sci. U. S. A.* **1965**, *54*, 1546–1552.

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

[15] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381–393.

[16] F. H. Westheimer, *Nature* **1986**, *319*, 534–536.

[17] H. Saito, *Nat. Rev. Mol. Cell Biol. 2022 239* **2022**, *23*, 582–582.

[18] T. R. Cech, B. L. Bass, *Ann. Rev. Biochem.* **1986**, *55*, 599–629.

[19] S. Altman, M. Baer, C. Guerrier-Takada, A. Vioque, *Trends Biochem. Sci.* **1986**, *11*, 515–518.

[20] W. Gilbert, *Nature* **1986**, *319*, 618.

[21] T. R. Cech, A. J. Zaug, P. J. Grabowski, *Cell* **1981**, *27*, 487–496.

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

[23] T. R. Cech, *Int. Rev. Cytol.* **1985**, *93*, 3–22.

[24] A. J. Zaug, T. R. Cech, *Science* **1986**, *231*, 470–475.

[25] A. J. Zaug, P. J. Grabowski, T. R. Cech, *Nature* **1983**, *301*, 578–583.

[26] A. J. Zaug, J. R. Kent, T. R. Cech, *Science* **1984**, *224*, 574–578.

[27] B. L. Bass, T. R. Cech, *Nature* **1984**, *308*, 820–826.

[28] A. J. Zaug, M. D. Been, T. R. Cech, *Nature* **1986**, *324*, 429–433.

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

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

[31] F. H. Westheimer, *Science* **1987**, *235*, 1173–1178.

[32] A. D. Ellington, X. Chen, M. Robertson, A. Syrett, *Int. J. Biochem. Cell Biol.* **2009**, *41*, 254–265.

[33] K. Le Vay, H. Mutschler, *Emerg. Top. Life Sci.* **2019**, *3*, 469–475.

[34] K. Adamala, J. W. Szostak, *Nat. Chem.* **2013**, *5*, 495–501.

[35] G. F. Joyce, J. W. Szostak, *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a034801.

[36] S. A. Benner, *Acc. Chem. Res.* **2004**, *37*, 784–797.

[37] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**, *120*, 4766–4805.

[38] S. L. Miller, *Science* **1953**, *117*, 528–529.

[39] H. C. Urey, *Proc. Natl. Acad. Sci.* **1952**, *38*, 351–363.

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

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

[42] S. Islam, M. W. Powner, *Chem* **2017**, *2*, 470–501.

[43] J. D. Sutherland, *Angew. Chem. Int. Ed.* **2016**, *55*, 104–121.

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

[45] H. J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11315–11320.

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

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

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

[49] 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.

[50] S. A. Benner, E. A. Bell, E. Biondi, R. Brasser, T. Carell, H.-J. Kim, S. J. Mojzsis, A. Omran, M. A. Pasek, D. Trail, *ChemSystemsChem* **2020**, *2*, e1900035.

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

[52] F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Kröger, L. Le Roy, A. J. MacDermott, et al., *Science* **2015**, *349*, aab0689-1.

[53] Y. Oba, T. Koga, Y. Takano, N. O. Ogawa, N. Ohkouchi, K. Sasaki, H. Sato, D. P. Glavin, J. P. Dworkin, H. Naraoka, et al., *Nat. Commun. 2023 141* **2023**, *14*, 1–9.

[54] J. M. Hollis, S. N. Vogel, L. E. Snyder, P. R. Jewell, F. J. Lovas, *Astrophys. J.* **2001**, *554*, L81–L85.

[55] 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.

[56] R. Martín-Domenech, V. M. Rivilla, I. Jimenez-Serra, D. Quenard, L. Testi, L. Testi, L. Testi, J. Martín-Pintado, *Mon. Not. R. Astron. Soc.* **2017**, *469*, 2230–2234.

[57] R. Breslow, *Tetrahedron Lett.* **1959**, *145*, 22–26.

[58] R. Larralde, M. P. Robertscn, S. L. Miller, *Proc. Natl. Acad. Sci.* **1995**, *92*, 8158–8160.

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

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

[61] H. J. Cleaves, *Evol. Educ. Outreach* **2012**, *5*, 342–360.

[62] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 1–9.

[63] M. P. Robertson, G. F. Joyce, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, 1.

[64] A. Pressman, C. Blanco, I. A. Chen, *Curr. Biol.* **2015**, *25*, R953–R963.

[65] D. M. Fialho, T. P. Roche, N. V. Hud, *Chem. Rev.* **2020**, *120*, 4806–4830.

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

[67] 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.

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

[69] K.-U. Schöning, P. Scholz, S. Guntha, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347–1351.

[70] R. Yi, R. Kern, P. Pollet, H. Lin, R. Krishnamurthy, C. L. Liotta, *Chem. Eur. J.* **2023**, *29*, e202202816.

[71] W. Zhang, S. C. Kim, C. P. Tam, V. S. Lelyveld, S. Bala, J. C. Chaput, J. W. Szostak, *Nucleic Acids Res.* **2021**, *49*, 646–656.

[72] 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.

[73] J. K. Watts, N. Martín-Pintado, I. Gómez-Pinto, J. Schwartzentruber, G. Portella, M. Orozco, C. González, M. J. Damha, *Nucleic Acids Res.* **2010**, *38*, 2498–2511.

[74] S. C. Kim, L. Zhou, W. Zhang, D. K. O'Flaherty, V. Rondo-Brovetto, J. W. Szostak, *J. Am. Chem. Soc.* **2020**, *142*, 2317–2326.

[75] I. Anosova, E. A. Kowal, M. R. Dunn, J. C. Chaput, W. D. V. Horn, M. Egli, *Nucleic Acids Res.* **2016**, *44*, 1007–1021.

[76] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S. Y. Peak-Chew, S. H. McLaughlin, et al., *Science* **2012**, *336*, 341–344.

[77] V. B. Pinheiro, D. Loakes, P. Holliger, *BioEssays* **2013**, *35*, 113–122.

[78] F. Eckstein, *Nucl. Acid Ther.* **2014**, *24*, 374–387.

[79] T. C. Roberts, R. Langer, M. J. A. Wood, *Nat. Rev. Drug Discov.* **2020**, *19*, 673–694.

[80] P. Li, Z. A. Sergueeva, M. Dobrikov, B. R. Shaw, *Chem. Rev.* **2007**, *107*, 4746–4796.

[81] J. kang Chen, R. G. Schultz, D. H. Lioyd, S. M. Gryaznov, *Nucleic Acids Res.* **1995**, *23*, 2661–2668.

[82] M. Faria, D. G. Spiller, C. Dubertret, J. S. Nelson, M. R. H. White, D. Scherman, C. Hélène, C. Giovannangeli, *Nat. Biotechnol. 2001 191* **2001**, *19*, 40–44.

[83] M. Egholm, O. Buchardt, P. E. Nielsen, R. H. Berg, *J. Am. Chem. Soc.* **1992**, *114*, 1895–1897.

[84] C. Sharma, S. K. Awasthi, *Chem. Biol. Drug Des.* **2017**, *89*, 16–37.

[85] K. E. Nelson, M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3868.

[86] P. E. Nielsen, *Chem. Biodivers.* **2007**, *4*, 1996–2002.

[87] Y. Ura, J. M. Beierle, L. J. Leman, L. E. Orgel, M. R. Ghadiri, *Science* **2009**, *325*, 73–77.

[88] A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn, P. Holliger, *Nature* **2015**, *518*, 427–430.

[89] S. C. Kim, D. K. O'Flaherty, C. Giurgiu, L. Zhou, J. W. Szostak, *J. Am. Chem. Soc.* **2021**, *143*, 3267–3279.

[90] F. R. Bowler, C. K. W. Chan, C. D. Duffy, B. Gerland, S. Islam, M. W. Powner, J. D. Sutherland, J. Xu, *Nat. Chem.* **2013**, *5*, 383–389.

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

[92] B. J. Cafferty, N. V. Hud, *Curr. Opin. Chem. Biol.* **2014**, *22*, 146–157.

[93] A. D. Keefe, S. L. Miller, *J. Mol. Evol.* **1995**, *41*, 693–702.

[94] R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 64–66.

[95] R. Lohrmann, L. E. Orgel, *Science* **1971**, *171*, 490–494.

[96] 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.

[97] A. M. Schoffstall, *Orig. Life* **1976**, *7*, 399–412.

[98] G. Costanzo, R. Saladino, C. Crestini, F. Ciciriello, E. Di Mauro, *J. Biol. Chem.* **2007**, *282*, 16729–16735.

[99] A. W. Schwartz, *Philos. Trans. R. Soc. B* **2006**, *361*, 1743–1749.

[100] A. Schwartz, C. Ponnamperuma, *Nature* **1968**, *218*, 443.

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

[102] A. W. Schwartz, *J. Chem. Soc. D* **1969**, 1393.

[103] I. Sibilska, Y. Feng, L. Li, J. Yin, *Orig. Life Evol. Biosph.* **2018**, *48*, 277–287.

[104] M. Karki, C. Gibard, S. Bhowmik, R. Krishnamurthy, *Life* **2017**, *7*, 1–28.

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

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

[107] H. Lin, E. I. Jiménez, J. T. Arriola, U. F. Müller, R. Krishnamurthy, *Angew. Chem. Int. Ed.* **2022**, *61*, e202113625.

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

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

[110] R. Lohrmann, *J. Mol. Evol.* **1977**, *10*, 137–154.

[111] J. Sulston, R. Lohrmann, L. E. Orgel, H. T. Miles, *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *59*, 726–733.

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

[113] J. P. Ferris, *Orig. Life Evol. Biosph.* **2002**, *32*, 311–332.

[114] W. Huang, J. P. Ferris, *J. Am. Chem. Soc.* **2006**, *128*, 8914–8919.

[115] T. Inoue, L. E. Orgel, *J. Mol. Biol.* **1982**, *162*, 201–217.

[116] P. K. Bridson, L. E. Orgel, *J. Mol. Biol.* **1980**, *144*, 567–577.

[117] M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed.* **2015**, *54*, 14559–14563.

[118] R. Lohrmann, L. E. Orgel, *Nature* **1973**, *244*, 418–420.

[119] M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed.* **2015**, *54*, 14564–14569.

[120] H. Griesser, P. Tremmel, E. Kervio, C. Pfeffer, U. E. Steiner, C. Richert, *Angew. Chem. Int. Ed.* **2017**, *56*, 1219–1223.

[121] H. Griesser, M. Bechthold, P. Tremmel, E. Kervio, C. Richert, *Angew. Chem. Int. Ed.* **2017**, *56*, 1224–1228.

[122] P. Tremmel, H. Griesser, U. E. Steiner, C. Richert, *Angew. Chem. Int. Ed.* **2019**, *58*, 13087–13092.

[123] M. Räuchle, G. Leveau, C. Richert, *European J. Org. Chem.* **2020**, *2020*, 6966–6975.

[124] A. Schimpl, R. M. Lemmon, M. Calvin, *Science* **1965**, *147*, 149–150.

[125] D. W. Nooner, E. Sherwood, M. A. More, J. Oró, *J. Mol. Evol.* **1977**, *10*, 211–220.

[126] P. BERG, *J. Biol. Chem.* **1958**, *233*, 608–611.

[127] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **2015**, *7*, 301–307.

[128] B. M. A. G. Piette, J. G. Heddle, *Trends Ecol. Evol.* **2020**, *35*, 397–406.

[129] M. Eigen, P. Schuster, *Naturwissenschaften* **1977**, *64*, 541–565.

[130] T. R. Cech, *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 4360–4363.

[131] D. P. Bartel, J. W. Szostak, *Science* **1993**, *261*, 1411–1418.

[132] E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **1995**, *269*, 364–370.

[133] N. Paul, G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12733–12740.

[134] D. E. Kim, G. F. Joyce, *Chem. Biol.* **2004**, *11*, 1505–1512.

[135] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229–1232.

[136] B. J. Lam, G. F. Joyce, *Nat. Biotechnol.* **2009**, *27*, 288–292.

[137] A. Mariani, C. Bonfio, C. M. Johnson, J. D. Sutherland, *Biochemistry* **2018**, *57*, 6382–6386.

[138] E. J. Hayden, G. von Kiedrowski, N. Lehman, *Angew. Chem. Int. Ed.* **2008**, *47*, 8424–8428.

[139] N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden, N. Lehman, *Nature* **2012**, *491*, 72–77.

[140] T. S. Jayathilaka, N. Lehman, *ChemBioChem* **2018**, *19*, 217–220.

[141] L. Zhou, D. K. O'Flaherty, J. W. Szostak, *J. Am. Chem. Soc.* **2020**, *142*, 15961–15965.

[142] F. Wachowius, P. Holliger, *ChemSystemsChem* **2019**, *1*, 12–15.

[143] D. W. Morgens, *J. Mol. Evol.* **2013**, *77*, 185–196.

[144] M. Root-Bernstein, R. Root-Bernstein, *J. Theor. Biol.* **2015**, *367*, 130–158.

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

[146] H. F. Noller, *Science* **2005**, *309*, 1508–1514.

[147] T. A. Steitz, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253.

[148] A. Sievers, M. Beringer, M. V. Rodnina, R. Wolfenden, *Proc. Natl Acad. Sci. USA* **2004**, *101*, 7897–7901.

[149] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 905–920.

[150] V. Ramakrishnan, *Cell* **2002**, *108*, 557–572.

[151] P. B. Moore, T. A. Steitz, *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a003780.

[152] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vonrheln, T. Hartsch, V. Ramakrishnan, *Nature* **2000**, *407*, 327–339.

[153] J. Harms, F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi, A. Yonath, *Cell* **2001**, *107*, 679–688.

[154] K. Tamura, *J. Biosci.* **2011**, *36*, 921–928.

[155] J. C. Bowman, N. V. Hud, L. D. Williams, *J. Mol. Evol.* **2015**, *80*, 143–161.

[156] G. F. Joyce, *Nature* **2002**, *418*, 214–221.

[157] J. Demongeot, H. Seligmann, *Biosystems* **2022**, *222*, 104796.

[158] J. Demongeot, H. Seligmann, *Gene* **2020**, *738*, 144436.

[159] F. H. C. Crick, S. Brenner, A. Klug, G. Pieczenik, *Orig. Life* **1976**, *7*, 389–397.

[160] H. F. Noller, V. Hoffarth, L. Zimniak, *Science* **1992**, *256*, 1416–1419.

[161] L. P. Gavrilova, O. E. Kostiashkina, V. E. Koteliansky, N. M. Rutkevitch, A. S. Spirin, *J. Mol. Biol.* **1976**, *101*, 537–552.

[162] D. R. Southworth, J. L. Brunelle, R. Green, *J. Mol. Biol.* **2002**, *324*, 611–623.

[163] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *Science* **1965**, *147*, 1462–1465.

[164] M. Di Giulio, *J. Theor. Biol.* **2004**, *226*, 89–93.

[165] M. Di Giulio, *J. Theor. Biol.* **2019**, *480*, 99–103.

[166] Y. L. J. Pang, K. Poruri, S. A. Martinis, *Wiley Interdiscip. Rev. RNA* **2014**, *5*, 461–480.

[167] M. B. Hoagland, E. B. Keller, P. C. Zamecnik, *J. Biol. Chem.* **1956**, *218*, 345–358.

[168] M. Delarue, *Curr. Opin. Struct. Biol.* **1995**, *5*, 48–55.

[169] M. Ibba, D. Soll, *Annu. Rev. Biochem.* **2000**, *69*, 617–650.

[170] R. B. Ganesh, S. J. Maerkl, *Front. Bioeng. Biotechnol.* **2022**, *10*, 971.

[171] C. S. Francklyn, E. A. First, J. J. Perona, Y. M. Hou, *Methods* **2008**, *44*, 100–118.

[172] M. Di Giulio, *J. Mol. Evol.* **1997**, *45*, 571–578.

[173] E. Janzen, C. Blanco, H. Peng, J. Kenchel, I. A. Chen, *Chem. Rev.* **2020**, *120*, 4879–4897.

[174] J. Morimoto, Y. Hayashi, K. Iwasaki, H. Suga, *Acc. Chem. Res.* **2011**, *44*, 1359–1368.

[175] K. Tamura, P. Schimmel, *Proc. Natl Acad. Sci. USA* **2003**, *100*, 8666–8669.

[176] L. F. Wu, M. Su, Z. Liu, S. J. Bjork, J. D. Sutherland, *J. Am. Chem. Soc.* **2021**, *143*, 11836–11842.

[177] Z. Liu, D. Beaufils, J. C. Rossi, R. Pascal, *Sci. Rep.* **2014**, *4*, 7440.

[178] R. M. Turk, M. Illangasekare, M. Yarus, *J. Am. Chem. Soc.* **2011**, *133*, 6044–6050.

[179] K. R. Lynn, *J. Phys. Chem.* **1965**, *69*, 687–689.

[180] E. Kinne-Saffran, R. K. H. Kinne, *Am. J. Nephrol.* **1999**, *19*, 290–294.

[181] P. Schimmel, *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 45–58.

[182] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110–7131.

[183] T. Suzuki, *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 375–392.

[184] M. López-Estepa, A. Ardá, M. Savko, A. Round, W. E. Shepard, M. Bruix, M. Coll, F. J. Fernández, J. Jiménez-Barbero, M. C. Vega, *PLoS One* **2015**, *10*, e0118606.

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

[186] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, et al., *Nucleic Acids Res.* **2013**, *41*.

[187] L. Perrochia, E. Crozat, A. Hecker, W. Zhang, J. Bareille, B. Collinet, H. Van Tilbeurgh, P. Forterre, T. Basta, *Nucleic Acids Res.* **2013**, *41*, 1953–1964.

[188] M. P. Schweizer, G. B. Chheda, L. Baczynskyj, R. H. Halit, *Biochemistry* **1969**, *8*, 3283–3289.

[189] D. M. Reddy, P. F. Crain, C. G. Edmonds, R. Gupta, T. Hashizume, K. O. Stetter, F. Widde, J. A. Mccloskey, *Nucleic Acids Res.* **1992**, *20*, 5607–5615.

[190] A. Nagao, M. Ohara, K. Miyauchi, S. I. Yokobori, A. Yamagishi, K. Watanabe, T. Suzuki, *Nat. Struct. Mol. Biol.* **2017**, *24*, 778–782.

[191] M. P. Schweizer, K. McGrath, L. Baczynskyj, *Biochem. Biophys. Res. Commun.* **1970**,

40, 1046–1052.

[192] K. Miyauchi, S. Kimura, T. Suzuki, *Nat. Chem. Biol.* **2013**, *9*, 105–111.

[193] Q. Qian, J. F. Curran, G. R. Björk, *J. Bacteriol.* **1998**, *180*, 1808–1813.

[194] M. Matuszewski, J. Wojciechowski, K. Miyauchi, Z. Gdaniec, W. M. Wolf, T. Suzuki, E. Sochacka, *Nucleic Acids Res.* **2017**, *45*, 2137–2149.

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

[196] J. Beenstock, F. Sicheri, *Nucleic Acids Res.* **2021**, *49*, 10818–10834.

[197] C. Deutsch, B. El Yacoubi, V. De Crécy-Lagard, D. Iwata-Reuyl, *J. Biol. Chem.* **2012**, *287*, 13666–13673.

[198] C. T. Lauhon, *Biochemistry* **2012**, *51*, 8950–8963.

[199] A. Kitamura, M. Nishimoto, T. Sengoku, R. Shibata, G. Jäger, G. R. Björk, H. Grosjean, S. Yokoyama, Y. Bessho, *J. Biol. Chem.* **2012**, *287*, 43950–43960.

[200] M. Blaise, H. D. Becker, G. Keith, C. Cambillau, J. Lapointe, R. Giegé, D. Kern, *Nucleic Acids Res.* **2004**, *32*, 2768–2775.

[201] H. Grosjean, V. De Crécy-Lagard, G. R. Björk, *Trends Biochem. Sci.* **2004**, *29*, 519–522.

[202] T. Muramatsu, S. Yokoyama, N. Horie, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, T. Miyazawa, *J. Biol. Chem.* **1988**, *263*, 9261–9267.

[203] I. Moukadiri, S. Prado, J. Piera, A. Velázquez-campoy, G. R. Björk, M. E. Armengod, *Nucleic Acids Res.* **2009**, *37*, 7177–7193.

[204] J. W. Szostak, *J. Syst. Chem.* **2012**, *3*, 2.

[205] B. R. Francis, *J. Mol. Evol. 2013 774* **2013**, *77*, 134–158.

[206] A. J. Doig, A. J. Doig, *FEBS J.* **2017**, *284*, 1296–1305.

[207] P. T. S. van der Gulik, D. Speijer, *Life* **2015**, *5*, 230–246.

[208] H. Grosjean, E. Westhof, *Nucleic Acids Res.* **2016**, *44*, 8020–8040.

[209] M. Di Giulio, *J. Theor. Biol.* **1998**, *191*, 191–196.

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

[211] M. P. Robertson, S. L. Miller, *Science* **1995**, *268*, 702–705.

[212] H. F. Noller, *RNA* **2004**, *10*, 1833–1837.

[213] J. W. Chin, *Nature* **2017**, *550*, 53–60.

[214] E. Janzen, Y. Shen, A. Vázquez-Salazar, Z. Liu, C. Blanco, J. Kenchel, I. A. Chen, *Nat. Commun.* **2022**, *13*, 1–12.

[215] A. L. Weber, J. C. Lacey, *J. Mol. Evol.* **1975**, *6*, 309–320.

[216] A. L. Weber, L. E. Orgel, *J. Mol. Evol.* **1978**, *11*, 9–16.

[217] M. Yarus, *Philos. Trans. R. Soc. B* **2011**, *366*, 2902–2909.

[218] M. Illangasekare, G. Sanchez, T. Nickles, M. Yarus, *Science* **1995**, *267*, 643–647.

[219] N. V. Chumachenko, Y. Novikov, M. Yarus, *J. Am. Chem. Soc.* **2009**, *131*, 5257–5263.

[220] R. M. Turk, N. V. Chumachenko, M. Yarus, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4585–4589.

[221] M. Illangasekare, M. Yarus, *RNA* **1999**, *5*, 1482–1489.

[222] O. Khersonsky, D. S. Tawfik, *Annu. Rev. Biochem.* **2010**, *79*, 471–505.

[223] E. A. Schultes, D. P. Bartel, *Science* **2000**, *289*, 448–452.

[224] H. Saito, H. Suga, *J. Am. Chem. Soc.* **2001**, *123*, 7178–7179.

[225] H. Saito, K. Watanabe, H. Suga, *RNA* **2001**, *7*, 1867–1878.

[226] M. Illangasekare, M. Yarus, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5470–5475.

[227] N. Lee, Y. Bessho, K. Wei, J. W. Szostak, H. Suga, *Nat. Struct. Biol.* **2000**, *7*, 28–33.

[228] H. Saito, D. Kourouklis, H. Suga, *EMBO J.* **2001**, *20*, 1797–1806.

[229] N. Li, F. Huang, *Biochemistry* **2005**, *44*, 4582–4590.

[230] H. Xiao, H. Murakami, H. Suga, A. R. Ferré-D'Amaré, *Nature* **2008**, *454*, 358–361.

[231] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357–359.

[232] Y. Goto, H. Murakami, H. Suga, *RNA* **2008**, *14*, 1390–1398.

[233] T. Kawakami, H. Murakami, H. Suga, *Chem. Biol.* **2008**, *15*, 32–42.

[234] A. Ohta, H. Murakami, E. Higashimura, H. Suga, *Chem. Biol.* **2007**, *14*, 1315–1322.

[235] Y. Goto, H. Suga, *J. Am. Chem. Soc.* **2009**, *131*, 5040–5041.

[236] A. D. Pressman, Z. Liu, E. Janzen, C. Blanco, U. F. Müller, G. F. Joyce, R. Pascal, I. A. Chen, *J. Am. Chem. Soc.* **2019**, *141*, 6213–6223.

[237] H. J. Cleaves, *J. Theor. Biol.* **2010**, *263*, 490–498.

[238] M. Frenkel-Pinter, M. Frenkel-Pinter, M. Samanta, G. Ashkenasy, L. J. Leman, L. J. Leman, *Chem. Rev.* **2020**, *120*, 4707–4765.

[239] A. L. Weber, S. L. Miller, *J. Mol. Evol.* **1981**, *17*, 273–284.

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

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

[242] A. S. Burton, J. C. Stern, J. E. Elsila, D. P. Glavin, J. P. Dworkin, *Chem. Soc. Rev.* **2012**, *41*, 5459–5472.

[243] X. Guo, M. Su, *Int. J. Mol. Sci.* **2023**, *24*, 197.

[244] K. Tamura, P. Schimmel, *Science* **2004**, *305*, 1253.

[245] K. Tamura, P. R. Schimmel, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13750–13752.

[246] N. S. M. D. Wickramasinghe, M. P. Staves, J. C. Lacey, *Biochemistry* **1991**, *30*, 2768–2772.

[247] S. J. Roberts, Z. Liu, J. D. Sutherland, *J. Am. Chem. Soc.* **2022**, *144*, 4254–4259.

[248] B. Jash, C. Richert, *Chem. Sci.* **2020**, *11*, 3487–3494.

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

[250] J. L. Shim, R. Lohrmann, L. E. Orgel, *J. Am. Chem. Soc.* **1974**, *96*, 5283–5284.

[251] A. Radakovic, T. H. Wright, V. S. Lelyveld, J. W. Szostak, *Biochemistry* **2021**, *60*, 477–488.

[252] A. Radakovic, S. DasGupta, T. H. Wright, H. R. M. Aitken, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116840119.

[253] N. Prywes, J. C. Blain, F. Del Frate, J. W. Szostak, *eLife* **2016**, *5*, 1–14.

[254] R. Lohrmann, L. E. Orgel, *Tetrahedron* **1978**, *34*, 853–855.

[255] L.-F. Wu, Z. Liu, S. J. Roberts, M. Su, J. W. Szostak, J. D. Sutherland, *J. Am. Chem. Soc.* **2022**, *144*, 13920–13927.

[256] K. J. Hertel, D. Herschlag, O. C. Uhlenbeck, *EMBO J.* **1996**, *15*, 3751–3757.

[257] R. M. Jimenez, J. A. Polanco, A. Lupták, *Trends Biochem. Sci.* **2015**, *40*, 648–661.

[258] C. H. Webb, N. J. Riccitelli, D. J. Ruminski, A. Lupták, *Science* **2009**, *326*, 953.

[259] M. J. Fedor, *Annu. Rev. Biophys.* **2009**, *38*, 271–299.

[260] C. E. Weinberg, Z. Weinberg, C. Hammann, *Nucleic Acids Res.* **2019**, *47*, 9480–9494.

[261] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, G. Bruening, *Science* **1986**, *231*, 1577–1580.

[262] A. C. Forster, R. H. Symons, *Cell* **1987**, *50*, 9–16.

[263] M. De, L. A. Pen˜a, P. Pen˜a, I. Garci´a, G. Garci´a-Robles, *RNA* **2010**, *16*, 1943–1950.

[264] J. Perreault, Z. Weinberg, A. Roth, O. Popescu, P. Chartrand, G. Ferbeyre, R. R. Breaker, *PLOS Comput. Biol.* **2011**, *7*, e1002031.

[265] C. Seehafer, A. Kalweit, G. Steger, S. Gräf, C. Hammann, *RNA* **2011**, *17*, 21–26.

[266] K. Salehi-Ashtiani, J. W. Szostak, *Nature* **2001**, *414*, 82–84.

[267] H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature* **1994**, *372*, 68–74.

[268] Y. I. Chi, M. Martick, M. Lares, R. Kim, W. G. Scott, S. H. Kim, *PLOS Biol.* **2008**, *6*, e234.

[269] A. R. Ferré-D'Amaré, W. G. Scott, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a003574.

[270] M. Anderson, E. P. Schultz, M. Martick, W. G. Scott, *J. Mol. Biol.* **2013**, *425*, 3790–3798.

[271] M. Roychowdhury-Saha, D. H. Burke, *RNA* **2007**, *13*, 841–848.

[272] D. McKay, *RNA* **1996**, *2*, 395–403.

[273] W. G. Scott, J. T. Finch, A. Klug, *Cell* **1995**, *81*, 991–1002.

[274] W. G. Scott, J. B. Murray, J. R. P. Arnold, B. L. Stoddard, A. Klug, *Science* **1996**, *274*, 2065–2069.

[275] H. Yang, F. Jossinet, N. Leontis, L. Chen, J. Westbrook, H. Berman, E. Westhof, *Nucleic Acids Res.* **2003**, *31*, 3450–3460.

[276] A. Khvorova, A. Lescoute, E. Westhof, S. D. Jayasena, *Nat. Struct. Mol. Biol.* **2003**, *10*, 708–712.

[277] M. Martick, W. G. Scott, *Cell* **2006**, *126*, 309–320.

[278] A. F. McKay, *Chem. Rev.* **1952**, *51*, 301–346.

[279] I. Fernández, P. Hervés, M. Parajó, *J. Phys. Org. Chem.* **2008**, *21*, 713–717.

[280] J. Vušurović, E. M. Schneeberger, K. Breuker, *ChemistryOpen* **2017**, *6*, 739–750.

[281] C. Riml, A. Lusser, E. Ennifar, R. Micura, *J. Org. Chem.* **2017**, *82*, 7939–7945.

[282] F. Himmelsbach, B. S. Schultz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59–72.

[283] G. Leszczynska, P. Leonczak, A. Dziergowska, A. Malkiewicz, *Nucleosides, Nucleotides and Nucleic Acids* **2013**, *32*, 599–616.

[284] T. Sasami, Y. Odawara, A. Ohkubo, M. Sekine, K. Seio, *Tetrahedron Lett.* **2007**, *48*, 5325–5329.

[285] S. P. Dutta, C. I. Hong, G. P. Murphy, A. Mittelman, G. B. Chheda, *Biochemistry* **2002**, *14*, 3144–3151.

Anhang I

Chemistry–A European Journal

Supporting Information

Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World

Milda Nainytė,[a] Felix Müller,[a] Giacomo Ganazzoli,[a] Chun-Yin Chan,[a] Antony Crisp,[a] Daniel Globisch,[b] and Thomas Carell*[a]

1. General Experimental Methods

Chemicals were purchased from Sigma-Aldrich, TCI, Fluka, ABCR, Carbosynth or Acros Organics and used without further purification. Some of the strands were purchased from Metabion or Ella Biotech. Reagent-grade dry solvents (Sigma-Aldrich, Acros Organics) were stored over molecular sieves and handled under inert gas atmosphere. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F254 TLC plates from Merck KGaA. Flash column chromatography was performed on Geduran® Si60 (40-63 µm) silica gel from Merck KGaA. NMR spectra were recorded on Bruker AVIIIHD 400 spectrometers (400 MHz). ^{1}H NMR shifts were calibrated to the residual solvent resonances: DMSO-d_6 (2.50 ppm), $CDCl_3$ (7.26 ppm), Acetone-d_6 (2.05 ppm), CD_2Cl_2 (5.32 ppm). ^{13}C NMR shifts were calibrated to the residual solvent: DMSO-d_6 (39.52 ppm), $CDCl_3$ (77.16 ppm), Acetone-d_6 (29.84 ppm), CD_2Cl_2 (53.84 ppm). All NMR spectra were analysed using the program MestreNova 10.0.1 from Mestrelab Research S. L. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of the Ludwigs-Maximilians-Universität München on the spectrometer MAT 90 (ESI) from Thermo Finnigan GmbH. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system. All substances were directly applied as solids or on the ATR unit. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow of 0.5 ml/min, a gradient of 0-30% of buffer B in 45 min was applied. Preparative RP-HPLC was performed on a HPLC Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel using a flow of 5 ml/min, a gradient of 0-25% of buffer B in 45 min was applied for the purifications. Oligonucleotides were purified using the following buffer system: 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 pH values of buffers were adjusted using a MP 220 pH-meter (Metter Toledo). Oligonucleotides were detected at wavelength: 260 nm. Melting profiles were measured on a JASCO V-650 spectrometer. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). For strands containing artificial bases, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a Bruker Autoflex II. For MALDI-TOF

measurements, the samples were desalted on a 0.025 µm VSWP filter (Millipore) against ddH_2O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

2. Synthesis of the Phosphoramidite Building-Blocks

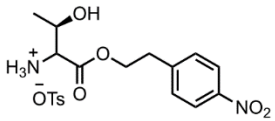
2.1. Synthesis of RNA building blocks

Npe-protection of carboxy group of amino acid

The reaction was performed according to the procedure published before. [1]

L-amino acid (1 eq.), 2-(4-nitrophenyl)ethanol (npe-OH, 3 eq.) and TsOH (3 eq.) were refluxed in toluene overnight in a *Dean-Stark* apparatus. The solution was cooled to room temperature and Et_2O was added. The oily residue was decanted, and the upper layer was removed to collect the oil. Precipitation of was induced by adding to the oil MeOH and Et_2O.

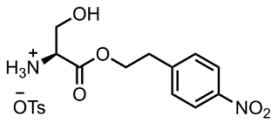
Compound 4



Yield: 70%; **IR**: $\tilde{\nu}$ = 3401 (w), 2930 (s), 2892 (s), 2858 (s), 1730 (s), 1510 (vs), 1465 (s), 1300 (vs), 1258 (s), 1167 (s), 1010 (s), 895 (w), 832 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.2 – 8.1 (m, 5H), 7.58 (d, *J* = 8.8 Hz, 2H), 7.48 (d, *J* = 8.2 Hz, 2H), 7.11 (d, *J* = 8.2 Hz, 2H), 5.6 (br. s., 1H), 4.45 (t, *J* = 6.2 Hz, 2H), 4.1 – 4.0 (m, 1H), 3.89 (d, *J* = 4.0 Hz, 1H), 3.10 (t, *J* = 6.2 Hz, 2H), 2.29 (s, 3H), 1.14 (d, *J* = 6.6 Hz, 3H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.2, 146.5, 146.3, 137.7, 130.4, 128.1, 125.6, 123.5, 65.4, 64.9, 57.9, 33.8, 20.7, 20.0; **HRMS (ESI)**: calculated for $C_{12}H_{17}N_2O_5^+$: m/z = 269.1137 $[M+H]^+$; found: m/z = 269.1140 $[M+H]^+$.

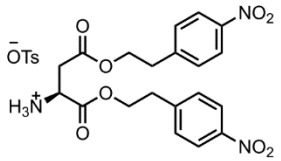
The analytical data is in agreement with the literature.[1]

Compound 5



Yield: 75%; **IR**: $\tilde{\nu}$ = 3400 (w), 2931 (s), 2894 (s), 2858 (s), 1730 (s), 1510 (vs), 1465 (s), 1300 (vs), 1258 (s), 1167 (s), 1010 (s), 895 (w), 832 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.34 (br. s, 3H), 8.16 (d, *J* = 8.9 Hz, 2H), 7.58 (d, *J* = 8.9 Hz, 2H), 7.50 (d, *J* = 8.0 Hz, 2H), 7.12 (d, *J* = 8.0 Hz, 2H), 4.49 – 4.36 (m, 2H), 4.11 (br. s, 1H), 3.74 – 3.73 (m, 2H), 3.08 (t, *J* = 6.4 Hz, 2H), 2.28 (s, 3H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.0, 146.3, 138.0, 130.4, 128.2, 125.6, 123.5, 65.4, 59.5, 54.2, 33.9, 20.9; **HRMS (ESI)**: calculated for $C_{11}H_{15}N_2O_5^+$: m/z = 255.0981 $[M+H]^+$; found: m/z = 255.0977 $[M+H]^+$.

Compound 6



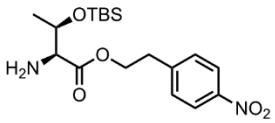
Yield: 82 %; **IR**: $\tilde{\nu}$ = 3402 (w), 2933 (s), 2894 (s), 2858 (s), 1730 (s), 1689 (s), 1514 (vs), 1469 (s), 1310 (vs), 1258 (s), 1167 (s), 1010 (s), 895 (w), 834 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.42 (s, 3H), 8.16 – 8.11 (m, 4H), 7.55 – 7.48 (m, 6H), 7.12 (d, *J* = 7.9 Hz, 2H), 4.37 (t, *J* = 6.3 Hz, 2H), 4.34 – 4.30 (m, 1H), 4.26 (q, *J* = 6.3 Hz, 2H), 3.01 (q, *J* = 6.5 Hz, 4H), 2.85 (qd, *J* = 17.5, 5.1 Hz, 2H), 2.28 (s, 3H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 169.0, 168.1, 146.4, 146.3, 145.3, 138.0, 130.3, 130.2, 128.2, 125.6, 123.5, 123.5, 65.7, 64.7, 48.4, 34.1, 33.8, 33.7, 20.8; **HRMS (ESI)**: calculated for $C_{20}H_{22}N_3O_8^+$: m/z = 432.1401 $[M+H]^+$; found: m/z = 432.1405 $[M+H]^+$.

General procedure of hydroxy group protection with TBSCl

The reaction was performed according to the procedure published before. [1]

Npe-protected ester (1 eq.) was dissolved in pyridine and treated with one half of TBSCl (3 eq.) and 1*H*-imidazole (3 eq.). After 10 min, the second half was added, and the reaction mixture was left to stir at room temperature overnight. The mixture was diluted with CH_2Cl_2 and washed successively with sat. $NaHCO_3$ solution and H_2O. The organic layer was dried, evaporated and purified by flash chromatography eluting with CH_2Cl_2/MeOH (10/1, v/v) to afford the target compound as an oil.

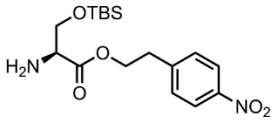
Compound 7



Yield: 94%; **IR**: $\tilde{\nu}$ = 3854 (w), 3745 (w), 2930 (w), 2856 (w), 1735 (vs), 1601 (s), 1518 (vs), 1472 (w), 1463 (w), 1374 (w), 1344 (vs), 1251 (s), 1155 (s), 1076 (s), 967 (s), 835 (s), 775 (s), 747 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.16 (d, *J* = 8.6 Hz, 2H), 7.38 (d, *J* = 8.6 Hz, 2H), 4.41 (dt, *J* = 11.0, 6.8 Hz, 1H), 4.29 – 4.16 (m, 2H), 3.24 (d, *J* = 2.8 Hz, 1H), 3.06 (t, *J* = 6.8 Hz, 2H), 1.20 (d, *J* = 6.3 Hz, 3H), 0.80 (s, 9H), -0.01 (s, 3H), -0.10 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 174.3, 147.0, 145.6, 129.9, 123.9, 69.6, 64.5, 60.9, 35.0, 25.7, 21.0, 17.9, -4.2, -5.2; **HRMS (ESI)**: calculated for $C_{18}H_{31}N_2O_5Si^+$: m/z = 383.2002 $[M+H]^+$; found: m/z = 383.1997 $[M+H]^+$.

The analytical data is in agreement with the literature.[1]

Compound 8

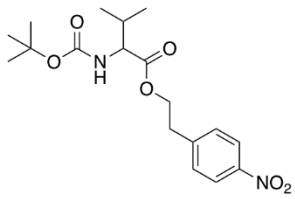


Yield: 96%; **IR**: $\tilde{\nu}$ = 3854 (w), 3745 (w), 2955 (w), 2930 (w), 2856 (w), 1735 (vs), 1601 (s), 1518 (vs), 1472 (w), 1374 (w), 1344 (vs), 1251 (s), 1155 (s), 1075 (s), 967 (s), 855 (w), 835 (s), 775 (s), 747 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.13 (d, *J* = 8.6 Hz, 2H), 7.36 (d, *J* = 8.6 Hz, 2H), 4.34 (q, *J* = 6.7 Hz, 2H), 3.82 (dd, *J* = 9.8, 3.7 Hz, 1H), 3.73 (dd, *J* = 9.8, 3.0 Hz, 1H), 3.47 (s, 1H), 3.04 (t, *J* = 6.7 Hz, 2H), 0.81 (s, 9H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 173.9, 146.9, 145.6, 129.8, 123.8, 65.4, 64.4, 56.5, 34.9, 25.7, 18.2, -5.5, -5.6; **HRMS (ESI)**: calculated for $C_{17}H_{29}N_2O_5Si^+$: m/z = 369.1846 $[M+H]^+$; found: m/z = 369.1842 $[M+H]^+$.

General procedure for npe-ester formation of Boc-protected amino acid

Boc-amino acid (1 eq.) was dissolved in CH_2Cl_2 under inert atmosphere and cooled to 0 °C. Then npeOH (1.3 eq.) and PPh_3 (1.3 eq.) were added followed by slow addition of DIAD (1.3 eq.). The reaction mixture was left to stir for 2 h at room temperature. Then the solution was washed with water, organic phase was dried over Na_2SO_4 and evaporated. The crude product was purified by flash chromatography to afford the target product.

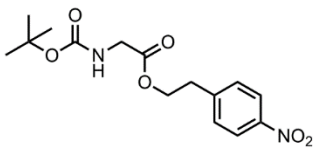
Compound 12



Eluent: Hex/EtOAc (4/1, v/v).

Yield: 96%; **IR**: $\tilde{\nu}$ = 3397 (w), 2975 (w), 1751 (s), 1709 (vs), 1519 (vs), 1391 (w), 1366 (w), 1345 (vs), 1159 (vs), 1056 (w), 905 (vs), 723 (vs) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.15 (d, *J* = 8.6 Hz, 2H), 7.38 (d, *J* = 8.6 Hz, 2H), 4.95 (d, *J* = 8.7 Hz, 1H), 4.38 (t, *J* = 6.6 Hz, 2H), 4.15 (dd, *J* = 9.0, 4.8 Hz, 1H), 3.06 (t, *J* = 6.6 Hz, 2H), 2.05 – 1.97 (m, 1H), 1.41 (s, 9H), 0.88 (d, *J* = 6.8 Hz, 3H), 0.78 (d, *J* = 6.8 Hz, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 172.4, 155.7, 146.9, 145.5, 129.8, 123.8, 79.9, 64.5, 58.6, 34.9, 31.2, 28.4, 19.0, 17.6; **HRMS (ESI)**: calculated for $C_{18}H_{27}N_2O_6^+$: m/z = 367.1869 $[M+H]^+$; found: m/z = 367.1874 $[M+H]^+$.

Compound 13

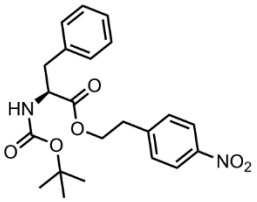


Eluent: Hex/EtOAc (3/1, v/v).

Yield: 85%; **IR**: $\tilde{\nu}$ = 3396 (w), 2978 (w), 2254 (w), 1750 (s), 1707 (vs), 1601 (w), 1518 (vs), 1391 (w), 1366 (w), 1345 (vs), 1250 (w), 1159 (vs), 1056 (w), 905 (vs), 727 (vs) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.10 (d, *J* = 8.6 Hz, 2H), 7.38 (d, *J* = 8.6 Hz, 2H), 4.97 (d, *J* = 9.2 Hz, 1H), 4.39 (t, *J* = 6.6 Hz, 2H), 3.87 (d, *J* = 5.8 Hz, 2H), 3.06 (t, *J* = 6.6 Hz, 2H), 1.43 (s, 9H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 170.3, 155.8, 147.0, 145.4, 129.9, 123.9, 80.3, 64.7, 42.4, 34.9, 28.4; **HRMS (ESI)**: calculated for $C_{15}H_{21}N_2O_6^+$: m/z = 325.1400 $[M+H]^+$; found: m/z = 325.1398 $[M+H]^+$.

The analytical data is in agreement with the literature.[2]

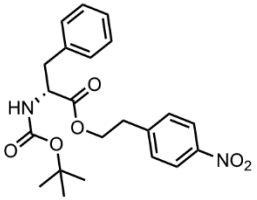
Compound 14a



Eluent: Hex/EtOAc (4/1, v/v).

Yield: 81%; **IR**: $\tilde{\nu}$ = 3426 (w), 3356 (w), 1740 (w), 1709 (vs), 1602 (w), 1518 (vs), 1495 (s), 1344 (vs), 1249 (w), 1159 (vs), 1056 (w), 855 (s), 733 (s), 698 (vs) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.15 (d, *J* = 8.7 Hz, 2H), 7.31 (d, *J* = 8.7 Hz, 2H), 7.25 – 7.20 (m, 3H), 7.04 (d, *J* = 5.6 Hz, 2H), 4.90 (d, *J* = 8.4 Hz, 1H), 4.54 (q, *J* = 6.4 Hz, 1H), 4.39 – 4.25 (m, 2H), 3.07 – 2.93 (m, 4H), 1.41 (s, 9H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.8, 155.0, 146.9, 145.3, 135.8, 129.8, 129.2, 128.6, 127.1, 123.8, 80.1, 64.7, 54.5, 38.4, 34.7, 28.3; **HRMS (ESI)**: calculated for $C_{22}H_{26}N_2O_6Na^+$: m/z = 437.1683 $[M+Na]^+$; found: m/z = 437.1684 $[M+Na]^+$.

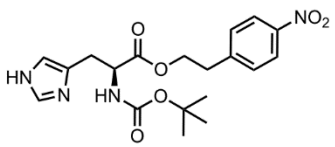
Compound 14b



Eluent: Hex/EtOAc (4/1, v/v).

Yield: 85%; **IR**: $\tilde{\nu}$ = 3443 (w), 3375 (w), 2977 (w), 2929 (w), 1740 (w), 1709 (vs), 1602 (w), 1517 (vs), 1495 (s), 1343 (vs), 1248 (w), 1157 (vs), 1055 (w), 855 (s), 747 (s), 698 (vs) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.16 (d, *J* = 8.7 Hz, 2H), 7.33 (d, *J* = 8.7 Hz, 2H), 7.30 – 7.20 (m, 3H), 7.07 (d, *J* = 6.2 Hz, 2H), 4.97 (d, *J* = 8.4 Hz, 1H), 4.56 (q, *J* = 6.2 Hz, 1H), 4.40 – 4.27 (m, 2H), 3.08 – 2.93 (m, 4H), 1.43 (s, 9H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 171.8, 155.1, 146.9, 145.4, 135.9, 129.8, 129.2, 128.6, 127.1, 123.8, 80.0, 64.7, 54.5, 38.4, 34.7, 28.3; **HRMS (ESI)**: calculated for $C_{22}H_{30}N_3O_6^+$: m/z = 432.2129 $[M+NH_4]^+$; found: m/z = 432.2131 $[M+NH_4]^+$.

Compound 19



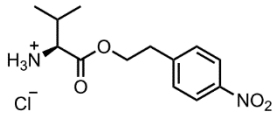
Boc-histidine (0.5 g, 1.96 mmol, 1 eq.), 2-(4-nitrophenyl)ethanol (0.655 g, 3.92 mmol, 2 eq.), DMAP (0.048 g, 0.39 mmol, 0.20 eq.) and HBTU (0.967 g, 2.55 mmol, 1.3 eq.) were dissolved in DMF (4 ml) under inert atmosphere. Diisopropylamine (686 µl, 4.90 mmol, 2.5 eq.) was added dropwise and the reaction mixture was stirred overnight at room temperature. The resulting solution was diluted with EtOAc (30 ml) and quenched with saturated NH_4Cl solution (15 ml). The organic layer was washed with water, dried and the solvents were removed *in vacuo*. The crude product was purified by flash chromatography on silica gel (2% to 5% MeOH in DCM) to obtain the target product as a pale-yellow foam.

Yield: 93%; **IR**: $\tilde{\nu}$ =2977 (w), 1699 (vs), 1600 (s), 1516 (vs), 1391 (w), 1365 (w), 1344 (vs), 1250 (w), 1160 (vs), 1108 (w), 1054 (w), 1016 (w), 855 (vs), 748 (w), 697 (w) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.18 – 8.07 (m, 2H), 7.58 (s, 1H), 7.41 – 7.30 (m, 2H), 6.72 (s, 1H), 5.77 (d, *J* = 8.2 Hz, 1H), 4.51 (q, *J* = 6.2 Hz, 1H), 4.34 (t, *J* = 6.6 Hz, 2H), 3.06 – 3.00 (m, 4H), 1.41 (s, 9H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 172.1, 155.7, 147.1, 145.7, 135.2, 134.2, 130.0, 123.9, 115.9, 80.2, 64.9, 53.6, 34.9, 29.7, 28.5; **HRMS (ESI)**: calculated for $C_{19}H_{25}N_4O_6^+$: m/z = 405.1769 $[M+H]^+$; found: m/z = 405.1765 $[M+H]^+$.

General procedure for deprotection of Boc-protecting group

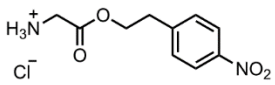
Npe-protected amino acid was dissolved in 4M HCl/Dioxane mixture at 0 °C. The reaction mixture was left to stir for 2 h and afterwards was evaporated to dryness. The resulting product was used for further steps without additional purification.

Compound 15



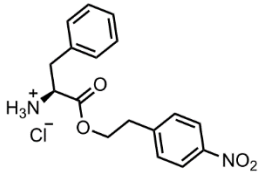
Yield: 99%; **IR**: $\tilde{\nu}$ = 3335 (w), 2964 (w), 2850 (w), 1741 (s), 1604 (w), 1516 (vs), 1464 (w), 1379 (vs), 1232 (vs), 1215 (s), 1170 (w), 1043 (w), 969 (w), 857 (s), 751 (s), 700 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.66 (br. s, 3H), 8.17 (d, *J* = 8.6 Hz, 2H), 7.62 (d, *J* = 8.6 Hz, 2H), 4.46 (dtd, *J* = 23.3, 11.1, 6.3 Hz, 2H), 3.76 (d, *J* = 4.5 Hz, 1H), 3.11 (t, *J* = 6.3 Hz, 2H), 2.10 (tt, *J* = 11.6, 5.8 Hz, 1H), 0.84 (d, *J* = 3.0 Hz, 3H), 0.82 (d, *J* = 3.0 Hz, 3H). **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.8, 146.3, 130.4, 123.4, 65.3, 57.2, 33.8, 29.2, 18.3, 17.4 **HRMS (ESI)**: calculated for $C_{13}H_{19}N_2O_4^+$: m/z = 267.1339 $[M+H]^+$; found: m/z = 267.1139 $[M+H]^+$.

Compound 16



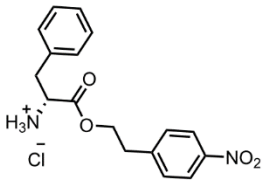
Yield: 99%; **IR**: $\tilde{\nu}$ = 2949 (w), 1746 (s), 1515 (vs), 1310 (vs), 1238 (vs), 1053 (w), 955 (s), 905 (s) 856 (s), 698 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.30 (br. s, 3H), 8.18 (d, *J* = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 4.44 (t, *J* = 6.4 Hz, 2H), 3.77 (s, 2H), 3.09 (t, *J* = 6.4 Hz, 2H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 167.6, 156.2, 146.3, 130.4, 123.5, 67.9, 33.9, 28.2, 22.0; **HRMS (ESI)**: calculated for $C_{10}H_{13}N_2O_4^+$: m/z = 225.0870 $[M+H]^+$; found: m/z = 225.0868 $[M+H]^+$.
The analytical data is in agreement with the literature.[2]

Compound 17a



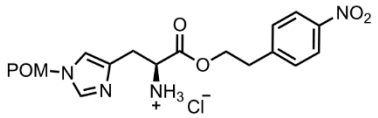
Yield: 99%; **IR**: $\tilde{\nu}$ = 3142 (w), 2988 (w), 2802 (w), 1740 (vs), 1601 (s), 1518 (vs), 1490 (vs), 1351 (vs), 1232 (vs), 1191 (s), 1102 (s), 981 (s), 856 (vs), 755 (vs), 736 (vs), 706 (vs) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.74 (br. s, 3H), 8.16 (d, *J* = 8.7 Hz, 2H), 7.49 (d, *J* = 8.7 Hz, 2H), 7.33 – 7.21 (m, 3H), 7.12 (dd, *J* = 7.8, 1.8 Hz, 2H), 4.33 (t, *J* = 6.3 Hz, 2H), 4.20 (dd, *J* = 7.8, 5.5 Hz, 1H), 3.16 (dd, *J* = 14.0, 5.5 Hz, 1H), 3.06 – 2.89 (m, 3H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.9, 146.3, 134.9, 130.4, 129.4, 128.5, 127.2, 123.4, 65.4, 53.3, 35.8, 33.7; **HRMS (ESI)**: calculated for $C_{17}H_{19}N_2O_4^+$: m/z = 315.1339 $[M+H]^+$; found: m/z = 315.1332 $[M+H]^+$.

Compound 17b



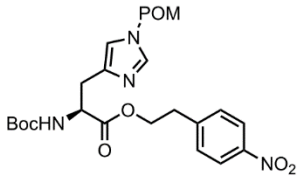
Yield: 99%; **IR**: $\tilde{\nu}$ = 3146 (w), 2988 (w), 2802 (w), 1740 (vs), 1602 (s), 1518 (vs), 1490 (vs), 1351 (vs), 1232 (vs), 1192 (s), 1102 (s), 856 (vs), 755 (vs), 736 (vs), 705 (vs) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ**: 8.91 (br. s., 3H), 8.12 (d, *J* = 8.7 Hz, 2H), 7.46 (d, *J* = 8.7 Hz, 2H), 7.27 – 7.20 (m, 3H), 7.12 (d, *J* = 6.4 Hz, 2H), 4.29 (t, *J* = 6.4 Hz, 2H), 4.19 – 4.08 (m, 1H), 3.21 (dd, *J* = 14.1, 5.0, 1H), 3.10 – 2.88 (m, 3H); **^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ**: 168.9, 146.3, 134.8, 130.4, 129.4, 128.5, 127.2, 123.4, 65.4, 53.3, 35.8, 33.7; **HRMS (ESI)**: calculated for $C_{17}H_{19}N_2O_4^+$: m/z = 315.1339 $[M+H]^+$; found: m/z = 315.1332 $[M+H]^+$.

Compound 21



Yield: 98%; **IR**: $\tilde{\nu}$ = 2960 (w), 1737 (vs), 1624 (w), 1598 (w), 1516 (vs), 1454 (w), 1415 (w), 1351 (vs), 1282 (w), 1188 (w), 1124 (vs), 1045 (w), 1008 (w), 854 (s), 774 (w), 749 (w), 701 (w) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ**: 8.69 (s, 3H), 8.21 – 8.15 (m, 2H), 7.59 – 7.55 (m, 2H), 7.51 (s, 1H), 6.04 (s, 2H), 4.39 (m, 3H), 3.56 (s, 1H), 3.17 (d, *J* = 6.9 Hz, 2H), 3.04 (td, *J* = 6.4, 3.9 Hz, 2H), 1.13 (s, 9H); **^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ**: 176.7, 168.2, 146.4, 146.1, 137.4, 130.4, 123.5, 120.0, 69.2, 66.4, 65.6, 51.1, 38.2, 33.7, 26.5; **HRMS (ESI)**: calculated for $C_{20}H_{27}N_4O_6^+$: m/z = 419.1925 $[M+H]^+$; found: 419.1919 $[M+H]^+$.

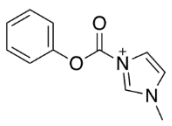
Compound 20



The histidine derivative **19** (0.745 g, 1.84 mmol, 1 eq.) was dissolved in DMF (12 mL) under N_2 atmosphere at 0 °C. Subsequently, K_2CO_3 (0.509 g, 3.68 mmol, 2 eq.) was added and the mixture was stirred for 40 min. Chloromethylpivalate (319 µl, 2.21 mmol, 1.2 eq.) was added dropwise at 0°C and the reaction mixture was left to warm to room temperature while stirring for 5 h. Catalytic amounts of KI were added and the mixture was stirred for another 1 h. The resulting suspension was diluted with EtOAc (75 ml) and quenched with saturated NH_4Cl solution (35 ml). The organic phase was washed with water, dried over Na_2SO_4 and concentrated *in vacuo*. The crude mixture was purified by flash chromatography on silica gel (40% isohexane in EtOAc to pure EtOAc). The pivalate protected histidine derivative was obtained as a yellow oil.

Yield: 55%; **IR**: 2850 (w), 2600 (w), 1738 (vs), 1624 (w), 1598 (w), 1573 (w), 1516 (vs), 1454 (w), 1414 (w), 1350 (vs), 1282 (w), 1191 (w), 1124 (vs), 1044 (w), 1008 (w), 854 (vs), 773 (w), 749 (w), 701 (w) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.19 – 8.14 (m, 2H), 7.54 (d, *J* = 1.4 Hz, 1H), 7.42 – 7.36 (m, 2H), 6.71 (s, 1H), 5.81 (d, *J* = 8.2 Hz, 1H), 5.73 (s, 2H), 4.51 (dt, *J* = 8.2, 5.3 Hz, 1H), 4.34 (t, *J* = 6.7 Hz, 2H), 3.04 (t, *J* = 6.7 Hz, 2H), 2.97 (t, *J* = 5.2 Hz, 2H), 1.42 (s, 9H), 1.14 (s, 9H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 177.9, 172.0, 155.7, 147.0, 145.8, 138.3, 130.0, 123.9, 117.3, 79.9, 77.4, 67.7, 64.7, 53.5, 38.9, 35.0, 30.1, 28.5, 27.0; **HRMS (ESI)**: calculated for $C_{25}H_{35}N_4O_8^+$: m/z = 519.2449 $[M+H]^+$; found: m/z = 519.2441 $[M+H]^+$.

Compound 22



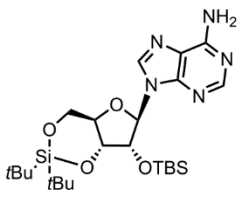
The reaction was conducted according to a published procedure.[3]

Phenyl chloroformate (4 ml, 31.9 mmol, 1 eq.) was dissolved in dry CH_2Cl_2 under nitrogen and cooled to 0 °C. Then *N*-methylimidazole (2.54 ml, 31.9 mmol, 1 eq.) was added dropwise. The mixture was allowed to stir at room temperature for 2 hours. Afterwards the reaction mixture was filtered, the precipitate was washed with CH_2Cl_2 and dried.

Yield: 95%; **IR**: $\tilde{\nu}$ = 2926 (w), 1783 (vs), 1588 (w), 1536 (w), 1372 (s), 1330 (s), 1232 (vs), 749 (vs), 689 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ**: 10.29 (s, 1H), 8.37 (s, 1H), 8.02 (s, 1H), 7.43 – 7.58 (m, 5H), 4.01 (s, 3H); **^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ**: 157.5, 135.6, 129.3, 123.1, 121.3, 119.5, 118.6, 115.3, 35.4.

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

Compound 24



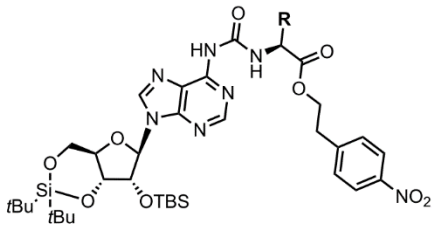
Compound was synthesized following the procedure published earlier.[4]

Adenosine **23** (1 g, 3.74 mmol, 1 eq.) was suspended in DMF and di-*tert*-butylsilyl ditriflate (1.46 ml, 4.49 mmol, 1.2 eq.) was added dropwise under stirring at 0 °C. The resulting solution was stirred at 0°C for 45 min. Then imidazole (1.27 g, 18.7 mmol, 5 eq.) was added and the reaction was warmed to room temperature over a period of 30 min. Then TBSCl (0.68 g, 4.49 mmol, 1.2 eq.) was added and the reaction was heated to 60 °C overnight. Subsequently, the reaction mixture was diluted with EtOAc and washed with water and brine. The organic layer was dried and evaporated. The residue was purified by flash chromatography (Hex/EtOAc, 1/1, v/v).

Yield: 76%; **IR**: $\tilde{\nu}$ = 3148 (w), 2933 (w), 2859 (w), 2361 (w), 1677 (s), 1604 (s), 1598 (w), 1576 (w), 1473 (w), 1426 (w), 1363 (w), 1329 (w), 1302 (w), 1258 (w), 1200 (w), 1166 (w), 1136 (w), 1105 (w), 1064 (vs), 1009 (s), 890 (w), 828 (vs), 786 (w), 754 (w), 729 (w) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.31 (s, 1H), 7.83 (s, 1H), 6.12 (br. s, 2H), 5.91 (s, 1H), 4.61 (d, *J* = 4.7 Hz, 1H), 4.50 (ddd, *J* = 16.5, 9.3, 4.7 Hz, 2H), 4.25 – 4.17 (m, 1H), 4.03 (dd, *J* = 10.5, 9.3 Hz, 1H), 1.07 (s, 9H), 1.04 (s, 9H), 0.92 (s, 9H), 0.16 (s, 3H), 0.14 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 155.5, 152.8, 149.3, 138.9, 120.4, 92.6, 75.9, 75.6, 74.8, 67.9, 27.6, 27.1, 26.0, 22.9, 20.5, 18.4, -4.2, -4.8; **HRMS (ESI)**: calculated for $C_{24}H_{44}N_5O_4Si_2^+$: m/z = 522.2932 $[M+H]^+$; found: m/z = 522.2926 $[M+H]^+$.

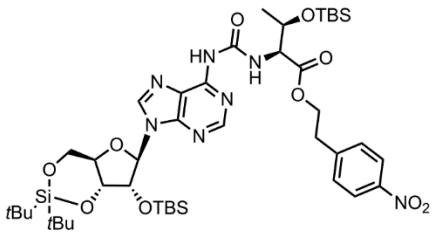
The analytical data is in agreement with the literature.[4]

General procedure for amino acid attachment to the adenosine derivative



The silyl-protected adenosine derivative **24** (1 eq.) was dissolved in dry CH_2Cl_2 under nitrogen atmosphere. 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**22**, 2 eq.) was added to the reaction mixture and the resulting suspension was stirred at room temperature for 2 hours (the solution in time becomes clear). Afterwards the protected amino acid (2 eq.) was added together with TEA (2 eq.) as a solution in CH_2Cl_2 and the resulting solution was stirred overnight at room temperature. The reaction was quenched by addition of saturated aqueous $NaHCO_3$ solution. The solution was extracted three times with CH_2Cl_2, and the organic phase was dried, filtered and concentrated *in vacuo*. The residue was purified by flash chromatography eluting with Hex/EtOAc to give product as white foam.

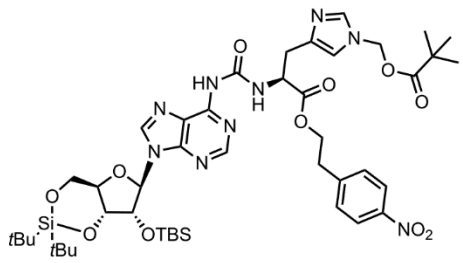
Compound 25



Eluent: Hex/EtOAc (4/3, v/v).

Yield: 91%; **IR**: $\tilde{\nu}$ = 3237 (w), 2931 (s), 2857 (s), 1737 (s), 1701 (vs), 1610 (s), 1520 (vs), 1465 (s), 1345 (s), 1250 (s), 1136 (w), 1057 (s), 998 (w), 894 (w), 840 (s), 777 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 10.05 (d, *J* = 9.0 Hz, 1H), 8.56 (s, 1H), 8.41 (s, 1H), 8.16 (s, 1H), 7.94 (d, *J* = 8.6 Hz, 2H), 7.32 (d, *J* = 8.6 Hz, 2H), 5.99 (s, 1H), 4.64 (d, *J* = 4.6 Hz, 1H), 4.60-4.46 (m, 3H), 4.33-4.24 (m, 2H), 4.20-4.29 (m, 1H), 4.05 (dd, *J* = 10.5, 9.1 Hz, 1H), 3.03 (t, *J* = 6.5 Hz, 2H), 1.25 (d, *J* = 6.5 Hz, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.90 (s, 9H), 0.19 (s, 3H), 0.16 (s, 3H), 0.07 (s, 3H), -0.04 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 171.1, 154.6, 151.3, 150.3, 149.8, 146.8, 145.7, 141.6, 129.9, 123.7, 121.1, 92.6, 76.0, 75.7, 75.0, 68.8, 68.0, 64.8, 59.8, 35.0, 27.7, 27.2, 26.1, 25.7, 22.9, 21.3, 20.6, 18.5, 18.0, -4.1, -4.8, -5.2; **HRMS (ESI)**: calculated for $C_{43}H_{72}N_7O_{10}Si_3^+$: m/z = 930.4643 $[M+H]^+$; found: m/z = 930.4640 $[M+H]^+$.

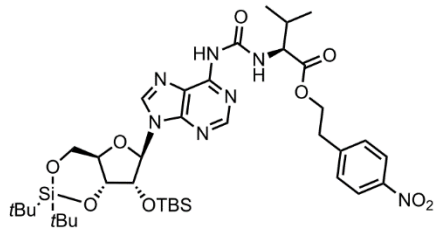
Compound 26



Eluent: 10% CH_2Cl_2 in EtOAc to pure EtOAc.

Yield: 86%; **IR**: $\tilde{\nu}$ = 3854 (w), 3745 (w), 3650 (w), 2932 (w), 2858 (w), 2361 (w), 2341 (w), 1735 (s), 1670 (s), 1654 (w), 1610 (w), 1587 (w), 1521 (vs), 1472 (s), 1395 (w), 1345 (vs), 1252 (w), 1166 (w), 1118 (vs), 1055 (vs), 999 (w), 894 (w), 826 (vs), 781 (s), 750 (w) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 10.05 (dd, *J* = 7.6, 3.7 Hz, 1H), 8.45 (s, 1H), 8.08 (s, 1H), 8.06 – 8.02 (m, 3H), 7.59 (d, *J* = 1.3 Hz, 1H), 7.41 – 7.35 (m, 2H), 6.86 (dd, *J* = 5.3, 1.3 Hz, 1H), 5.96 (d, *J* = 7.0 Hz, 1H), 5.74 (d, *J* = 2.6 Hz, 2H), 4.94 – 4.85 (m, 1H), 4.60 (dd, *J* = 6.0, 4.6 Hz, 1H), 4.53 – 4.46 (m, 2H), 4.42 (tq, *J* = 6.5, 1.7 Hz, 2H), 4.24 (tdd, *J* = 9.8, 5.0, 2.8 Hz, 1H), 4.08 – 4.00 (m, 1H), 3.23 – 3.12 (m, 2H), 3.07 (t, *J* = 6.5 Hz, 2H), 1.09 (s, 9H), 1.08 (s, 9H), 1.05 (s, 9H), 0.94 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 177.8, 171.7, 153.6, 151.3, 150.2, 149.8, 146.9, 145.8, 141.2, 138.4, 138.1, 123.0, 123.8, 121.1, 117.3, 92.6, 76.0, 75.7, 74.9, 67.9, 67.7, 64.7, 55.5, 38.8, 35.0, 30.7, 27.7, 27.2, 26.9, 26.1, 22.9, 20.5, 18.5, 1.3, -4.2; **HRMS (ESI)**: calculated for $C_{45}H_{68}N_9O_{11}Si_2^+$: m/z = 966.4571 $[M+H]^+$; found: m/z = 966.4576 $[M+H]^+$.

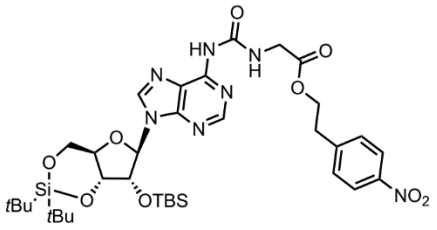
Compound 27



Eluent: Hex/EtOAc (4/3, v/v).

Yield: 78%; **IR**: $\tilde{\nu}$ = 3230 (w), 2960 (w), 2960 (w), 2858 (w), 1741 (s), 1702 (vs), 1611 (s), 1520 (vs), 1466 (s), 1345 (vs), 1250 (s), 1139 (vs), 1057 (vs), 999 (s), 894 (s), 810 (vs), 781 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 10.01 (d, *J* = 8.4 Hz, 1H), 8.50 (s, 1H), 8.17 (s, 1H), 8.07 (d, *J* = 8.3 Hz, 2H), 7.28 (d, *J* = 8.3 Hz, 2H), 5.98 (s, 1H), 4.61 (d, *J* = 4.6 Hz, 1H), 4.56 – 4.38 (m, 5H), 4.20-4.29 (m, 1H), 4.07 (dd, *J* = 10.5, 9.1 Hz, 1H), 3.09 (t, *J* = 6.5 Hz, 2H), 2.28 – 2.21 (m, 1H), 1.08 (s, 9H), 1.05 (s, 9H), 1.00 (d, *J* = 6.5 Hz, 3H), 0.95 (s, 12H), 0.18 (s, 3H), 0.16 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 171.1, 154.1, 152.2, 151.1, 149.8, 146.9, 145.6, 141.5, 129.9, 126.4, 123.7, 121.1, 92.6, 75.9, 75.7, 74.9, 67.9, 64.8, 58.8, 35.0, 30.9, 27.7, 27.2, 26.1, 22.9, 20.4, 19.5, 18.4, 18.0, -4.2, -4.9; **HRMS (ESI)**: calculated for $C_{38}H_{60}N_7O_9Si_2^+$: m/z = 814.3991 $[M+H]^+$; found: m/z = 814.3976.

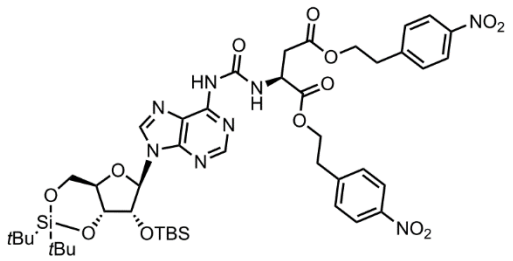
Compound 28



Eluent: Hex/EtOAc (4/3, v/v).

Yield: 72%; **IR**: $\tilde{\nu}$ = 3239 (w), 2932 (s), 2858 (s), 1749 (s), 1703 (vs), 1611 (s), 1520 (vs), 1468 (s), 1345 (s), 1252 (s), 1141 (w), 1055 (s), 990 (w), 894 (w), 826 (s), 750 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 9.95 (br. s, 1H), 8.50 (s, 1H), 8.21 – 8.03 (m, 3H), 7.38 (d, *J* = 8.6 Hz, 2H), 5.98 (s, 1H), 4.60 (d, *J* = 4.7 Hz, 1H), 4.57 – 4.40 (m, 4H), 4.30 – 4.17 (m, 3H), 4.10 – 4.03 (m, 1H), 3.09 (t, *J* = 6.6 Hz, 2H), 1.08 (s, 9H), 1.05 (s, 9H) 0.94 (s, 9H), 0.17 (s, 3H), 0.16 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 170.0, 154.2, 151.2, 150.2, 149.9, 147.0, 145.5, 141.4, 129.9, 123.8, 121.1, 92.5, 76.0, 75.7, 74.9, 67.9, 42.2, 35.0, 27.6, 27.2, 26.0, 22.9, 20.5, 18.5, -4.1, -4.9; **HRMS (ESI)**: calculated for $C_{35}H_{54}N_7O_9Si_2^+$: m/z = 772.3522 $[M+H]^+$; found: m/z = 772.3504 $[M+H]^+$.

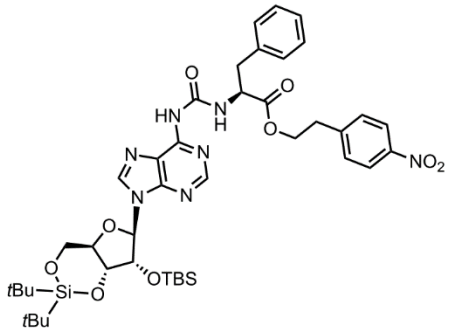
Compound 29



Eluent: Hex/EtOAc (2/1, v/v).

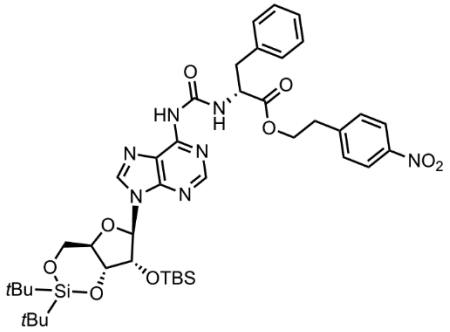
Yield: 87%; **IR**: $\tilde{\nu}$ = 3229 (w), 2953 (w), 2929 (w), 2857 (w), 1735 (s), 1693 (s), 1607 (s), 1589 (s), 1517 (vs), 1469 (s), 1391 (w), 1310 (vs), 1292 (w), 1251 (w), 1210 (w), 835 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 10.28 (s, 1H), 8.42 (s, 1H), 8.22 – 8.12 (m, 1H), 8.07 (d, *J* = 8.6 Hz, 2H), 7.96 (d, *J* = 8.6 Hz, 2H), 7.37 – 7.30 (m, 4H), 6.00 (s, 1H), 4.93 – 4.89 (m, 1H), 4.64 (d, *J* = 4.6 Hz, 1H), 4.51 (ddd, *J* = 9.2, 4.9, 2.7 Hz, 2H), 4.46 – 4.40 (m, 2H), 4.39 – 4.22 (m, 3H), 4.09 – 4.05 (m, 1H), 3.08 – 2.94 (m, 6H), 1.07 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.19 (s, 3H), 0.17 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 179.8, 170.7, 153.7, 149.8, 146.9, 146.8, 145.5, 145.4, 129.8, 123.8, 123.6, 120.9, 92.6, 75.9, 75.7, 74.9, 67.9, 65.0, 64.5, 49.7, 36.5, 34.9, 34.8, 27.6, 27.1, 26.0, 22.8, 20.5, 18.4, -4.2, -4.9; **HRMS (ESI)**: calculated for $C_{45}H_{63}N_8O_{13}Si_2^+$: m/z = 979.4053 $[M+H]^+$; found: m/z = 979.4056.

Compound 30a



Yield: 68%; **IR**: $\tilde{\nu}$ = 3190 (w), 2933 (w), 2858 (w), 1742 (w), 1702 (vs), 1612 (s), 1587 (w), 1521 (vs), 1469 (vs), 1345 (vs), 1253 (s), 1057 (s), 1000 (w), 828 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CD_2Cl_2) δ**: 10.00 (d, *J* = 7.5 Hz, 1H), 8.94 (s, 1H), 8.37 (d, *J* = 1.2 Hz, 2H), 8.00 (d, *J* = 8.7 Hz, 2H), 7.39 – 7.26 (m, 5H), 7.21 – 7.17 (m, 2H), 6.02 (s, 1H), 4.89 – 4.76 (m, 1H), 4.60 (d, *J* = 4.6 Hz, 1H), 4.54 – 4.46 (m, 2H), 4.40 (td, *J* = 6.5 Hz, 1.7 Hz, 2H), 4.31 – 4.22 (m, 1H), 4.16 – 4.06 (m, 1H), 3.18 (d, *J* = 6.8 Hz, 2H), 3.03 (t, *J* = 6.5 Hz, 2H), 1.10 (s, 9H), 1.06 (s, 9H), 0.96 (s, 9H), 0.20 (s, 3H), 0.18 (s, 3H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.5, 153.8, 150.7, 150.2, 149.8, 146.7, 145.9, 142.2, 136.4, 129.8, 129.5, 128.5, 127.1, 123.4, 120.8, 92.2, 75.8, 75.7, 74.8, 67.7, 64.6, 54.9, 37.8, 34.7, 27.3, 26.9, 25.7, 22.6, 20.2, 18.2, -4.6, -5.2; **HRMS (ESI)**: calculated for $C_{42}H_{60}N_7O_9Si_2^+$: m/z = 862.3986 $[M+H]^+$; found: m/z = 862.3995 $[M+H]^+$.

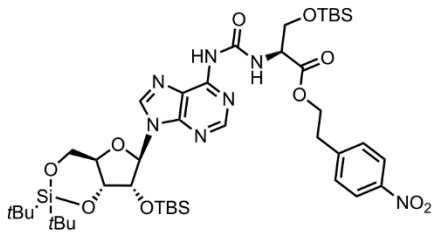
Compound 30b



Eluent: Hex/EtOAc (4/1, v/v).

Yield: 71%; **IR**: $\tilde{\nu}$ = 3192 (w), 2933 (w), 2858 (w), 1740 (w), 1700 (vs), 1611 (s), 1520 (vs), 1466 (vs), 1345 (vs), 1252 (s), 1166 (w), 1139 (w), 1054 (vs), 998 (s), 893 (s), 826 (vs), 736 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CD_2Cl_2) δ**: 10.01 (d, *J* = 7.6 Hz, 1H), 8.96 (s, 1H), 8.37 (s, 2H), 8.00 (d, *J* = 8.7 Hz, 1H), 7.37 – 7.22 (m, 5H), 7.22 – 7.14 (m, 2H), 6.03 (s, 1H), 4.90 – 4.81 (m, 1H), 4.61 (d, *J* = 4.5 Hz, 1H), 4.54 – 4.45 (m, 2H), 4.39 (t, *J* = 6.4 Hz, 2H), 4.32 – 4.21 (m, 1H), 4.16 – 4.05 (m, 1H), 3.19 (d, *J* = 6.2 Hz, 2H), 3.02 (t, *J* = 6.4 Hz, 2H), 1.09 (s, 9H), 1.06 (s, 9H), 0.96 (s, 9H), 0.20 (s, 3H), 0.18 (s, 3H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.4, 153.6, 150.8, 150.2, 149.8, 146.7, 145.9, 141.8, 136.4, 129.8, 129.5, 128.5, 127.1, 123.4, 120.8, 92.2, 75.9, 75.7, 74.8, 67.7, 64.6, 54.9, 37.9, 34.7, 27.3, 26.8, 25.7, 22.6, 20.2, 18.2, -4.6, -5.3.; **HRMS (ESI)**: calculated for $C_{42}H_{60}N_7O_9Si_2^+$: m/z = 862.3986 $[M+H]^+$; found: m/z = 862.3996 $[M+H]^+$.

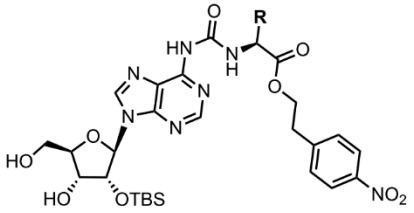
Compound 31



Eluent: Hex/EtOAc (1/1, v/v).

Yield: 85%; **IR**: $\tilde{\nu}$ = 3238 (w), 2931 (s), 2859 (s), 1737 (s), 1701 (vs), 1610 (s), 1520 (vs), 1470 (s), 1345 (s), 1251 (s), 1136 (w), 1057 (s), 998 (w), 899 (w), 840 (s), 778 (s) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 10.18 (d, *J* = 8.3 Hz, 1H), 8.61 (s, 1H), 8.44 (s, 1H), 8.19 (s, 1H), 8.02 (d, *J* = 8.6 Hz, 2H), 7.35 (d, *J* = 8.7 Hz, 2H), 5.99 (s, 1H), 4.72 (dt, *J* = 8.4, 2.8 Hz, 1H), 4.61 (d, *J* = 4.6 Hz, 1H), 4.50 (td, *J* = 9.8, 4.9 Hz, 2H), 4.43 (t, *J* = 6.5 Hz, 2H), 4.24 (td, *J* = 10.0, 5.0 Hz, 1H), 4.14 (dd, *J* = 10.1, 2.7 Hz, 1H), 4.07 (dd, *J* = 10.5, 9.1 Hz, 1H), 3.91 (dd, *J* = 10.1, 3.1 Hz, 1H), 3.06 (t, *J* = 6.5 Hz, 2H), 1.08 (s, 9H), 1.05 (s, 9H), 0.94 (s, 9H), 0.88 (s, 10H), 0.18 (s, 3H), 0.16 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 170.6, 153.9, 151.2, 150.3, 149.8, 146.9, 145.6, 141.6, 129.8, 123.7, 121.1, 92.5, 75.9, 75.7, 74.9, 67.9, 64.8, 64.6, 55.7, 34.9, 27.6, 27.1, 26.0, 25.7, 22.8, 20.5, 18.4, 18.2, -4.2, -4.8, -5.3, -5.6; **HRMS (ESI)**: calculated for $C_{42}H_{70}N_7O_{10}Si_3^+$: m/z = 916.4492 $[M+H]^+$; found: m/z = 916.4501 $[M+H]^+$.

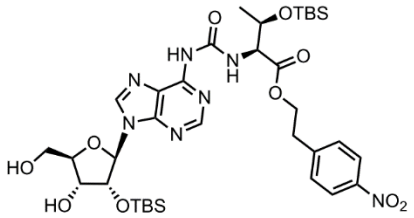
General procedure for deprotection 3'-5'-silyl protecting group:



The modified adenosine (0.86 mmol) was dissolved in CH_2Cl_2 under N_2 atmosphere and transferred into a plastic flask. Pyridine (1 ml) was added and the solution was cooled in an ice-bath. Then Py*$(HF)_n$ (140 µl) was added and the mixture was stirred at 0 °C for 2 h. The reaction was quenched with sat. $NaHCO_3$ and extracted with CH_2Cl_2. Organic phase was washed with

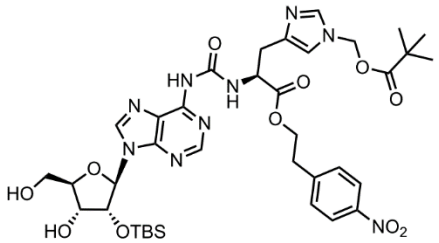
water and dried over Na_2SO_4. The solvents were removed in vacuo. The crude product was purified by flash chromatography eluting with CH_2Cl_2/MeOH (9/1, v/v) to afford the product as a colourless foam.

Compound 32



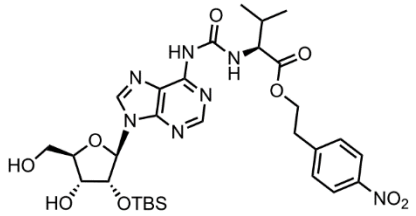
Yield: 95%; **IR**: $\tilde{\nu}$ = 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (vs), 1469 (s), 1345 (vs), 1313 (w), 1250 (vs), 1129 (w), 1093 (s), 835 (vs), 760 (vs) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 9.90 (d, *J* = 9.1 Hz, 1H), 8.50 (s, 1H), 8.46 (s, 1H), 8.10 – 8.08 (m, 3H), 7.37 (d, *J* = 8.6 Hz, 2H), 5.85 (d, *J* = 7.0 Hz, 1H), 5.10 (dd, *J* = 7.0, 4.7 Hz, 1H), 4.59 (dd, *J* = 9.1, 1.5 Hz, 1H), 4.53 – 4.41 (m, 2H), 4.41 – 4.34 (m, 2H), 4.34 – 4.26 (m, 1H), 3.98 (dd, *J* = 13.0, 1.5 Hz, 1H), 3.78 (dd, *J* = 13.0, 1.5 Hz, 1H), 3.07 (t, *J* = 6.7 Hz, 2H), 1.24 (d, *J* = 6.7 Hz, 3H), 0.89 (s, 9H), 0.81 (s, 9H), 0.05 (s, 3H), -0.06 (s, 3H), -0.14 (s, 3H), -0.34 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 170.9, 154.0, 151.3, 150.9, 149.3, 147.1, 145.6, 143.0, 130.0, 123.9, 91.5, 87.7, 74.9, 72.7, 68.8, 65.13, 63.3, 59.8, 35.0, 25.7, 21.3, 18.1, 18.0, -4.0, -5.0, -5.1, -5.2; **HRMS (ESI)**: calculated for $C_{35}H_{56}N_7O_{10}Si_2^+$: m/z = 790.3622 $[M+H]^+$; found: m/z = 790.3612 $[M+H]^+$.

Compound 33



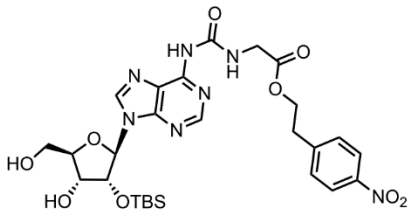
Yield: 75%; **IR**: $\tilde{\nu}$ = 3854 (w), 3745 (w), 3650 (w), 3230 (w), 2930 (w), 2857 (w), 2361 (w), 2341 (w), 1740 (s), 1699 (vs), 1611 (w), 1587 (w), 1520 (vs), 1472 (s), 1395 (w), 1345 (vs), 1253 (w), 1119 (vs), 1091 (w), 1058 (w), 1030 (w), 983 (w), 914 (w), 856 (vs), 779 (vs), 747(w), 670 (w) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 10.05 – 9.98 (m, 1H), 8.50 (d, *J* = 3.1 Hz, 1H), 8.18 – 8.10 (m, 3H), 8.01 (d, *J* = 9.1 Hz, 1H), 7.60 (dd, *J* = 8.2, 1.3 Hz, 1H), 7.44 – 7.37 (m, 2H), 6.84 (d, *J* = 1.4 Hz, 1H), 5.99 – 5.93 (m, 1H), 5.81 (dd, *J* = 7.3, 5.4 Hz, 1H), 5.77 – 5.71 (m, 2H), 5.08 (dt, *J* = 7.3, 5.1 Hz, 1H), 4.90 (dt, *J* = 7.2, 5.4 Hz, 1H), 4.42 (q, *J* = 6.3 Hz, 2H), 4.39 – 4.34 (m, 2H), 3.98 – 3.92 (m, 1H), 3.82 – 3.72 (m, 1H), 3.23 – 3.13 (m, 2H), 3.08 (dt, *J* = 13.8, 6.6 Hz, 2H), 2.83 (s, 1H), 1.09 (s, 9H), 0.80 (s, 9H), -0.18 (s, 3H), -0.40 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 177.8, 171.8, 153.2, 151.1, 150.9, 149.3, 147.0, 145.8, 143.0, 138.2, 138.1, 130.0, 123.9, 123.1, 117.4, 91.4, 87.8, 74.7, 73.0, 67.7, 65.0, 63.5, 53.5, 38.8, 35.0, 30.6, 26.9, 25.7, 18.0, -5.1, -5.2; **HRMS (ESI)**: calculated for $C_{37}H_{52}N_9O_{11}Si^+$: m/z = 826.3550 $[M+H]^+$; found: m/z = 826.3559 $[M+H]^+$.

Compound 34



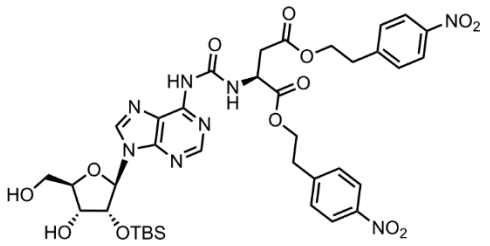
Yield: 95%; **IR**: $\tilde{\nu}$ = 3245 (w), 2930 (w), 2857 (w), 1743 (s), 1695 (vs), 1590 (s), 1518 (vs), 1471 (s), 1344 (vs), 1252 (s), 1212 (w), 1188 (s), 1130 (w), 1085 (s), 836 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.01 (d, *J* = 8.4 Hz, 1H), 8.50 (s, 1H), 8.17 (s, 1H), 8.07 (d, *J* = 8.3 Hz, 2H), 7.28 (d, *J* = 8.3 Hz, 2H), 5.98 (s, 1H), 4.61 (d, *J* = 4.6 Hz, 1H), 4.56 – 4.38 (m, 6H), 4.20-4.29 (m, 1H), 4.07 (dd, *J* = 10.5, 9.1 Hz, 1H), 3.09 (t, *J* = 6.5 Hz, 2H), 1.00 (d, *J* = 6.5 Hz, 3H), 0.96 (d, *J* = 6.5 Hz, 3H), 0.95 (s, 9H), 0.18 (s, 3H), 0.16 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 172.2, 154.4, 151.6, 149.5, 147.0, 145.5, 143.3, 136.5, 129.9, 123.7, 122.0, 91.2, 87.6, 74.7, 72.8, 64.7, 63.4, 58.7, 35.0, 30.9, 25.6, 19.4, 17.9, -5.2, -5.3; **HRMS (ESI)**: calculated for $C_{30}H_{44}N_7O_9Si^+$: m/z = 674.2970 $[M+H]^+$; found: m/z = 674.2974 $[M+H]^+$.

Compound 35



Yield: 97%; **IR**: $\tilde{\nu}$ = 3250 (w), 2932 (w), 2859 (w), 1740 (s), 1680 (vs), 1595 (s), 1520 (vs), 1470 (s), 1340 (vs), 1254 (s), 1213 (w), 1168 (s), 1130 (w), 1085 (s), 836 (vs) cm^{-1}; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 9.92 (t, *J* = 5.7 Hz, 1H), 8.92 (br. s, 1H), 8.54 (s, 1H), 8.25 (s, 1H), 8.13 (d, *J* = 8.7 Hz, 2H), 7.39 (d, *J* = 8.7 Hz, 2H), 5.87 (d, *J* = 7.2 Hz, 1H), 5.07 (dd, *J* = 7.2, *J* = 4.7 Hz, 1H), 4.44 (t, *J* = 6.7 Hz, 2H), 4.37 (m, 2H), 4.20 (t, *J* = 5.2 Hz, 2H), 3.97 (d, *J* = 13.0 Hz, 1H), 3.77 (d, *J* = 13.0 Hz, 1H), 3.09 (t, *J* = 6.7 Hz, 2H), 2.94 (br. s, 1H), 0.78 (s, 9H), -0.19 (s, 3H), -0.38 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$) δ**: 169.9, 154.0, 150.8, 149.5, 147.0, 145.4, 143.6, 129.9, 123.9, 122.0, 91.2, 87.6, 74.8, 72.8, 64.9, 63.3, 42.2, 34.9, 25.6, 17.9, -5.2; **HRMS (ESI)**: calculated for $C_{27}H_{37}N_7O_9Si^+$: m/z = 632.2495 $[M + H]^+$; found m/z = 632.2492.

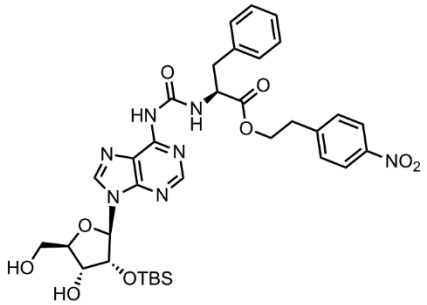
Compound 36



Yield: 98%; **IR**: $\tilde{\nu}$ = 3229 (w), 2953 (w), 2929 (w), 2857 (w), 1735 (s), 1693 (s), 1607 (s), 1589 (s), 1517 (vs), 1469 (s), 1391 (w), 1310 (vs), 1292 (w), 1251 (w), 1210 (w), 835 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.30 (d, *J* = 7.9 Hz, 1H), 9.41 (s, 1H), 8.77 (s, 1H), 8.50 (s, 1H), 8.10 (d, *J* = 8.8 Hz, 2H), 8.06 (d, *J* = 8.8 Hz, 2H), 7.54 (dd, *J* = 8.7, 7.4 Hz, 4H), 6.12 (d, *J* = 5.8 Hz, 1H), 5.07 (s, 1H), 5.00 – 4.95 (m, 1H), 4.90 (dt, *J* = 7.9, 5.3 Hz, 1H), 4.49 – 4.31 (m, 5H), 4.23 (q, *J* = 2.4 Hz, 1H), 4.06 (s, 1H), 3.92 (dd, *J* = 12.4, 2.1 Hz, 1H), 3.79 (d, *J* = 12.2 Hz, 1H), 3.10 (dt, *J* = 12.6, 6.4 Hz, 4H), 3.02 (d, *J* = 5.5 Hz, 2H), 0.79 (s, 9H), -0.08 (s, 3H), -0.21 (s,

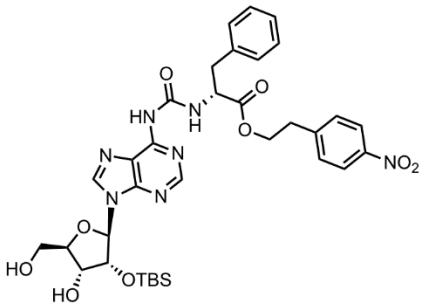
3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 171.3, 171.2, 154.2, 151.4, 150.9, 150.6, 147.6, 147.4, 147.2, 144.3, 136.6, 131.0, 124.1, 121.9, 90.6, 87.6, 76.8, 72.4, 65.8, 65.1, 62.9, 50.6, 37.1, 35.2, 27.2, 26.0, 18.6, -4.9, -5.1; **HRMS (ESI)**: calculated for $C_{37}H_{47}N_8O_{13}Si^+$: m/z = 839.3032 $[M+H]^+$; found: m/z = 839.3041 $[M+H]^+$.

Compound 37a



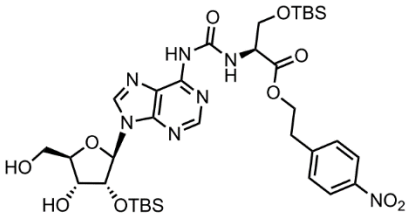
Yield: 80%; **IR**: $\tilde{\nu}$ = 3240 (w), 2929 (w), 2857 (w), 1742 (w), 1699 (vs), 1613 (s), 1520 (vs), 1471 (s), 1345 (vs), 1255 (w), 839 (s) cm^{-1}; **^{1}H NMR (400 MHz, CD_2Cl_2) δ**: 9.83 (d, *J* = 7.4 Hz, 1H), 8.63 (s, 1H), 8.40 (s, 1H), 8.18 (s, 1H), 8.10 (d, *J* = 8.7 Hz, 2H), 7.38 (d, *J* = 8.7 Hz, 2H), 7.33 – 7.21 (m, 2H), 7.19 – 7.14 (m, 2H), 5.87 (d, *J* = 7.2 Hz, 1H), 5.64 (dd, *J* = 11.6 Hz, *J* = 2.3 Hz, 1H), 5.07 (dd, *J* = 7.2 Hz, *J* = 4.7 Hz, 1H), 4.87 – 4.80 (m, 1H), 4.45 – 4.32 (m, 4H), 3.93 (d, *J* = 12.9 Hz, 1H), 3.79 – 3.69 (m, 1H), 3.17 (d, *J* = 6.2 Hz, 2H), 3.05 (t, *J* = 6.6 Hz, 2H), 2.93 (s, 1H), 0.79 (s, 9H), -0.19 (s, 3H), -0.38 (s, 3H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.4, 153.1, 150.7, 150.5, 149.8, 149.3, 146.8, 145.8, 143.3, 136.3, 129.8, 129.4, 128.3, 127.2, 123.5, 91.0, 87.6, 74.8, 72.7, 64.7, 63.1, 54.9, 37.8, 34.7, 25.3, 17.7, -5.6, -5.7; **HRMS (ESI)**: calculated for $C_{34}H_{44}N_7O_9Si^+$: m/z = 722.2965 $[M+H]^+$; found: m/z = 722.2971 $[M+H]^+$.

Compound 37b



Yield: 78%; **IR**: $\tilde{\nu}$ = 3229 (w), 2930 (w), 2858 (w), 1740 w), 1697 (vs), 1612 (s), 1519 (vs), 1470 (s), 1345 (vs), 1254 (s), 1216 (w), 838 (vs), 781 (s), 736 (vs), 700 (s) cm^{-1}; **^{1}H NMR (400 MHz, CD_2Cl_2) δ**: 9.79 (br. s, 1H), 8.58 (br. s, 1H), 8.39 (s, 1H), 8.17 (t, *J* = 4.0 Hz, 1H), 8.05 (d, *J* = 8.7 Hz, 2H), 7.35 (d, *J* = 8.7 Hz, 2H), 7.32 – 7.23 (m, 2H), 7.21-7.15 (m, 2H), 5.87 (d, *J* = 7.1 Hz, 1H), 5.61 (dd, *J* = 11.6 Hz, 2.4 Hz, 1H), 5.05 (dd, *J* = 7.2 Hz, 4.7 Hz, 1H), 4.82 (q, *J* = 6.2 Hz, 1H), 4.46 – 4.30 (m, 4H), 3.93 (d, *J* = 12.9 Hz, 1H), 3.79 – 3.68 (m, 1H), 3.17 (d, *J* = 6.2 Hz, 2H), 3.03 (t, *J* = 6.5 Hz, 3H), 2.88 (s, 1H), 0.80 (s, 9H), -0.17 (s, 3H), -0.36 (s, 3H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.4, 153.1, 150.7, 150.5, 149.8, 149.3, 146.8, 145.8, 143.3, 136.3, 129.8, 129.4, 128.5, 128.3, 127.2, 123.5, 91.0, 87.6, 74.8, 72.7, 64.7, 63.1, 54.9, 37.8, 34.7, 25.3, 17.7, -5.6, -5.7; **HRMS (ESI)**: calculated for $C_{34}H_{43}N_7O_9SiNa^+$: m/z = 744.2784 $[M+Na]^+$; found: m/z = 744.2776 $[M+Na]^+$.

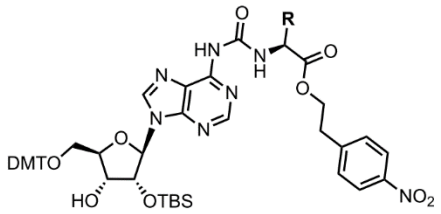
Compound 38



Yield: 96%; **IR**: $\tilde{\nu}$ = 3240 (w), 2950 (w), 2927 (w), 2856 (w), 1737 (w), 1695 (s), 1611 (s), 1589 (s), 1520 (vs), 1469 (s), 1346 (vs), 1313 (w), 1250 (vs), 1130 (w), 1093 (s), 835 (vs), 780 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.11 (d, *J* = 8.2 Hz, 1H), 9.13 (s, 1H), 8.72 (s, 1H), 8.52 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 6.13 (d, *J* = 6.1 Hz, 1H), 5.07 (dd, *J* = 8.7, 3.8 Hz, 1H), 4.99 (dd, *J* = 6.1, 4.7 Hz, 1H), 4.66 (dt, *J* = 8.3, 3.0 Hz, 1H), 4.45 (td, *J*

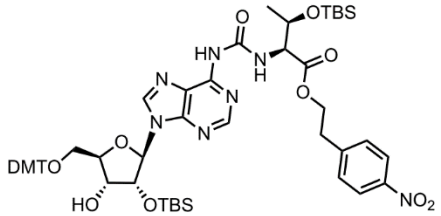
= 6.5, 1.2 Hz, 2H), 4.39 (td, *J* = 4.3, 2.5 Hz, 1H), 4.22 (p, *J* = 2.8 Hz, 1H), 4.15 (dd, *J* = 10.3, 2.9 Hz, 1H), 4.04 (d, *J* = 3.8 Hz, 1H), 3.98 (dd, *J* = 10.2, 3.2 Hz, 1H), 3.94 – 3.87 (m, 1H), 3.78 (ddd, *J* = 12.4, 8.5, 2.5 Hz, 1H), 3.16 (t, *J* = 6.3 Hz, 2H), 0.88 (s, 9H), 0.79 (s, 9H), 0.07 (s, 3H), 0.02 (s, 3H), -0.08 (s, 3H), -0.23 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 171.0, 154.1, 151.4, 151.0, 147.4, 144.1, 131.1, 124.2, 122.1, 90.5, 87.8, 76.8, 72.6, 65.7, 64.3, 63.0, 56.4, 35.3, 27.2, 26.1, 26.0, 18.7, 18.6, -4.9, -5.2, -5.3, -5.6; **HRMS (ESI)**: calculated for $C_{34}H_{54}N_7O_{10}Si_2Na^+$: m/z = 798.3290 $[M+Na]^+$; found: m/z = 798.3288 $[M+Na]^+$.

General procedure for DMT protection:



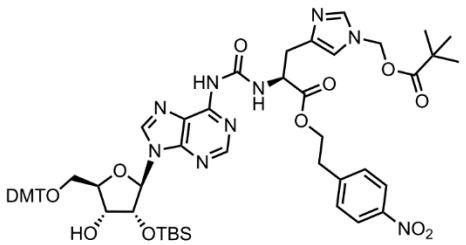
The 3'-5'-deprotected adenosine derivative (1 eq.) was dissolved in pyridine under N_2 atmosphere. DMT chloride (1.5 eq.) was added in two portions and the mixture was stirred at room temperature overnight. Then the volatiles were evaporated, and crude product was purified by flash chromatography eluting with CH_2Cl_2/MeOH (10/1/, v/v) with an addition of 0.1% of pyridine, unless otherwise specified, to afford the DMT protected derivative as white foam.

Compound 39



Yield: 90%; **IR**: $\tilde{\nu}$ = 3350 (w), 2930 (w), 2856 (w), 1729 (w), 1684 (s), 1608 (s), 1521 (vs), 1464 (s), 1345 (vs), 1248 (vs), 1174 (w), 1094 (w), 10033 (s), 827 (vs), 777 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.00 (d, *J* = 9.0 Hz, 1H), 8.95 (s, 1H), 8.55 (s, 1H), 8.40 (s, 1H), 7.98 (d, *J* = 8.7 Hz, 2H), 7.59 – 7.46 (m, 4H), 7.35 (dd, *J* = 9.0, 3.0 Hz, 4H), 7.28 – 7.13 (m, 4H), 6.83 (dd, *J* = 9.0, 3.0 Hz, 4H), 6.16 (d, *J* = 4.6 Hz, 1H), 5.16 (t, *J* = 4.6 Hz, 1H), 4.61 – 4.48 (m, 3H), 4.45 – 4.35 (m, 2H), 4.33 – 4.26 (m, 1H), 3.99 (d, *J* = 5.8 Hz, 1H), 3.75 (s, 6H), 3.51 – 3.44 (m, 2H), 3.12 (t, *J* = 6.2 Hz, 2H), 1.27 (d, *J* = 6.2 Hz, 3H), 0.91 (s, 9H), 0.85 (s, 9H), 0.09 (s, 3H), 0.06 (s, 3H), -0.05 (s, 6H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 171.6, 159.6, 154.8, 151.5, 151.3, 147.5, 146.1, 136.7, 131.1, 130.9, 129.7, 129.0, 128.5, 127.5, 126.1, 124.1, 121.9, 113.8, 90.2, 87.1, 84.8, 76.3, 71.9, 69.6, 65.6, 64.4, 60.3, 55.5, 35.3, 26.1, 25.9, 21.5, 18.7, 18.5, -4.1, -4.6, -4.8, -5.2; **HRMS (ESI)**: calculated for $C_{56}H_{74}N_7O_{12}Si_2^+$: m/z = 1092.4929 $[M+H]^+$; found: m/z = 1092.4937 $[M+H]^+$.

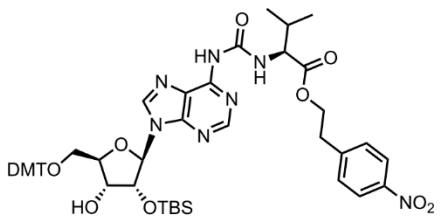
Compound 40



Eluent: 10% CH_2Cl_2 in EtOAc to pure EtOAc containing 0.1% of pyridine.

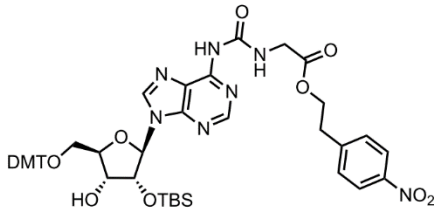
Yield: 85%; **IR**: $\tilde{\nu}$ = 3854 (w), 3746 (w), 3650 (w), 3630 (w), 2931 (w), 2361 (w), 2341 (w), 1740 (s), 1700 (vs), 1654 (w), 1609 (s), 1587 (w), 1559 (w), 1508 (vs), 1472 (s), 1396 (w), 1345 (s), 1300 (w), 1250 (vs), 1176 (s), 1118 (vs), 1032 (vs), 986 (w), 912 (w), 835 (vs), 781 (s), 751 (w), 700 (w) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.11 – 10.06 (m, 1H), 8.61 – 8.55 (m, 1H), 8.47 (s, 1H), 8.45 (s, 1H), 8.13 – 8.05 (m, 2H), 7.71 (d, *J* = 1.4 Hz, 1H), 7.63 – 7.55 (m, 2H), 7.53 – 7.46 (m, 2H), 7.40 – 7.32 (m, 4H), 7.28 (td, *J* = 8.2, 7.7, 2.0 Hz, 2H), 7.25 – 7.18 (m, 1H), 7.06 (s, 1H), 6.88 – 6.81 (m, 4H), 6.13 (d, *J* = 4.4 Hz, 1H), 5.88 (s, 2H), 5.10 (t, *J* = 4.6 Hz, 1H), 4.75 (dq, *J* = 8.0, 4.2, 3.3 Hz, 1H), 4.56 – 4.51 (m, 1H), 4.47 – 4.36 (m, 2H), 4.30 – 4.25 (m, 1H), 4.03 – 3.93 (m, 1H), 3.77 (s, 3H), 3.76 (s, 3H), 3.50 – 3.40 (m, 2H), 3.16 – 3.11 (m, 2H), 3.12 – 3.04 (m, 2H), 1.05 (s, 9H), 0.86 (s, 9H), 0.06 (s, 3H), -0.04 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 177.9, 172.1, 159.6, 153.9, 151.7, 151.3, 147.7, 147.6, 146.1, 143.3, 139.2, 138.8, 136.7, 131.1, 131.0, 130.0, 129.0, 128.6, 127.6, 124.6, 124.1, 118.2, 113.9, 90.1, 87.1, 84.7, 76.4, 71.9, 71.8, 68.8, 66.2, 64.4, 55.5, 39.2, 35.3, 31.0, 27.0, 26.1, 18.7, -4.6, -4.8; **HRMS (ESI)**: calculated for $C_{58}H_{70}N_9O_{13}Si^+$: m/z = 1128.4857 $[M+H]^+$; found: m/z = 1128.4885 $[M+H]^+$.

Compound 41



Yield: 70%; **IR**: $\tilde{\nu}$ = 2929 (w), 1735 (w), 1684 (w), 1569 (s), 1508 (vs), 1464 (s), 1344 (s), 1249 (vs), 1176 (w), 1015 (s), 800 (vs) 749 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.02 (d, *J* = 9.0 Hz, 1H), 8.55 (s, 1H), 8.51 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 7.52 – 7.47 (m, 2H), 7.37 (dd, *J* = 9.0, 2.2 Hz, 4H), 7.30 – 7.13 (m, 4H), 6.86 – 6.82 (m, 4H), 6.16 (d, *J* = 4.6 Hz, 1H), 5.16 (t, *J* = 4.6 Hz, 1H), 4.57 – 4.39 (m, 5H), 4.30 (q, *J* = 4.6 Hz, 1H), 4.00 (d, *J* = 5.8 Hz, 1H), 3.76 (s, 6H), 3.47 (d, *J* = 4.2 Hz, 2H), 3.16 (t, *J* = 6.5 Hz, 2H), 2.31 – 2.30 (m, 1H), 0.99 (d, *J* = 6.5 Hz, 3H), 0.95 (d, *J* = 6.5 Hz, 3H), 0.85 (s, 9H), 0.06 (s, 3H), -0.04 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 172.2, 159.5, 154.3, 151.6, 151.3, 147.4, 146.1, 143.5, 136.6, 131.0, 130.0, 128.9, 128.9, 127.5, 126.1, 124.2, 121.7, 113.8, 90.3, 87.0, 84.7, 76.4, 71.9, 65.2, 64.3, 59.5, 55.4, 35.3, 31.5, 26.1, 19.6, 18.7, 18.2, -4.6, -4.8; **HRMS (ESI)**: calculated for $C_{51}H_{62}N_7O_{11}Si^+$: m/z = 976.4277 $[M+H]^+$; found: m/z = 976.4287 $[M+H]^+$.

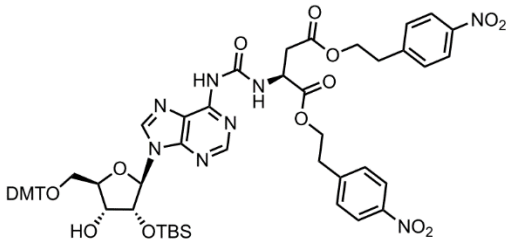
Compound 42



Yield: 85%; **IR**: $\tilde{\nu}$ = 2931 (w), 1748 (w), 1703 (s), 1609 (s), 1588 (w), 1509 (s), 1469 (s), 1345 (vs), 1250 (vs), 1177 (vs), 1035 (w), 835 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 9.93 (t, *J* = 5.5 Hz, 1H), 9.05 (br. s, 1H), 8.58 (s, 1H), 8.46 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.58 (d, *J* = 8.7 Hz, 2H), 7.54 – 7.47 (m, 2H), 7.41 – 7.33 (m, 4H), 7.31 – 7.23 (m, 2H), 7.23 – 7.16 (m, 1H), 6.88 – 6.78 (m, 4H), 6.16 (d, *J* = 4.4 Hz, 1H), 5.10 (dd, *J* = 4.7 Hz, 1H), 4.53 (t, *J* = 5.1 Hz, 1H), 4.44 (t, *J* = 6.4 Hz, 2H), 4.30 (dt, *J* = 4.4 Hz, 1H), 4.13 (d, *J* = 5.5 Hz, 2H), 4.00 (d, *J* = 5.7 Hz,

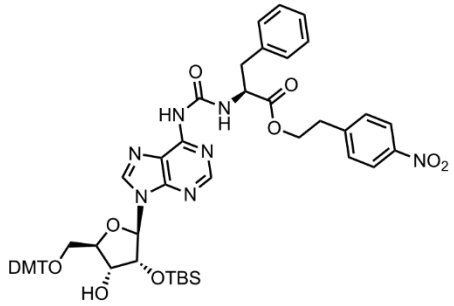
1H), 3.75 (s, 6H), 3.53 – 3.42 (m, 2H), 3.13 (t, *J* = 6.4 Hz, 2H), 0.85 (s, 9H), 0.05 (s, 3H), -0.05 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 170.6, 159.6, 154.6, 151.7, 151.3, 151.2, 147.6, 147.4, 146.1, 143.6, 136.7, 131.1, 130.9, 129.1, 129.0, 128.6, 127.5, 124.2, 121.7, 113.8, 90.0, 87.1, 84.8, 76.4, 72.0, 65.2, 64.4, 55.5, 42.6, 35.3, 26.1, 18.7, -4.7, -4.8; **HRMS (ESI)**: calculated for $C_{48}H_{55}N_7O_{11}Si^+$: m/z = 934.3802 $[M + H]^+$; found: m/z = 934.3812 $[M + H]^+$.

Compound 43



Yield: 95%; **IR**: $\tilde{\nu}$ = 2930 (w), 1734 (s), 1698 (s), 1607 (s), 1509 (vs), 1466 (s), 1370 (vs), 1248 (s), 1174 (s), 1032 (s), 829 (vs), 699 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.29 (d, *J* = 7.9 Hz, 1H), 9.18 (s, 1H), 8.60 (s, 1H), 8.42 (s, 1H), 8.10 – 8.04 (m, 4H), 7.58 – 7.51 (m, 4H), 7.39 – 7.36 (m, 4H), 7.30 – 7.25 (m, 2H), 7.21 – 7.18 (m, 2H), 6.86 – 6.82 (m, 5H), 6.17 (d, *J* = 4.2 Hz, 1H), 5.10 (t, *J* = 4.5 Hz, 1H), 4.85 (dt, *J* = 7.9, 5.2 Hz, 1H), 4.56 (q, *J* = 5.2 Hz, 1H), 4.46 – 4.28 (m, 5H), 4.02 (d, *J* = 5.9 Hz, 1H), 3.75 (d, *J* = 2.1 Hz, 6H), 3.53 – 3.44 (m, 2H), 3.13 – 3.06 (m, 4H), 3.00 – 2.96 (m, 2H), 0.86 (s, 9H), 0.07 (s, 3H), 0.03 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 171.3, 171.2, 159.5, 154.2, 151.5, 150.6, 147.4, 147.2, 146.1, 143.6, 136.7, 131.0, 130.0, 129.0, 128.6, 127.5, 124.2, 113.8, 90.1, 87.0, 84.6, 76.4, 71.8, 65.7, 65.1, 64.3, 55.4, 50.6, 37.1, 35.2, 26.1, 18.6, -4.6, -4.8; **HRMS (ESI)**: calculated for $C_{58}H_{64}N_8O_{15}SiNa^+$: m/z = 1163.4158 $[M+Na]^+$; found: m/z = 1163.4159 $[M+Na]^+$.

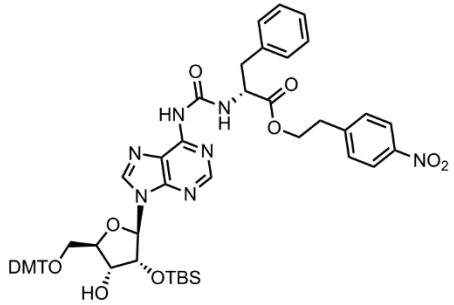
Compound 44a



Eluent: CH_2Cl_2/MeOH (20/1, v/v).

Yield: 82%; **IR**: $\tilde{\nu}$ = 3245 (w), 2952 (w), 2931 (w), 1741 (w), 1699 (s), 1608 (s), 1587 (w), 1519 (s), 1509 (s), 1469 (w), 1345 (s), 1249 (s), 1176 (s), 1034 (w), 834 (s), 783 (w), 735 (vs), 701 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CD_2Cl_2) δ**: 9.83 (d, *J* = 7.5 Hz, 1H), 8.29 (s, 1H), 8.17 (s, 1H), 8.06 (d, *J* = 8.7 Hz, 2H), 8.00 (s, 1H), 7.46 (d, *J* = 6.8 Hz, 2H), 7.38 – 7.32 (m, 7H), 7.31 – 7.20 (m, 7H), 7.18 – 7.14 (m, 2H), 6.82 (d, *J* = 8.9 Hz, 4H), 6.02 (d, *J* = 5.0 Hz, 1H), 5.02 (t, *J* = 5.0 Hz, 1H), 4.80 (q, *J* = 6.3 Hz, 1H), 4.44 – 4.32 (m, 3H), 4.24 (q, *J* = 3.8 Hz, 1H), 3.77 (s, 6H), 3.48 (dd, *J* = 10.7 Hz, *J* = 3.2 Hz, 1H), 3.39 (dd, *J* = 10.7 Hz, *J* = 4.3 Hz, 1H), 3.16 (d, *J* = 6.3 Hz, 2H), 3.03 (t, *J* = 6.5 Hz, 2H), 2.67 (d, *J* = 4.7 Hz, 1H), 0.85 (s, 9H), 0.01 (s, 3H), -0.12 (s, 3H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.5, 153.3, 150.8, 150.6, 149.8, 149.4, 146.8, 145.8, 143.6, 136.2, 135.9, 129.9, 129.4, 128.5, 128.3, 127.2, 123.7, 123.6, 121.8, 91.0, 87.5, 74.8, 72.7, 64.8, 63.1, 54.8, 37.8, 34.7, 25.3, 17.7, -5.7; **HRMS (ESI)**: calculated for $C_{55}H_{62}N_7O_{11}Si^+$: m/z = 1024.4271 $[M+H]^+$; found: m/z = 1024.429 $[M+H]^+$.

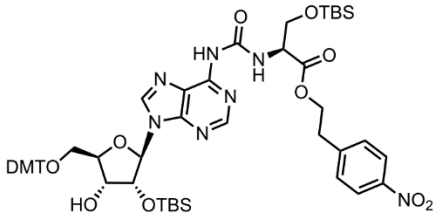
Compound 44b



Eluent: CH_2Cl_2/MeOH (20/1, v/v).

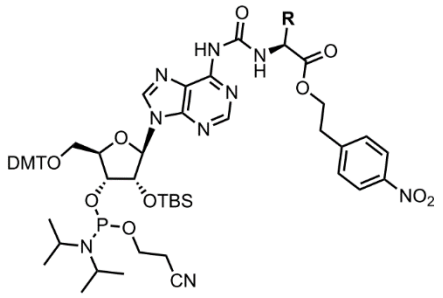
Yield: 85%; **IR**: $\tilde{\nu}$ = 3231 (w), 2930 (w), 2858 (w), 1741 (w), 1697 (s), 1612 (s), 1587 (w), 1518 (vs), 1470 (s), 1344 (vs), 1253 (w), 1180 (w), 1134 (w), 1085 (w), 836 (s), 780 (s), 734 (vs), 698 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CD_2Cl_2) δ**: 9.83 (d, *J* = 7.5 Hz, 1H), 8.58 (br. s, 1H), 8.32 (s, 1H), 8.19 (s, 1H), 8.06 (d, *J* = 8.7 Hz, 2H), 7.99 (s, 1H), 7.46 (d, *J* = 6.9 Hz, 2H), 7.38 – 7.31 (m, 7H), 7.31 – 7.21 (m, 7H), 7.20 – 7.16 (m, 2H), 6.82 (d, *J* = 8.9 Hz, 4H), 6.05 (d, *J* = 4.7 Hz, 1H), 4.94 (t, *J* = 5.0 Hz, 1H), 4.81 (q, *J* = 6.3 Hz, 1H), 4.44 – 4.33 (m, 3H), 4.24 (q, *J* = 3.8 Hz, 1H), 3.77 (s, 6H), 3.48 (dd, *J* = 10.7 Hz, 3.1 Hz, 1H), 3.41 (dd, *J* = 10.7 Hz, 4.2 Hz, 1H), 3.16 (d, *J* = 6.9 Hz, 2H), 3.03 (t, *J* = 6.5 Hz, 2H), 2.66 (d, *J* = 5.0 Hz, 1H), 0.86 (s, 9H), 0.02 (s, 3H), -0.09 (s, 3H); **^{13}C NMR (101 MHz, CD_2Cl_2) δ**: 171.4, 153.4, 150.7, 150.5, 149.8, 149.4, 146.8, 145.9, 143.7, 136.3, 129.9, 129.5, 128.5, 128.3, 127.2, 123.5, 123.5, 121.8, 91.0, 87.5, 74.9, 72.7, 64.8, 63.1, 54.9, 37.8, 34.7, 25.3, 17.8, -5.6; **HRMS (ESI)**: calculated for $C_{55}H_{62}N_7O_{11}Si^+$: m/z = 1024.4271 $[M+H]^+$; found: m/z = 1024.4291 $[M+H]^+$.

Compound 45



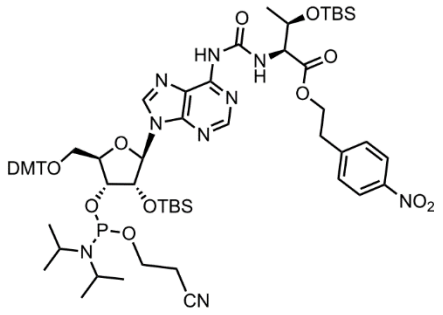
Yield: 90%; **IR**: $\tilde{\nu}$ = 2930 (w), 1740 (s), 1694 (s), 1607 (w), 1582 (w), 1519 (vs), 1465 (w), 1345 (s), 1249 (s), 1176 (w), 1110 (w), 1034 (w), 820 (vs), 778 (s), 699 (w) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.13 (d, *J* = 8.2 Hz, 1H), 8.99 (s, 1H), 8.58 (s, 1H), 8.44 (s, 1H), 8.09 (d, *J* = 8.7 Hz, 2H), 7.58 (d, *J* = 8.7 Hz, 2H), 7.51 (dd, *J* = 8.4, 1.3 Hz, 2H), 7.39 (d, *J* = 2.5 Hz, 2H), 7.37 (d, *J* = 2.4 Hz, 2H), 7.32 – 7.18 (m, 4H), 6.87 – 6.83 (m, 4H), 6.18 (d, *J* = 4.7 Hz, 1H), 5.13 (t, *J* = 5.0 Hz, 1H), 4.62 (dt, *J* = 8.2, 2.8 Hz, 1H), 4.53 (q, *J* = 5.0 Hz, 1H), 4.45 (t, *J* = 6.3 Hz, 2H), 4.30 (td, *J* = 4.6, 3.4 Hz, 1H), 4.14 (dd, *J* = 10.2, 2.8 Hz, 1H), 3.96 (dd, *J* = 10.2, 2.8 Hz, 1H), 3.77 (s, 6H), 3.48 (qd, *J* = 10.5, 4.1 Hz, 2H), 3.16 (t, *J* = 6.3 Hz, 2H), 0.89 (s, 9H), 0.85 (s, 9H), 0.07 (s, 3H), 0.06 (s, 3H), 0.03 (s, 3H), -0.06 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 170.0, 159.5, 154.1, 151.6, 150.6, 147.4, 146.1, 143.6, 136.7, 131.1, 130.9, 129.0, 128.6, 127.6, 124.6, 124.2, 121.8, 113.8, 89.9, 87.1, 84.9, 76.5, 72.0, 65.8, 64.3, 56.4, 55.4, 35.3, 27.2, 26.1, 18.7, 18.4, -4.6, -4.9, -5.3, -5.5; **HRMS (ESI)**: calculated for $C_{55}H_{71}N_7O_{12}Si_2Na^+$: m/z = 1100.4597 $[M+Na]^+$; found: m/z = 1100.4620 $[M+Na]^+$.

General synthesis of RNA phosphoramidites



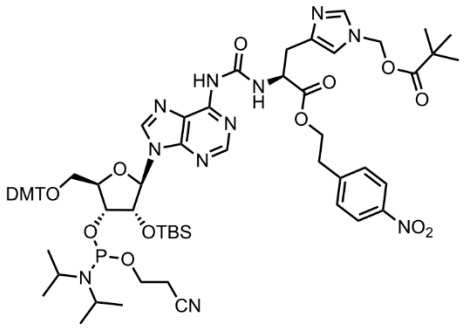
In an oven-dried flask under argon atmosphere, 5'-DMT protected compound (1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 °C. Hünig's base (4 eq.) was added dropwise followed by the addition of 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (2.5 eq.). The solution was stirred at room temperature for 2 h. The reaction was quenched by addition of sat. $NaHCO_3$ solution, and then extracted three times with CH_2Cl_2. The organic phase was dried over Na_2SO_4, filtered and concentrated in vacuo. The residue was purified by flash chromatography, eluting with Hex/EtOAc (1/1, v/v) containing 0.1% of pyridine, unless otherwise specified. The phosphoramidite was obtained as a mixture of diastereomers as white foam.

Compound 46



Yield: 85%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.16, 148.45; **HRMS (ESI)**: calculated for $C_{65}H_{91}N_9O_{13}PSi_2^+$: m/z = 1292.6007 $[M+H]^+$; found: m/z = 1292.6033 $[M+H]^+$.

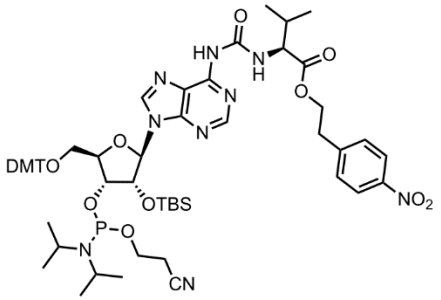
Compound 47



Eluent: 30% CH_2Cl_2 in EtOAc containing 0.1% of pyridine.

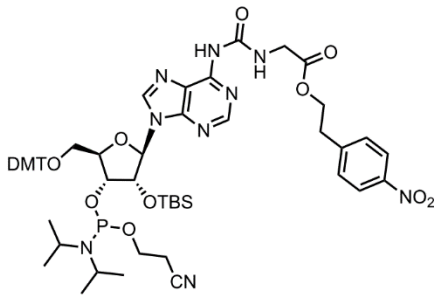
Yield: 66%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.26, 148.61; **HRMS (ESI)**: calculated for $C_{67}H_{87}N_{11}O_{14}PSi^+$: m/z = 1328.5935 $[M+H]^+$; found: m/z = 1328.5944 $[M+H]^+$.

Compound 48



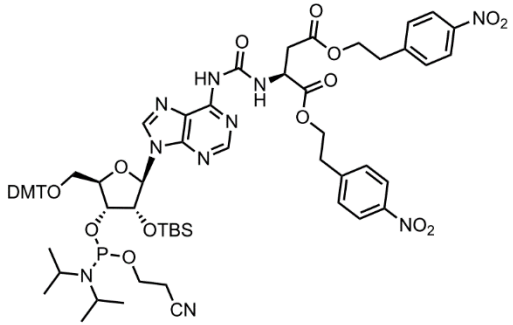
Yield: 87%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.22, 148.65; **HRMS (ESI)**: calculated for $C_{60}H_{79}N_9O_{12}PSi^+$: m/z = 1176.5355 $[M+H]^+$; found: m/z = 1176.5359 $[M+H]^+$.

Compound 49



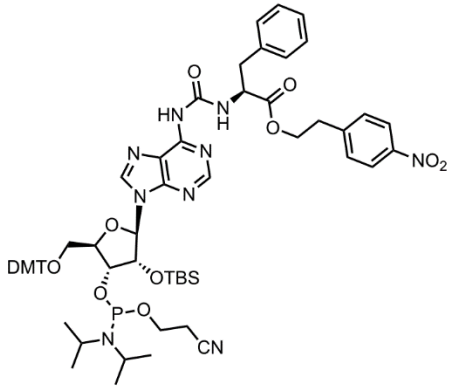
Yield: 86%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.32, 148.60; **HRMS (ESI)**: calculated for $C_{57}H_{72}N_9O_{12}SiPSi^+$: m/z = 1134.4880 $[M+H]^+$; found: m/z = 1134.4894 $[M + H]^+$.

Compound 50



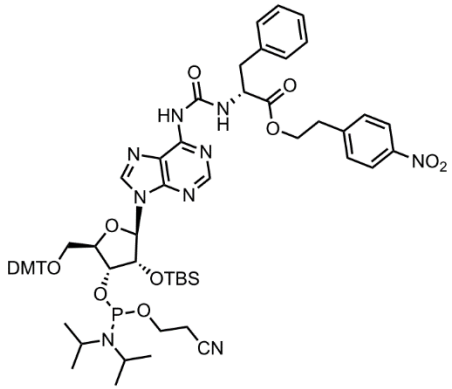
Yield: 65%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.15, 148.67; **HRMS (ESI)**: calculated for $C_{67}H_{82}N_{10}O_{16}PSi^+$: m/z = 1341.5417 $[M+H]^+$; found: m/z = 1341.5437 $[M+H]^+$.

Compound 51a



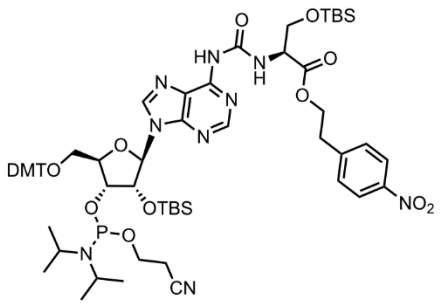
Yield: 67%; **^{31}P NMR (162 MHz, CD_2Cl_2) δ**: 150.7, 149.1; **HRMS (ESI)**: calculated for $C_{64}H_{79}N_9O_{12}PSi^+$: m/z = 1224.5350 $[M+H]^+$; found: m/z = 1224.5374 $[M+H]^+$.

Compound 51b



Yield: 67%; **^{31}P NMR (162 MHz, CD_2Cl_2) δ**: 150.6, 148.9; **HRMS (ESI)**: calculated for $C_{64}H_{79}N_9O_{12}PSi^+$: m/z = 1224.5350 $[M+H]^+$; found: m/z = 1224.5383 $[M+H]^+$.

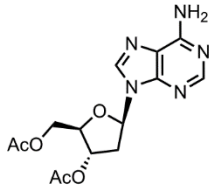
Compound 52



Yield: 70%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.34, 148.45; **HRMS (ESI)**: calculated for $C_{64}H_{89}N_9O_{13}PSi_2^+$: m/z = 1278.5856 $[M+H]^+$; found: m/z = 1278.5877 $[M+H]^+$.

2.2. Synthesis of DNA building block

Compound 54



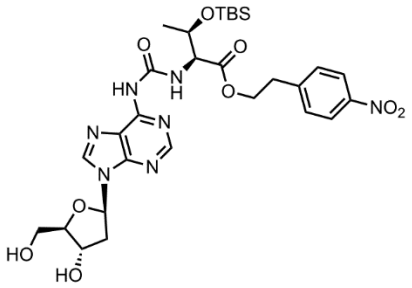
The compound was synthesized according to the published procedure.[5]

Acetic anhydride (1.1 ml, 11.5 mmol, 6.2 eq.) was added to a mixture of deoxyadenosine monohydrate **53** (0.5 g, 1.85 mmol, 1 eq.), pyridine (7 ml) and 4-(dimethylamino)pyridine (25 mg, 0.4 mmol, 0.1 eq.). The reaction mixture was stirred at room temperature for 4 hours. Subsequently, iced water was added, and the mixture was concentrated and co-evaporated with toluene. The compound was used for further steps without additional purification.

Yield: 99%; **^{1}H NMR (400 MHz, $CDCl_3$) δ**: 8.23 (s, 1H), 7.92 (s, 1H), 7.26 (s, 2H), 6.34 (dd, *J* = 7.9, 6.0 Hz, 1H), 5.34 – 5.32 (m, 1H), 4.32 – 4.24 (m, 3H), 2.87 – 2.80 (m, 1H), 2.57 – 2.52 (m, 1H), 2.04 (s, 3H), 2.00 (s, 3H). **HRMS (ESI)**: calculated for $C_{14}H_{18}N_5O_5^+$: m/z = 336.1302 $[M+H]^+$; found: m/z = 336.1305 $[M+H]^+$.

The analytical data is in agreement with the literature.[5]

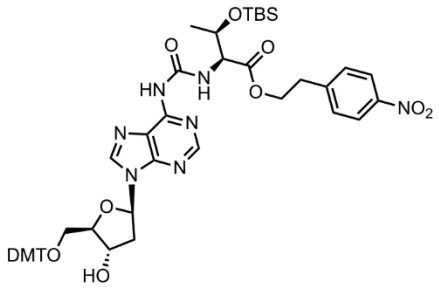
Compound 55



The acetyl-protected deoxyadenosine derivative **54** (0.5 g, 1.5 mmol, 1 eq.) was dissolved in dry CH_2Cl_2 under nitrogen atmosphere. 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**22**; 0.71 g, 3 mmol, 2 eq.) was added to the reaction mixture and the resulting suspension was stirred at room temperature for 2 hours (the solution in time becomes clear). Afterwards the protected threonine derivative **7** (1.1 g, 3 mmol, 2 eq.) was added together with TEA (415 µl, 3 mmol, 2 eq.) as a solution in CH_2Cl_2 and the resulting solution was stirred overnight at room temperature. The reaction was quenched by addition of saturated aqueous $NaHCO_3$ solution. The solution was extracted three times with CH_2Cl_2, and the organic phase was dried, filtered and concentrated *in vacuo*. The acetyl groups were immediately deprotected with 7N NH_3/MeOH (2 ml). After stirring 2 hours at room temperature, the reaction mixture was evaporated. The residue was purified via flash chromatography eluting with CH_2Cl_2/MeOH (10/1, v/v).

Yield: 65%; **IR**: $\tilde{\nu}$ = 3227 (w), 2929 (w), 2855 (w), 1734 (w), 1686 (vs), 1610 (s), 1587 (s), 1518 (vs), 1465 (vs), 1344 (vs), 1312 (w), 1248 (s), 1213 (w), 1094 (vs), 939 (s), 827 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.02 (d, *J* = 8.6 Hz, 1H), 9.40 (s, 1H), 8.70 (s, 1H), 8.39 (s, 1H), 7.91 (d, *J* = 8.2 Hz, 2H), 7.47 (d, *J* = 8.2 Hz, 2H), 6.54 (t, *J* = 6.7 Hz, 1H), 4.67 (s, 1H), 4.51 (d, *J* = 9.3 Hz, 2H), 4.38 (t, *J* = 5.6 Hz, 2H), 4.09 (s, 1H), 3.86 – 3.67 (m, 2H), 3.08 (t, *J* = 5.6 Hz, 2H), 2.90 (dt, *J* = 13.0, 6.7 Hz, 1H), 2.52 – 2.41 (m, 1H), 2.01 (s, 1H), 1.93 (s, 1H), 1.23 (d, *J* = 5.9 Hz, 3H), 0.88 (s, 9H), 0.05 (s, 3H), -0.06 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 171.5, 155.1, 151.1, 150.9, 147.4, 144.2, 130.9, 124.0, 122.0, 89.8, 86.8, 72.6, 69.5, 65.5, 63.3, 41.4, 35.2, 30.6, 26.0, 21.5, 20.5, 18.4, -4.2, -5.2; **HRMS (ESI)**: calculated for $C_{29}H_{42}N_7O_9Si^+$: m/z = 660.2813 $[M+H]^+$; found: m/z = 660.2812 $[M+H]^+$.

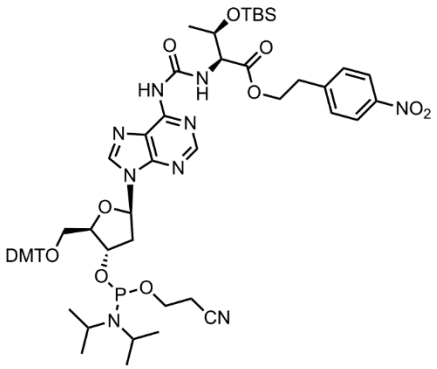
Compound 56



The 3'-5'-deprotected adenosine derivative **55** (0.34 g, 0.52 mmol, 1 eq.) was dissolved in pyridine under N_2 atmosphere. DMT chloride (0.26 g, 0.77 mmol, 1.5 eq.) was added in two portions and the mixture was stirred at room temperature overnight. Then the volatiles were evaporated, and crude product was purified by flash chromatography eluting with CH_2Cl_2/MeOH (10/1/, v/v) with an addition of 0.1% of pyridine to afford the DMT protected derivative.

Yield: 72%; **IR**: $\tilde{\nu}$ = 2960 (w), 2930 (w), 1734 (w), 1696 (s), 1609 (s), 1586 (w), 1520 (vs), 1509 (vs), 1467 (s), 1345 (vs), 1304 (w), 1250 (vs), 1175 (s), 1095 (w), 1034 (s), 940 (w), 828 (vs), 777 (s), 699 (w) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.02 (d, *J* = 9.1 Hz, 1H), 8.98 (s, 1H), 8.52 (s, 1H), 8.34 (s, 1H), 7.96 (d, *J* = 8.7 Hz, 2H), 7.51 (d, *J* = 8.7 Hz, 2H), 7.42 (d, *J* = 7.1 Hz, 2H), 7.31 – 7.28 (m, 4H), 7.24 – 7.12 (m, 3H), 6.79 – 6.74 (m, 4H), 6.55 (t, *J* = 6.4 Hz, 1H), 4.76 (q, *J* = 4.2 Hz, 1H), 4.64 – 4.58 (m, 1H), 4.55 (ddd, *J* = 10.6, 7.7, 1.5 Hz, 2H), 4.41 (t, *J* = 6.2 Hz, 2H), 4.25 – 4.18 (m, 1H), 3.74 (d, *J* = 3.3 Hz, 6H), 3.43 (dd, *J* = 10.2, 5.9 Hz, 1H), 3.35 (dd, *J* = 10.2, 4.0 Hz, 1H), 3.17 – 3.10 (m, 3H), 2.58 – 2.52 (m, 1H), 1.29 (d, *J* = 6.3 Hz, 3H), 0.93 (s, 9H), 0.11 (s, 3H), -0.01 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 171.6, 159.4, 154.9, 151.2, 151.0, 147.4, 146.1, 143.9, 136.8, 131.0, 130.9, 130.8, 128.9, 128.4, 127.4, 124.0, 122.0, 113.7, 87.7, 86.8, 86.1, 72.6, 69.6, 65.4, 65.1, 60.3, 55.4, 40.0, 35.3, 26.0, 21.5, 18.4, -4.1, -5.2; **HRMS (ESI)**: calculated for $C_{50}H_{60}N_7O_{11}Si^+$: m/z = 962.4120 $[M+H]^+$; found: m/z = 962.4136 $[M+H]^+$.

Compound 57



In an oven-dried flask under argon atmosphere, 5'-DMT protected compound **56** (0.1 g, 0.1 mmol, 1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 °C. Hünig's base (72 µl, 0.4 mmol, 4 eq.) was added dropwise followed by the addition of 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (60 µl, 0.25 mmol, 2.5 eq.). The solution was stirred at room temperature for 2 h. The reaction was quenched by addition of sat. $NaHCO_3$ solution, and then extracted three times with CH_2Cl_2. The organic phase was dried over Na_2SO_4, filtered and concentrated in vacuo. The residue was purified by flash chromatography, eluting with Hex/EtOAc (1/2, v/v) containing 0.1% of pyridine. The phosphoramidite was obtained as a mixture of diastereomers as white foam.

Yield: 62%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 148.00, 146.59; **HRMS (ESI)**: calculated for $C_{59}H_{76}N_9NaO_{12}PSi^+$: m/z = 1184.5018 $[M+Na]^+$; found: m/z = 1184.5021 $[M+Na]^+$.

3. Synthesis and Purification of Oligonucleotides

All of the oligonucleotides used in this study were synthesized on a 1 µmol scale using a DNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleosides were purchased from Glen Research and Sigma-Aldrich. Oligonucleotides containing non-canonical nucleosides were synthesized in DMT-OFF mode using phosphoramidites (Bz-A, Dmf-G, Ac-C, U) with BTT in CH_3CN as an activator, DCA in CH_2Cl_2 as a deblocking solution and Ac_2O in pyridine/THF as a capping reagent. For deprotection of npe-group the solid support was treated with DBU solution (10% in THF, 1 ml) for 2 h at room temperature. Subsequently, the supernatant was removed, and the beads were washed with THF and dried under vacuum. The solid support was suspended in a mixture of aqueous ammonia and methylamine (1:1, 1 ml) at room temperature for 1 h. The supernatant was removed, and the beads were washed with water. The supernatant and washings were combined, and the solvents were evaporated under reduced pressure. The residue was subsequently heated with a solution of triethylamine trihydrofluoride (125 µl) in DMSO (50 µl) at 65 °C for 1.5 h. The RNA was precipitated by addition of aqueous NaOAc solution (3M, 25 µl) and *n*-butanol (1 ml). To ensure complete precipitation, the sample was incubated at -80 °C for 1 h. After centrifugation (4 °C, 4000 rpm, 15 min), the supernatant was removed, and the precipitated RNA was lyophilized. The oligonucleotides were further purified by reverse-phase HPLC using a Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel. Oligonucleotides were purified using the following buffer system: 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. A flow rate of 5 ml/min with a gradient of 0-25 % of buffer B in 30 min was applied for the purifications. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow of 0.5 mL/min, a gradient of 0-30% of buffer B in 45 min was applied. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). For strands containing non-canonical base, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. The structural integrity of the synthesized oligonucleotides was analyzed by MALDI-TOF mass measurements.

Sequences of synthesized strands:

ON1: 5' GUCt^6ACCUGA 3'

ON2: 5' GUCg^6ACCUGA 3'

ON3: 5' GUCval^6ACCUGA 3'

ON4: 5' GUChis^6ACCUGA 3'

ON5: 5' GUCasp^6ACCUGA 3'

ON6: 5' GUC-L-phe^6ACCUGA 3'

ON7: 5' GUC-D-phe^6ACCUGA 3'

ON8: 5' AUCGt^6ACUACGt^6AAUCGCt^6AACCG 3'

ON9: 5' AGAUGUG-ser^6A-asp^6A-his^6A-GAGAUGA 3'

ODN1: 5' d(GTCt^6dACCTGA) 3'

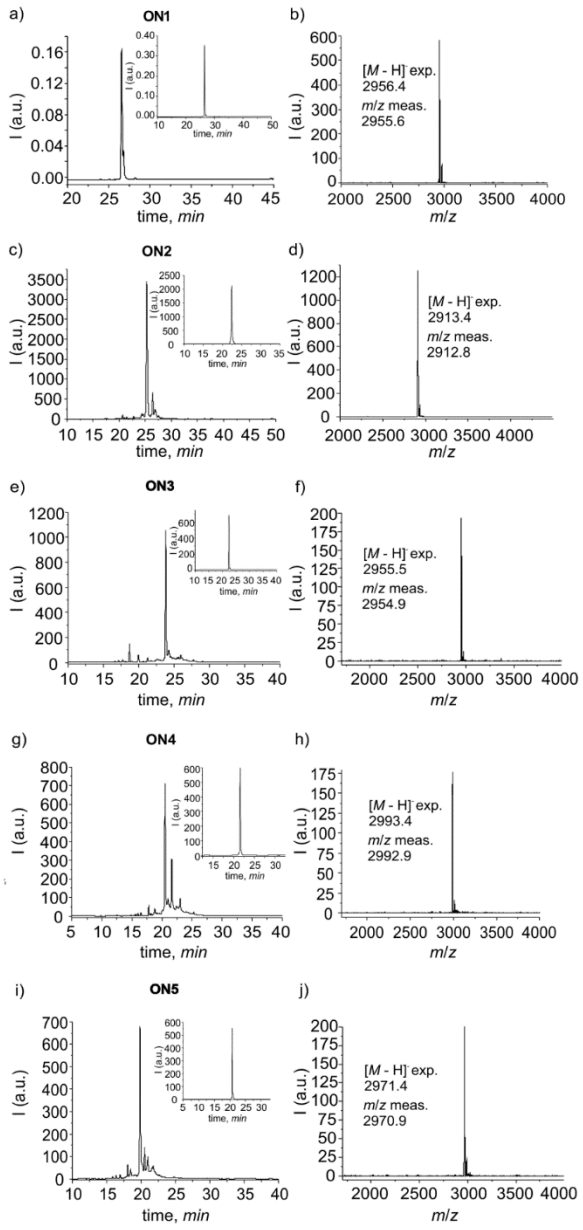


Figure S1. HPLC and MALDI data of oligonucleotides **ON1**-**ON5**.

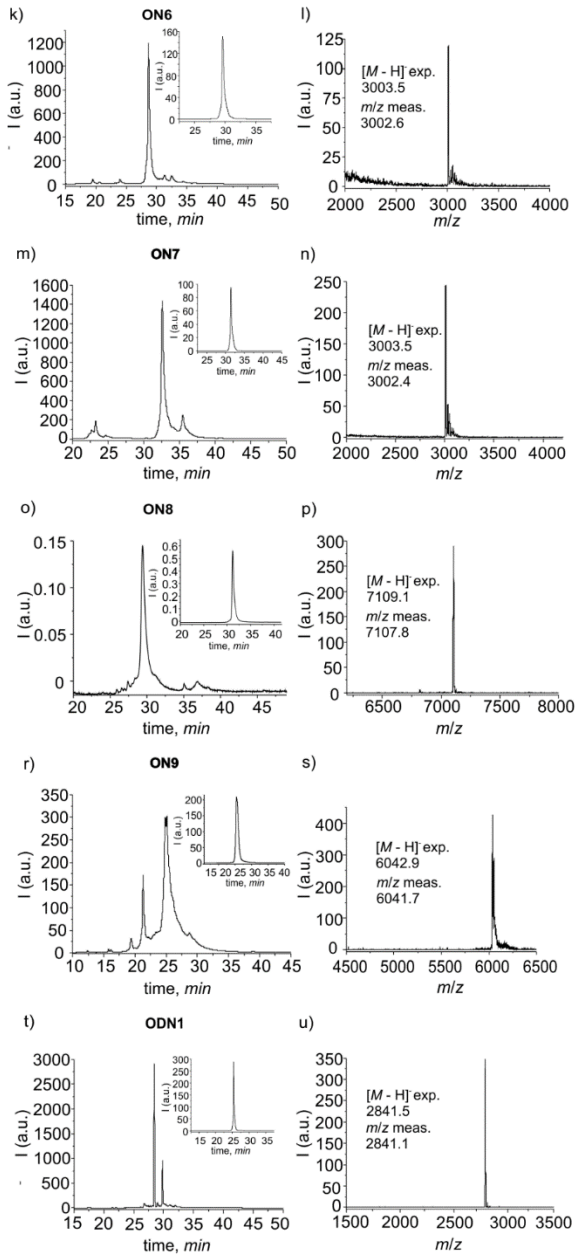


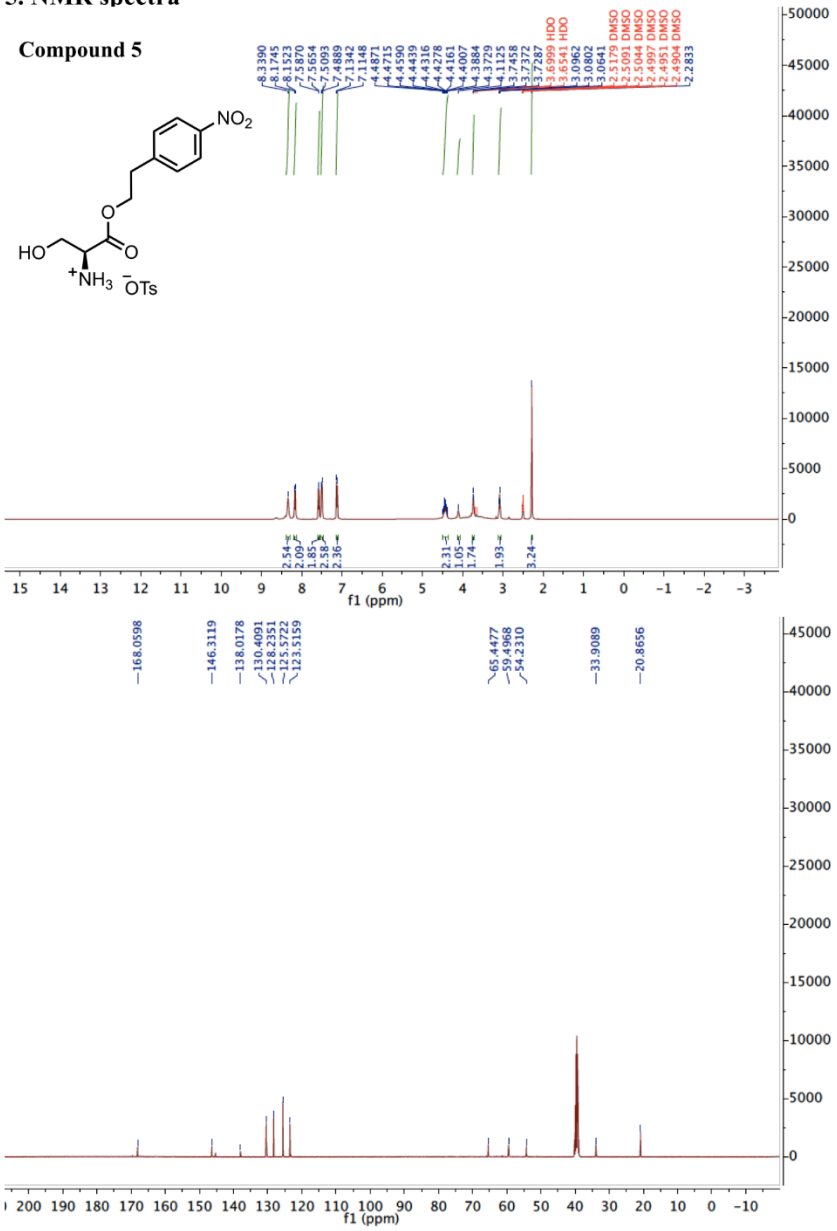
Figure S2. HPLC and MALDI data of oligonucleotides **ON6**-**ODN1**.

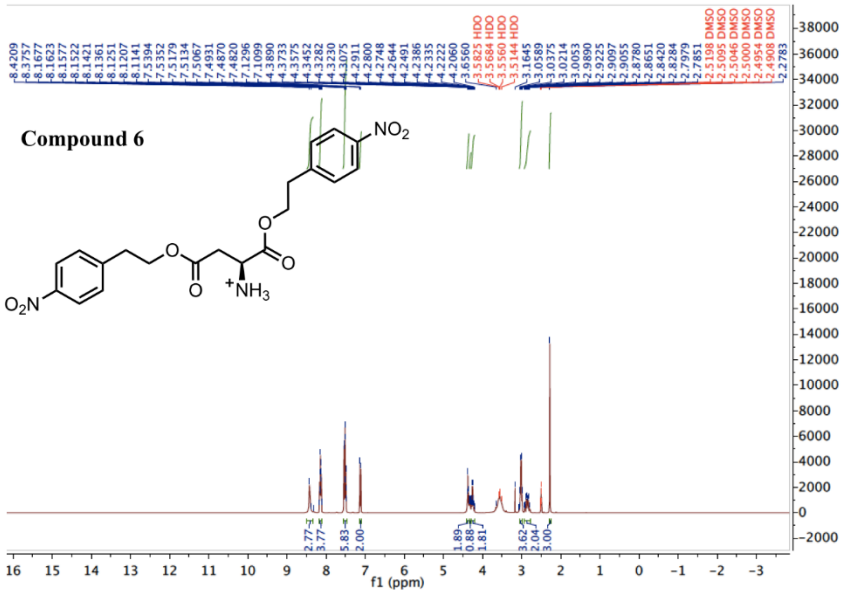
4. UV Melting Curve Measurements

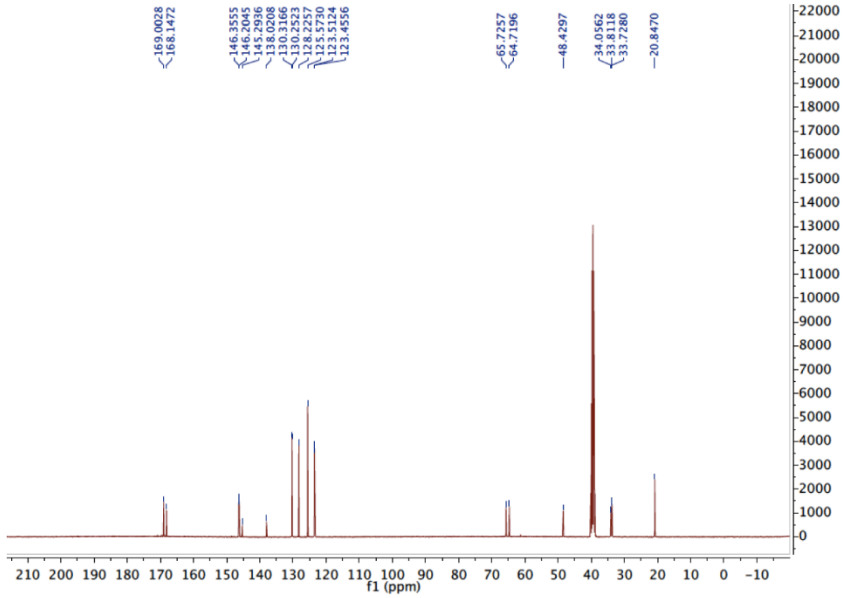
The UV melting curves were measured on JASCO V-650 spectrometer using 10 mm QS cuvettes, purchased from Hellma Analytics. A solution (80 µL) of equimolar amounts of oligonucleotides (4 µM each) in the buffer solution containing 10 mM sodium phosphate buffer (pH 7.0) and 150 mM NaCl was heated at 90 °C for 2 min and gradually cooled to 4 °C prior to the measurement. Melting profiles were recorded at temperatures between 5 and 75 °C with a ramping and scanning rate of 1 °C/min at 260 nm. All samples were measured at least three times. T_m values from each measurement were calculated using the “fitting curve” method and presented as an average of three independent measurements.

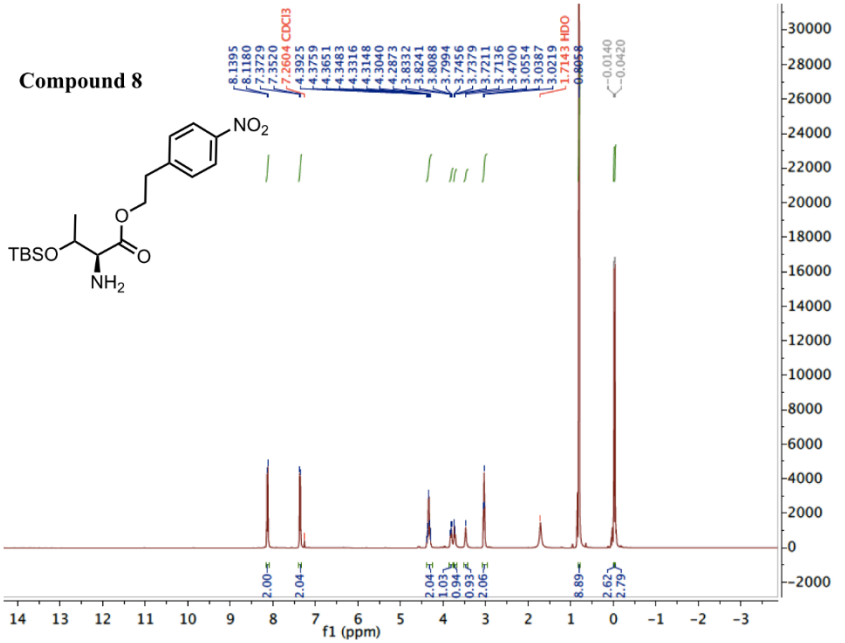
5. NMR spectra

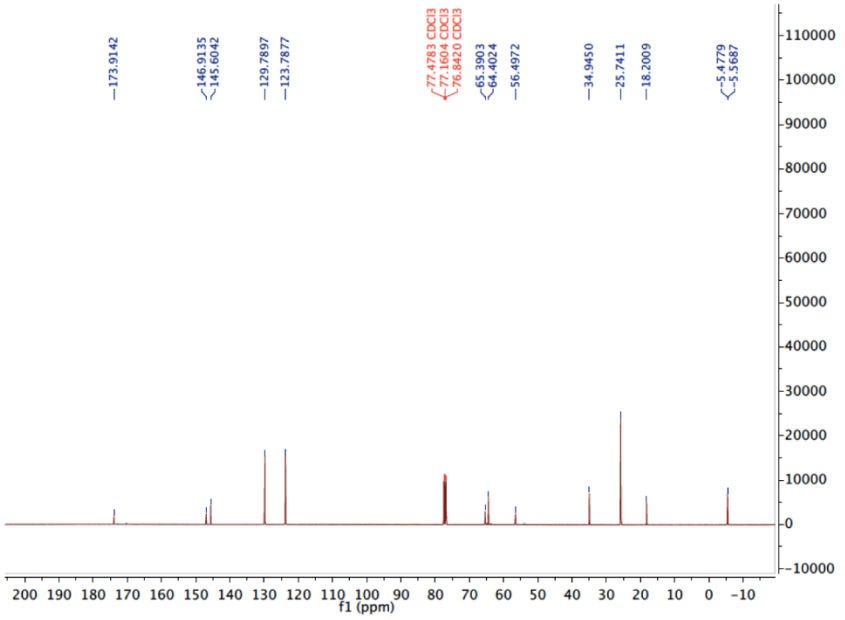
Compound 5



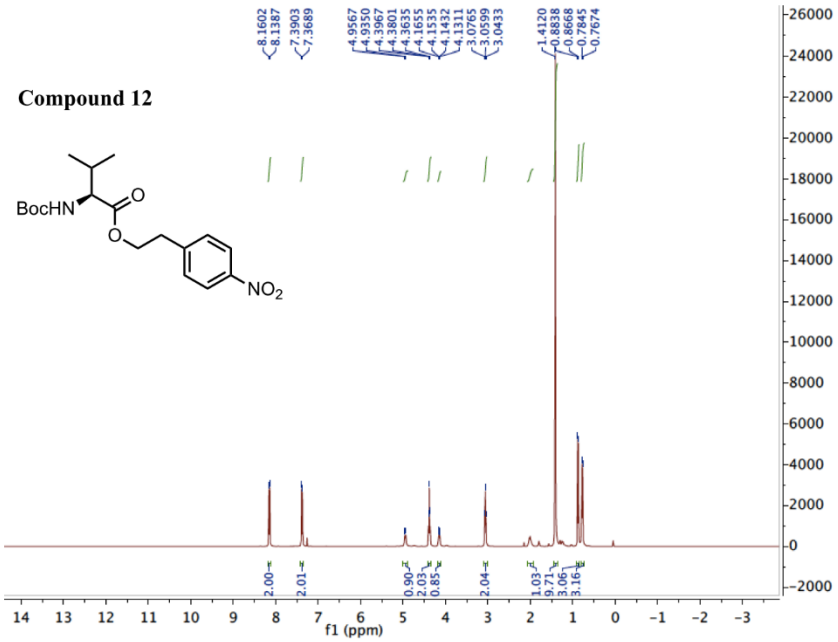

Compound 6
NO2
O2N
+NH3
f1 (ppm)

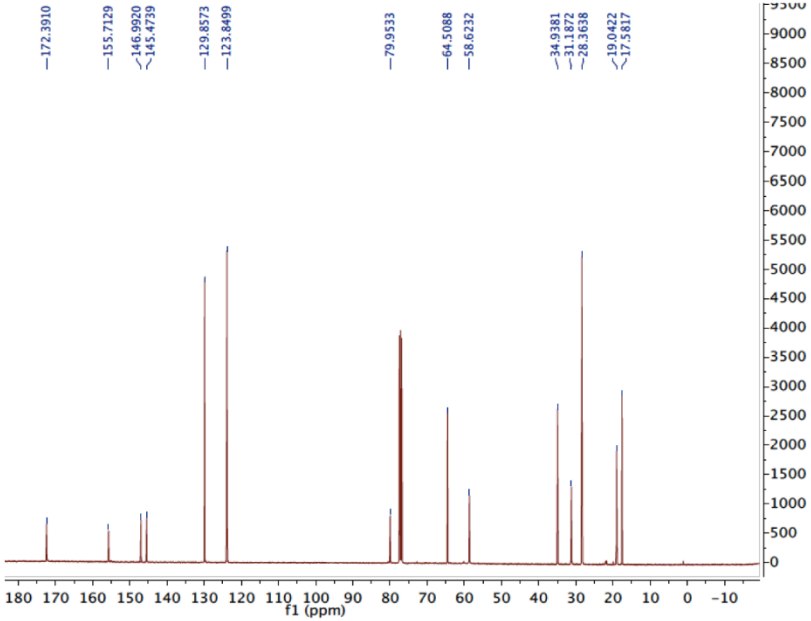

f1 (ppm)


Compound 8
NO2
TBSO
NH2
f1 (ppm)

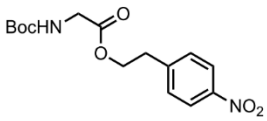

f1 (ppm)

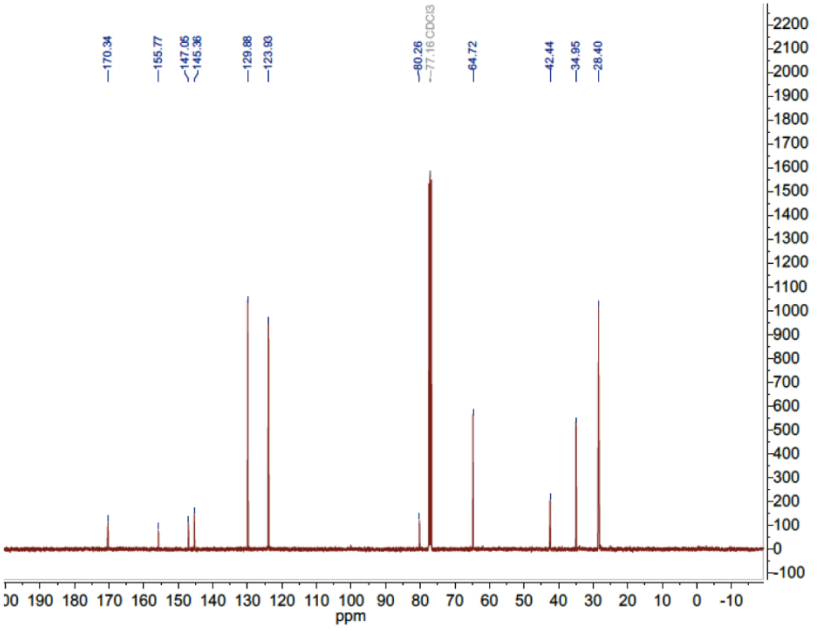
Compound 12

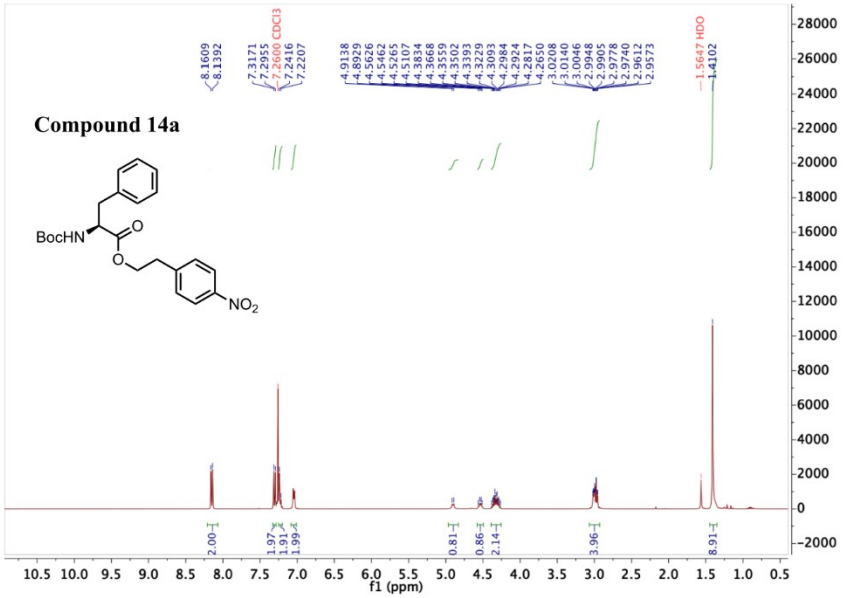


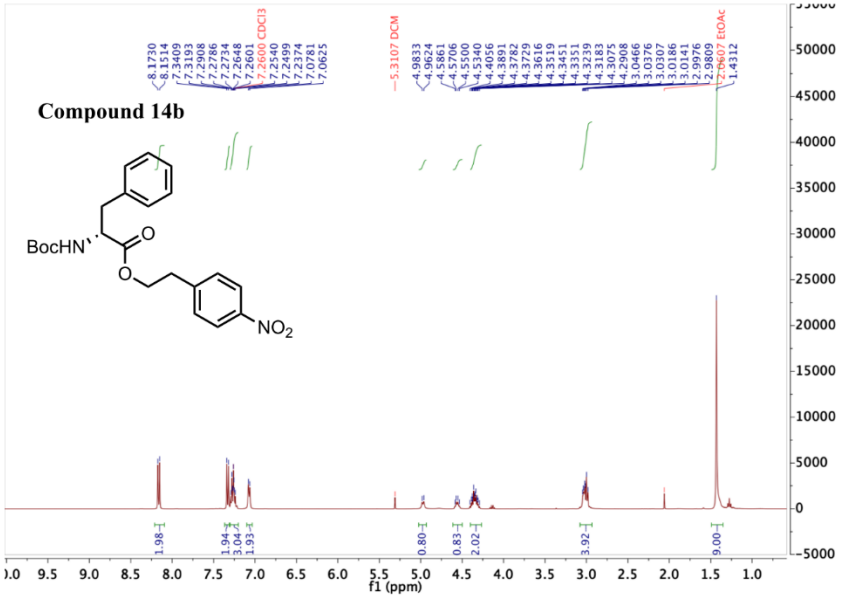


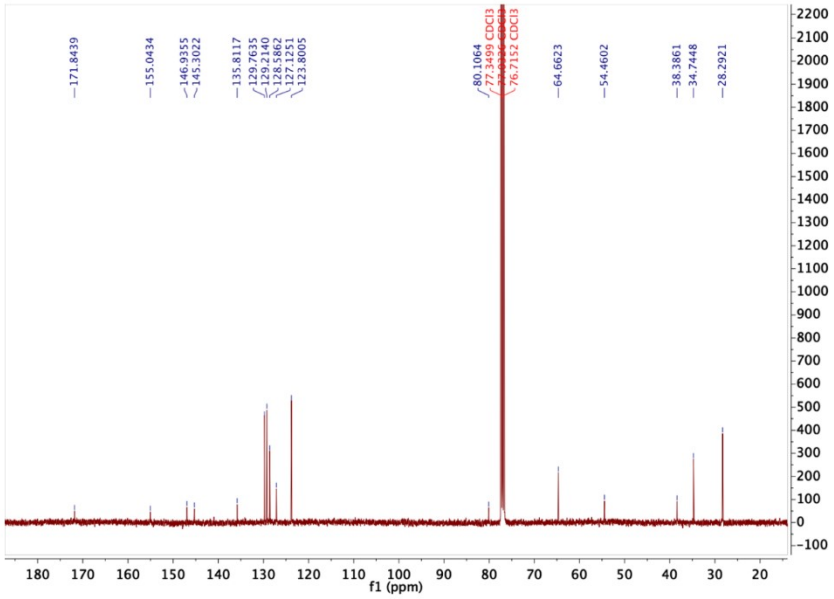
Compound 13

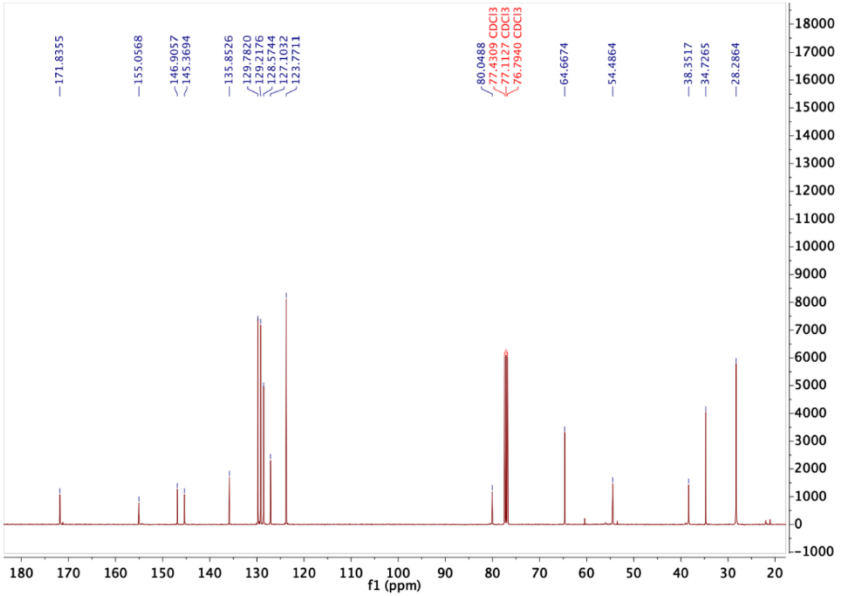


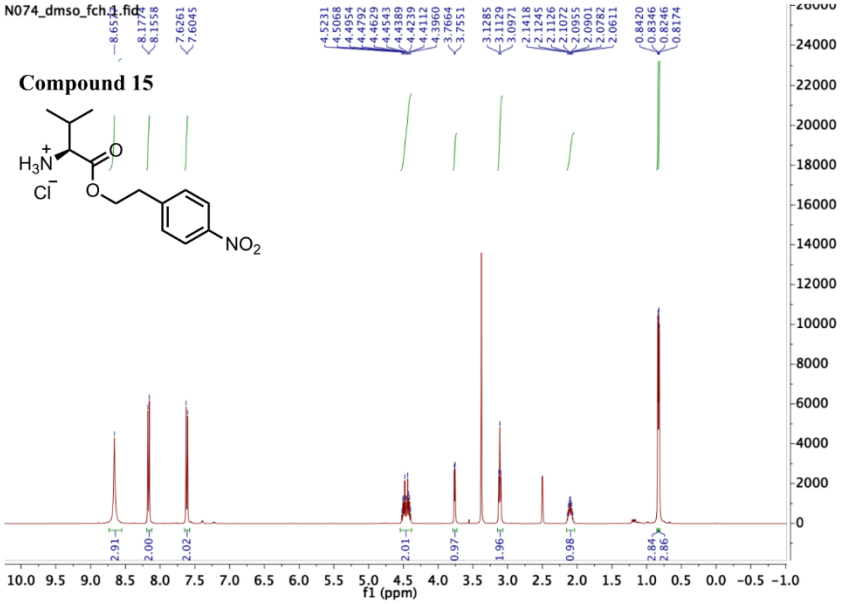


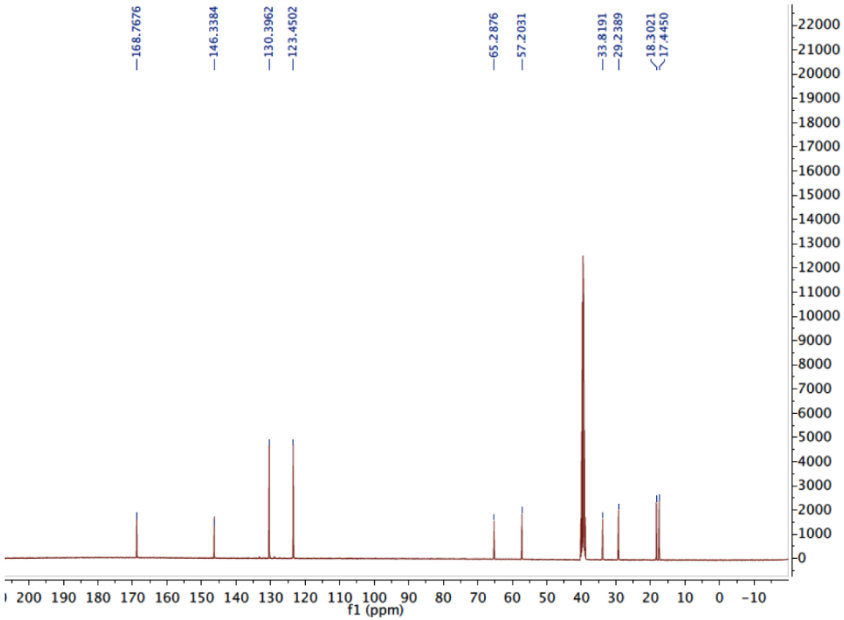

Compound 14a
BocHN
O
O
NO2
f1 (ppm)

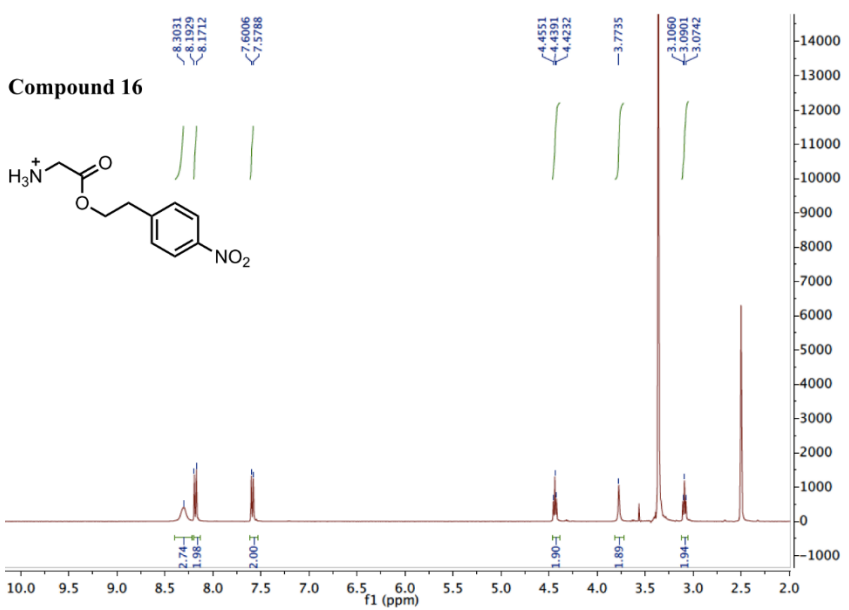

Compound 14b
BocHN
O
O
NO2
f1 (ppm)

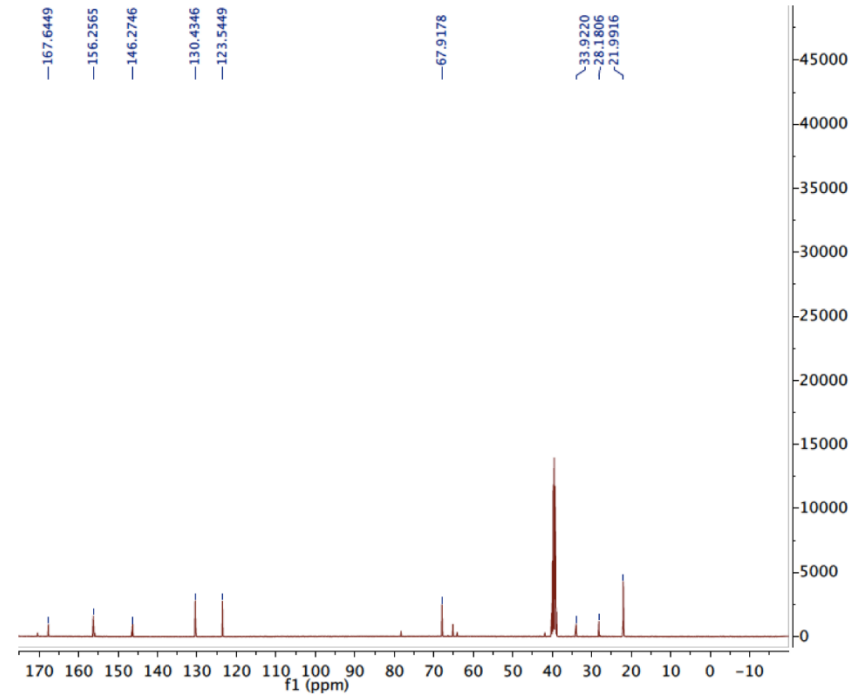

f1 (ppm)

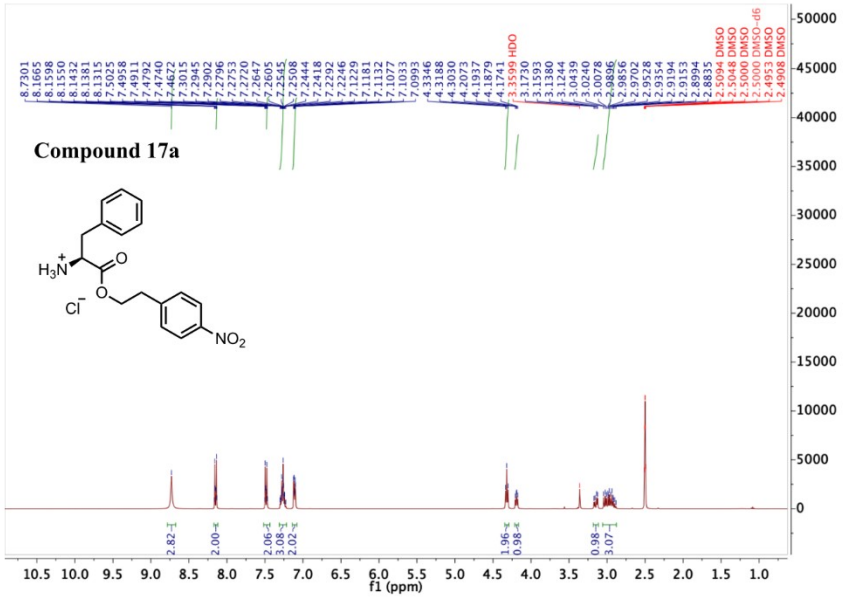

f1 (ppm)

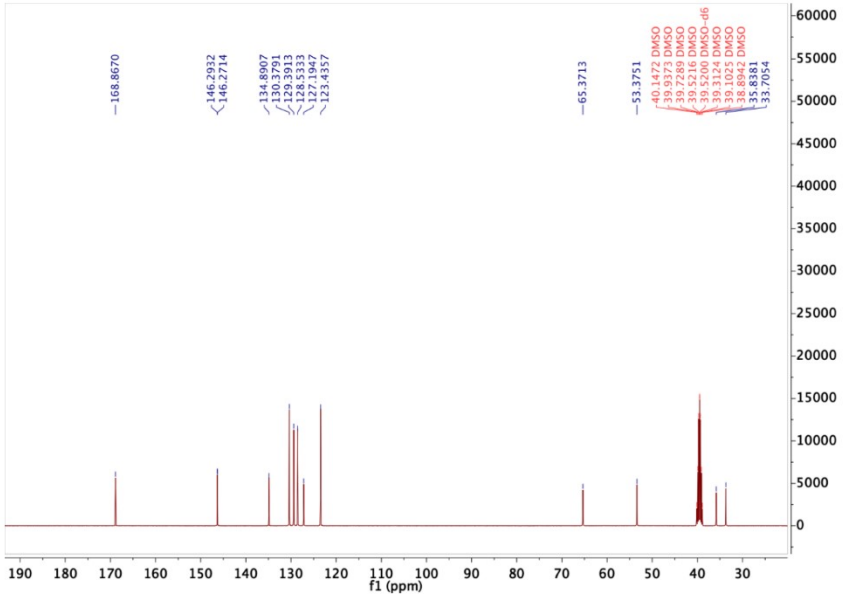

N074_dmso_fch_1.fid
Compound 15
f1 (ppm)

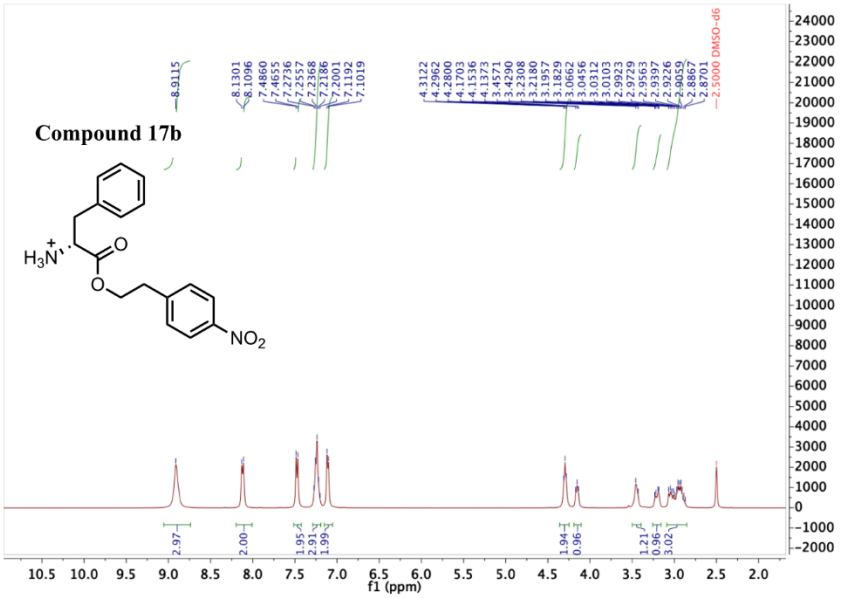

f1 (ppm)

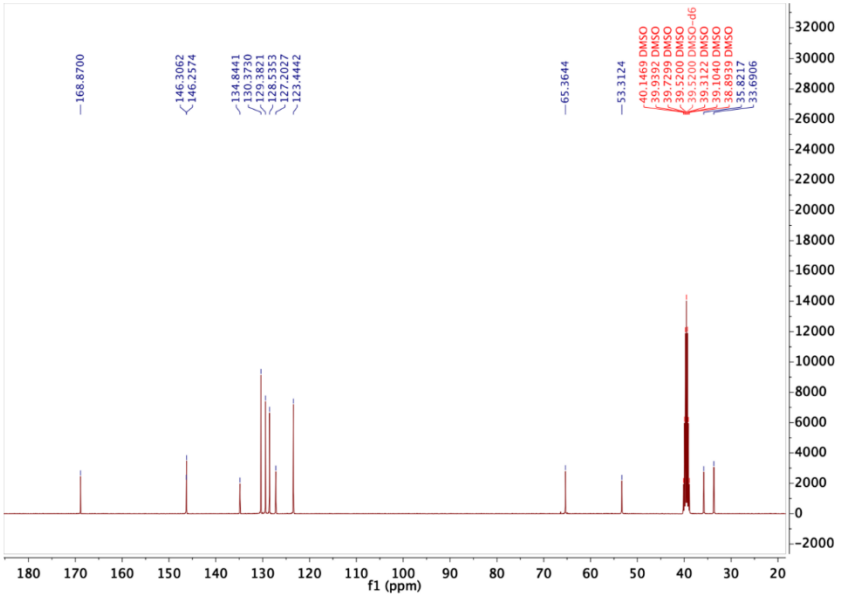

Compound 16
f1 (ppm)


f1 (ppm)

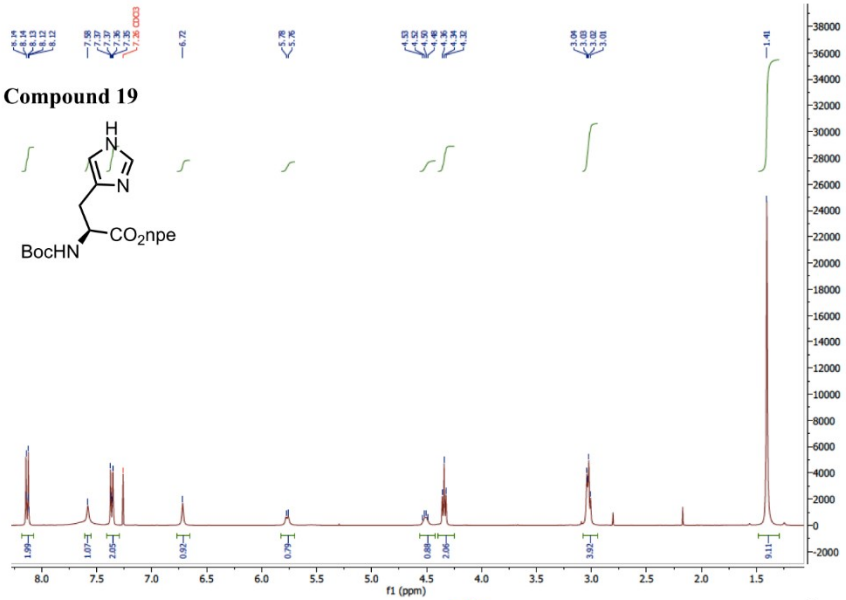

Compound 17a
f1 (ppm)

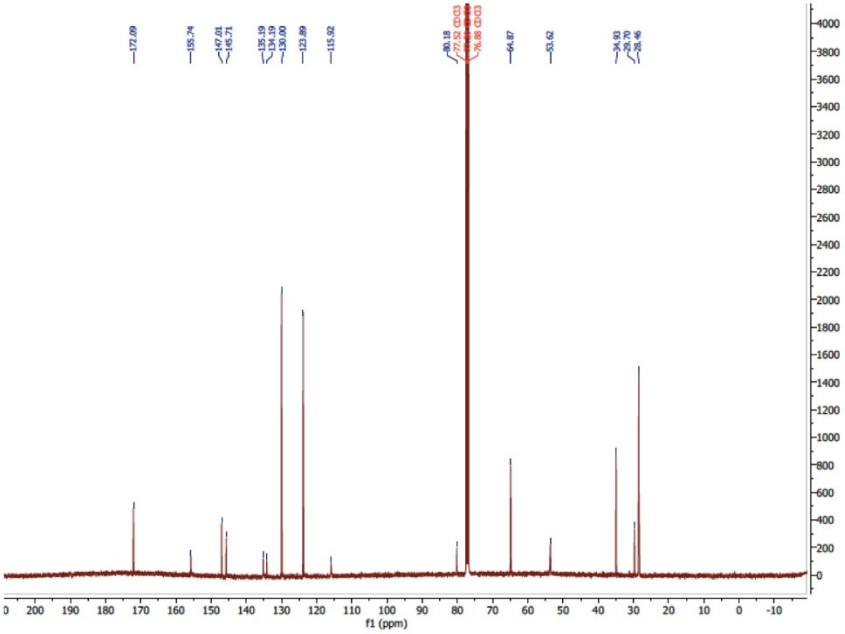

f1 (ppm)


Compound 17b
f1 (ppm)

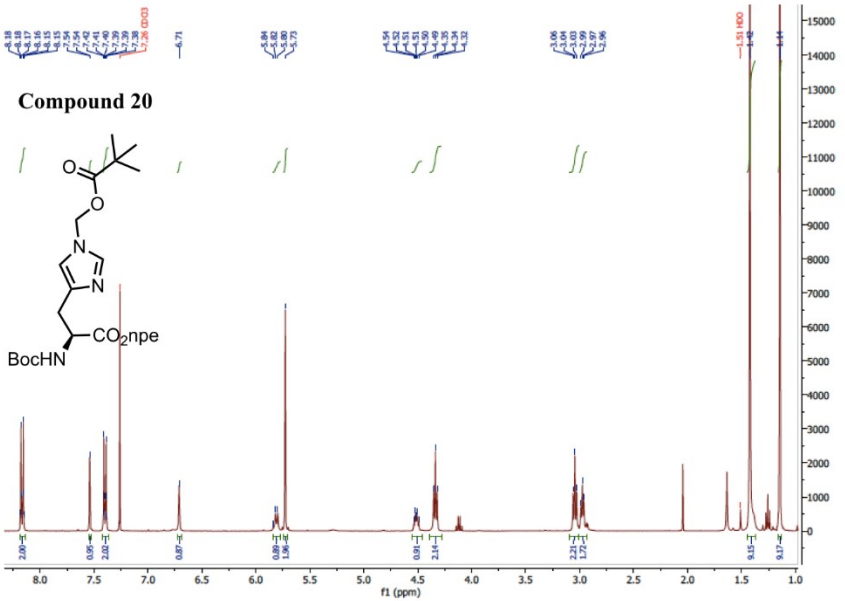

f1 (ppm)

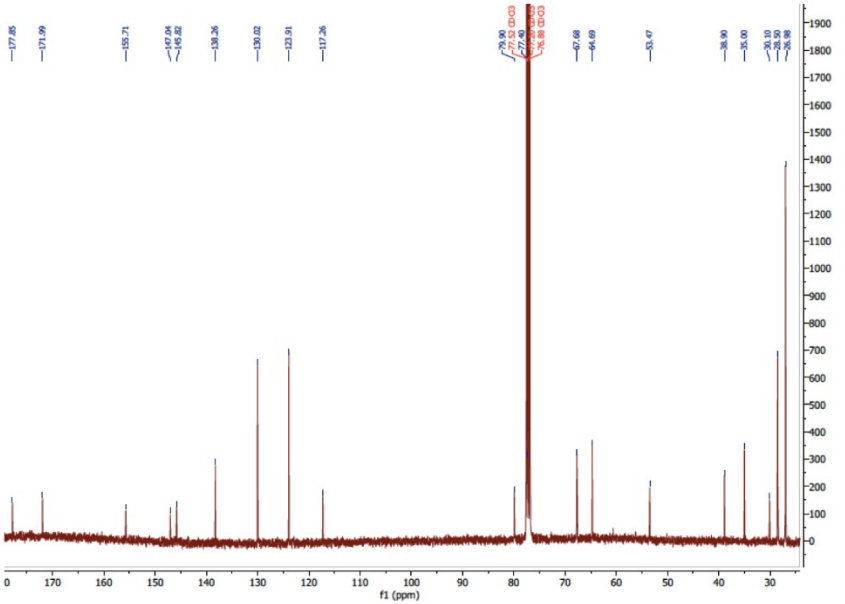
Compound 19



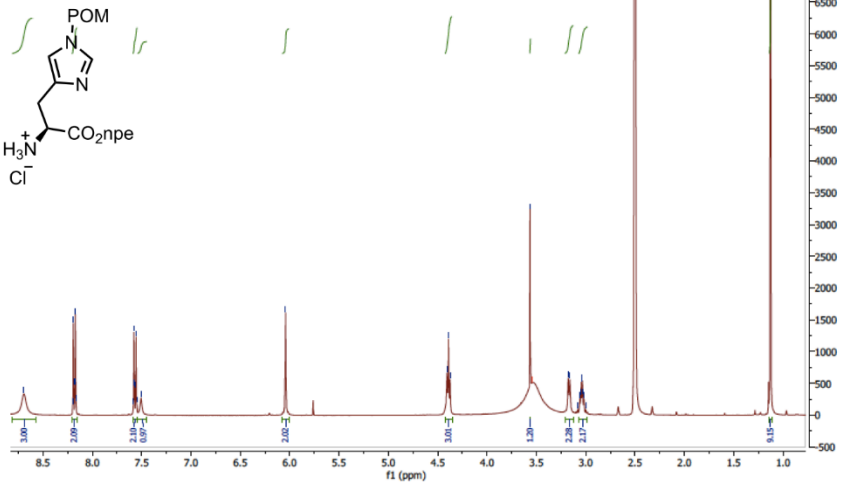


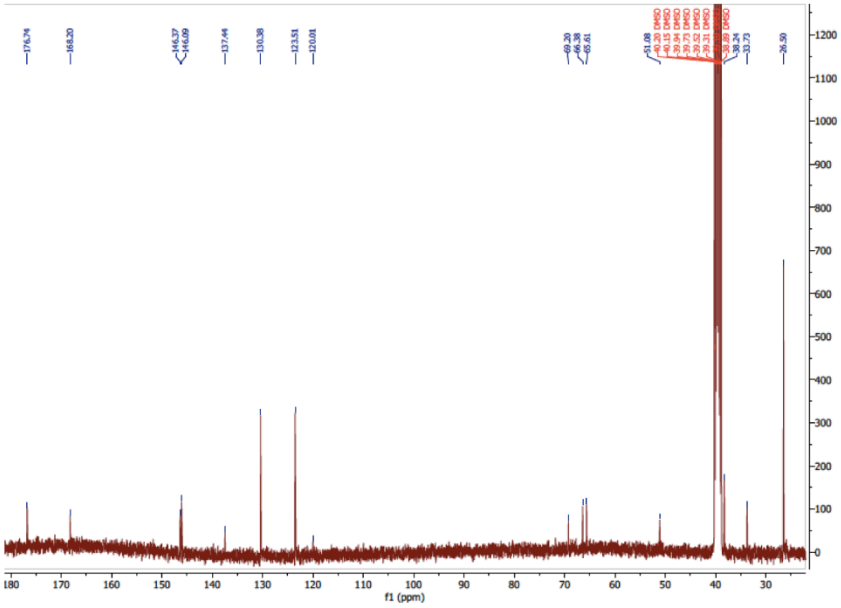
Compound 20



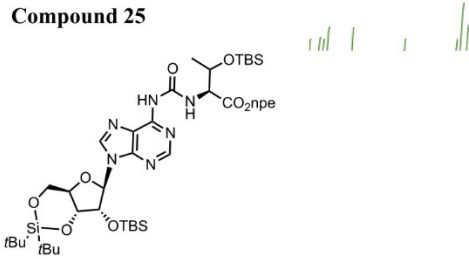


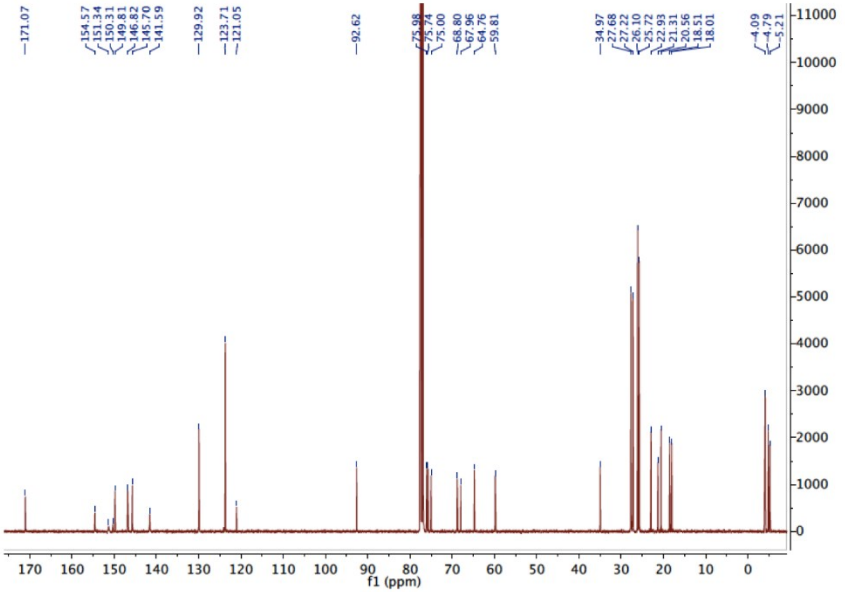
Compound 21



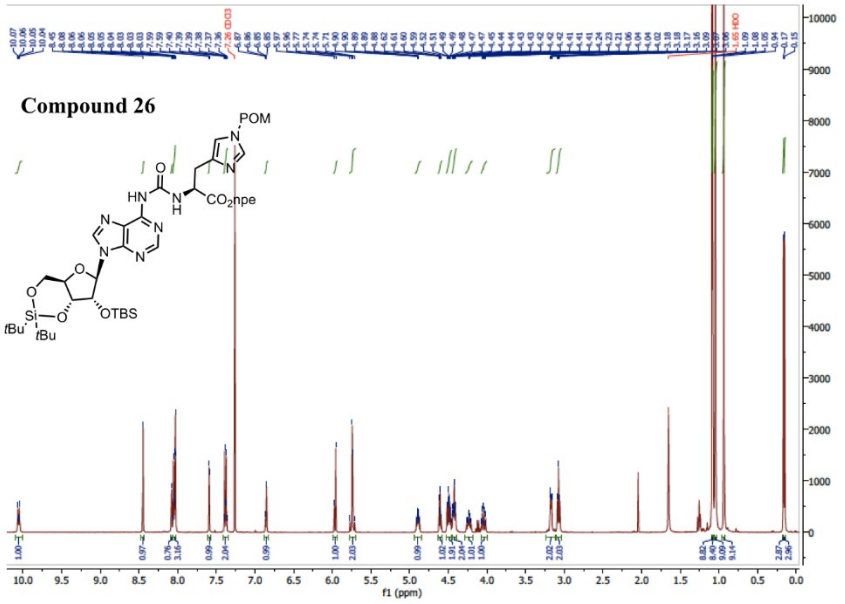


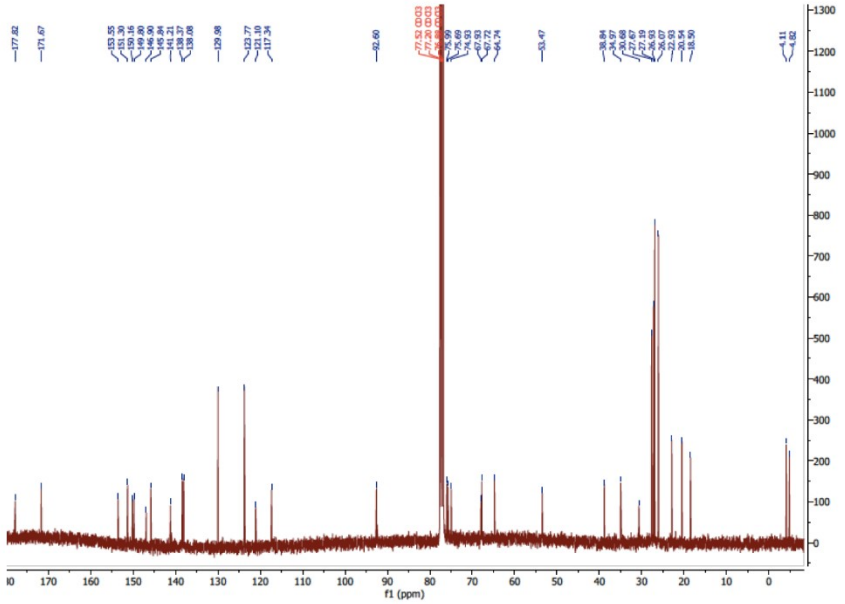
Compound 25



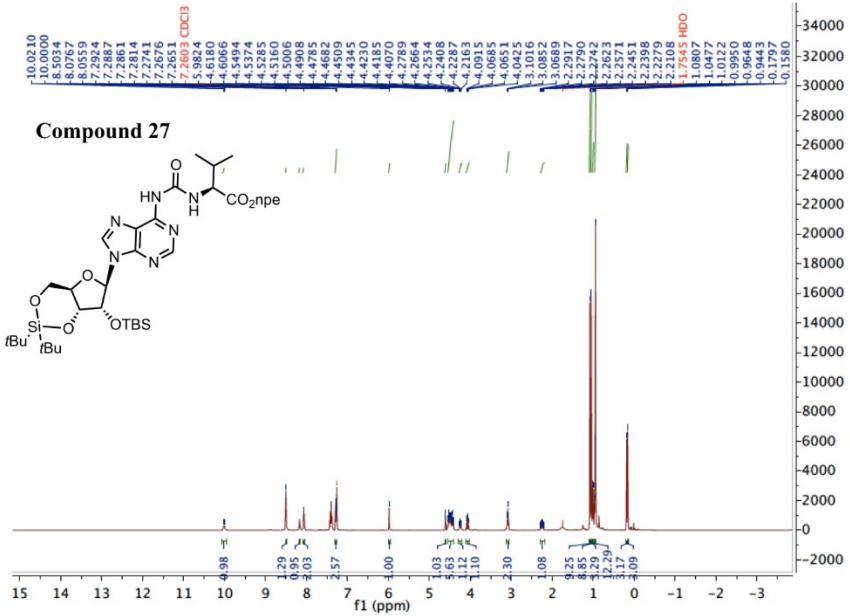


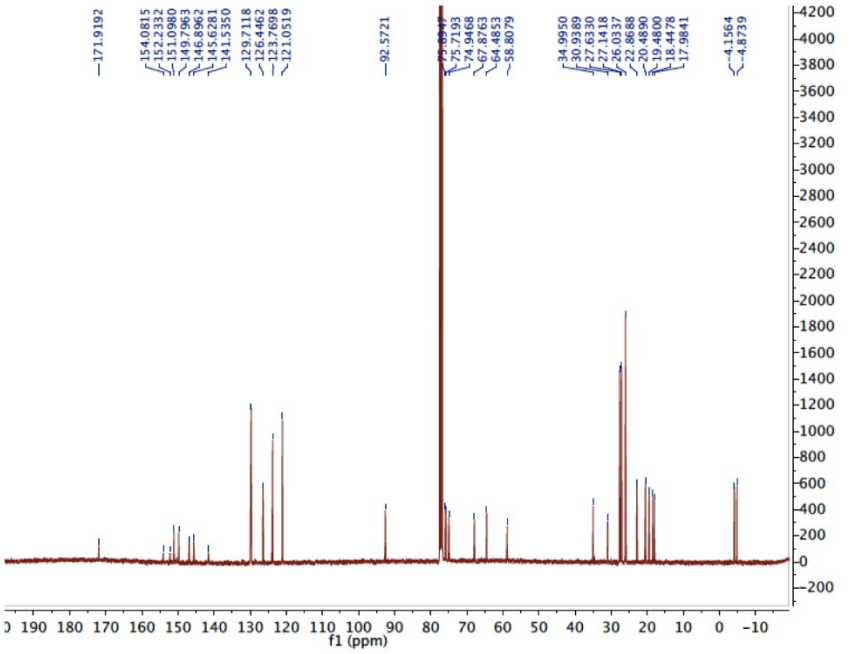
Compound 26



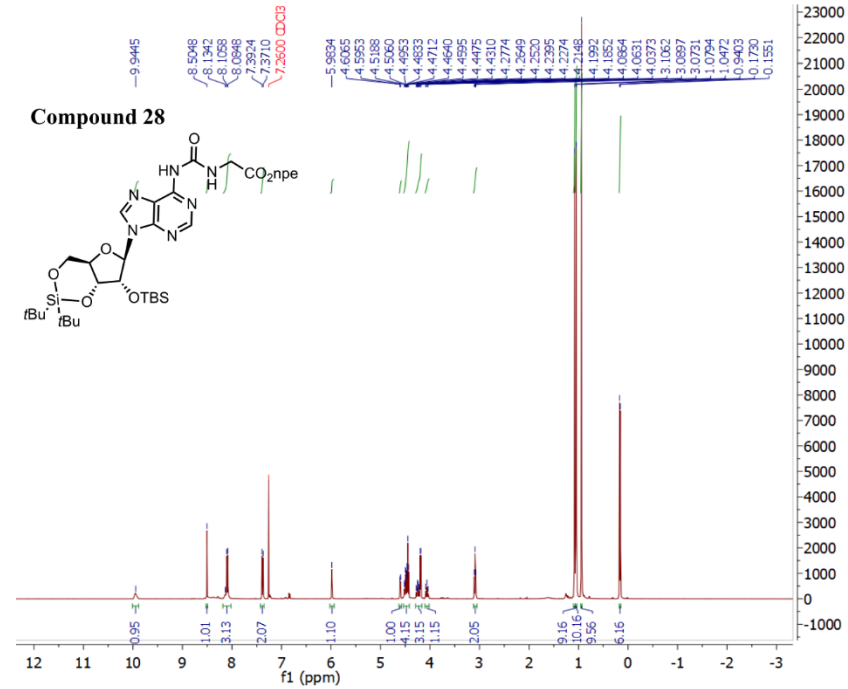


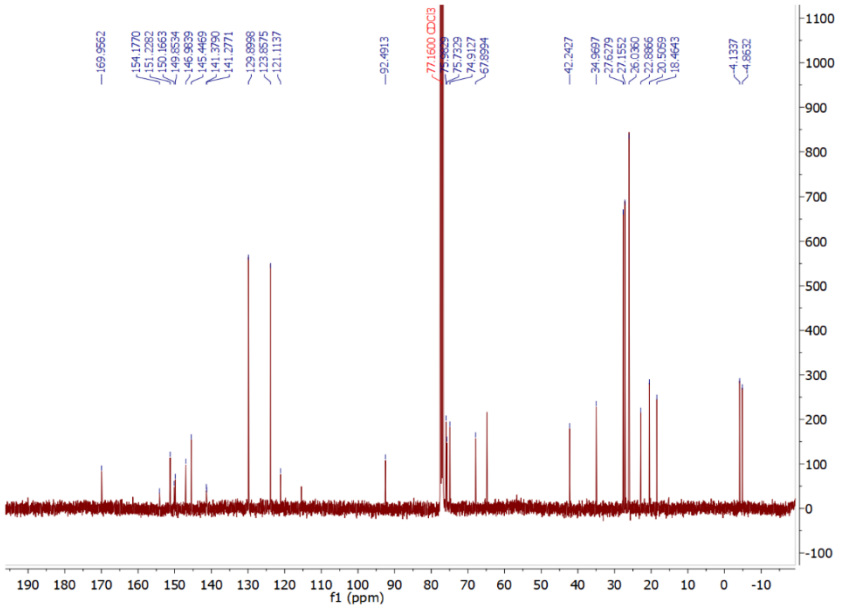
Compound 27



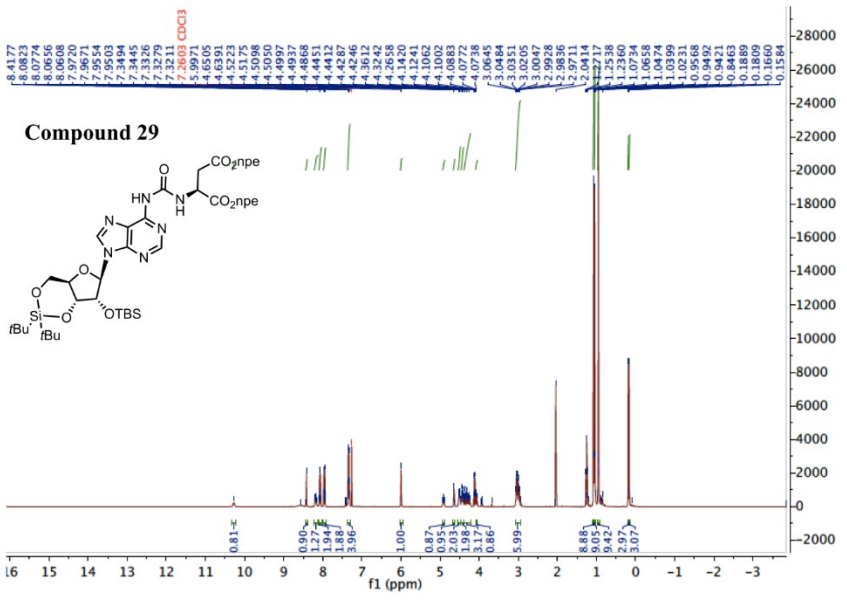


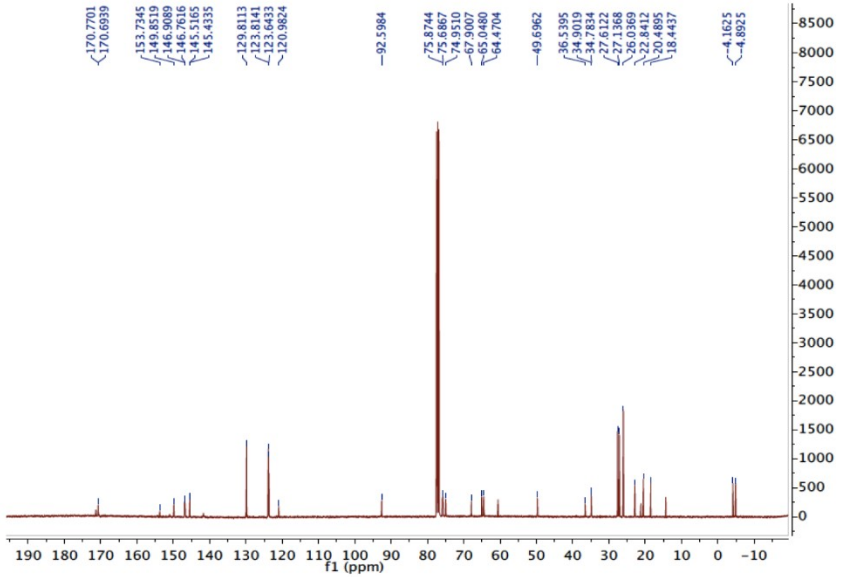
Compound 28

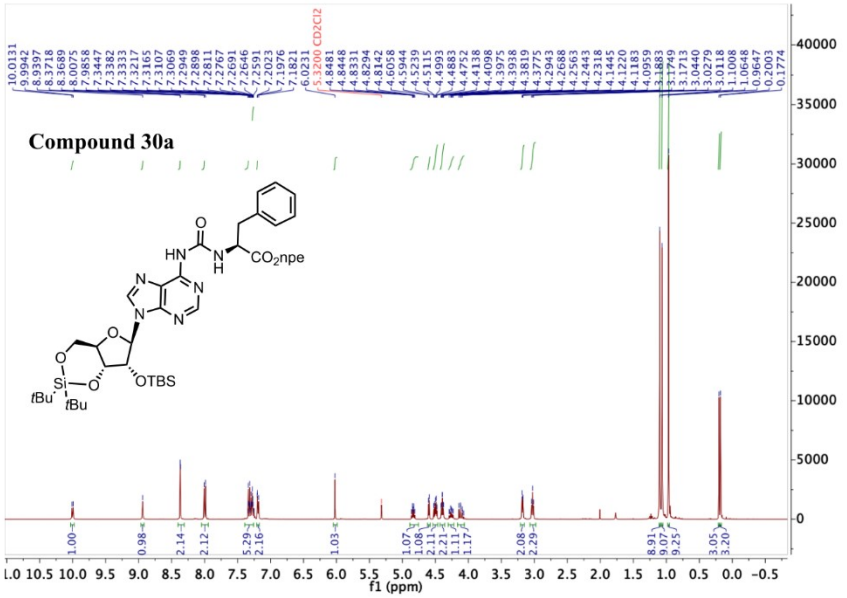


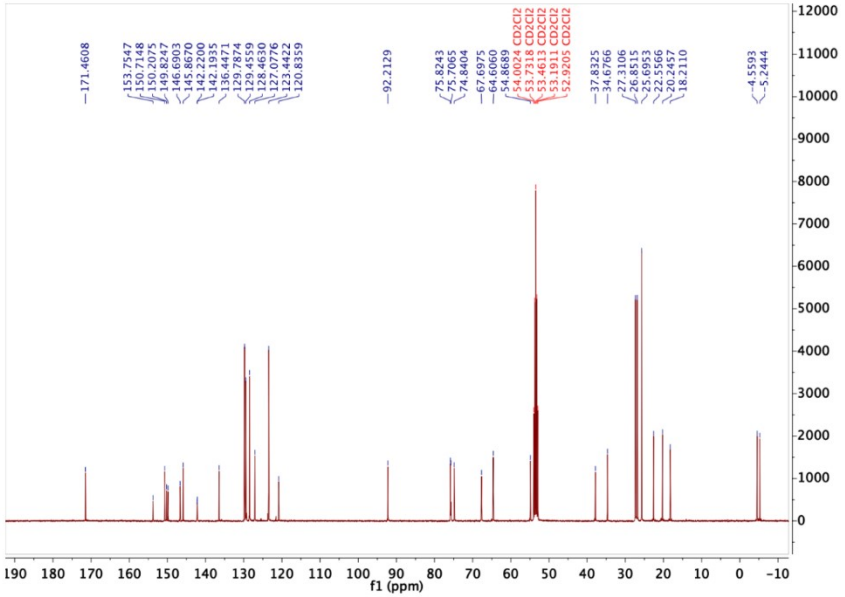


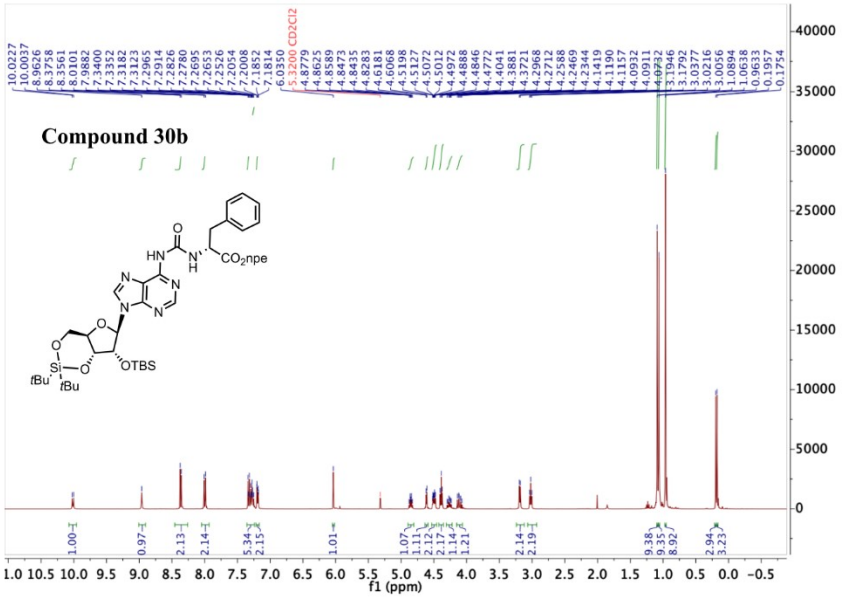
Compound 29

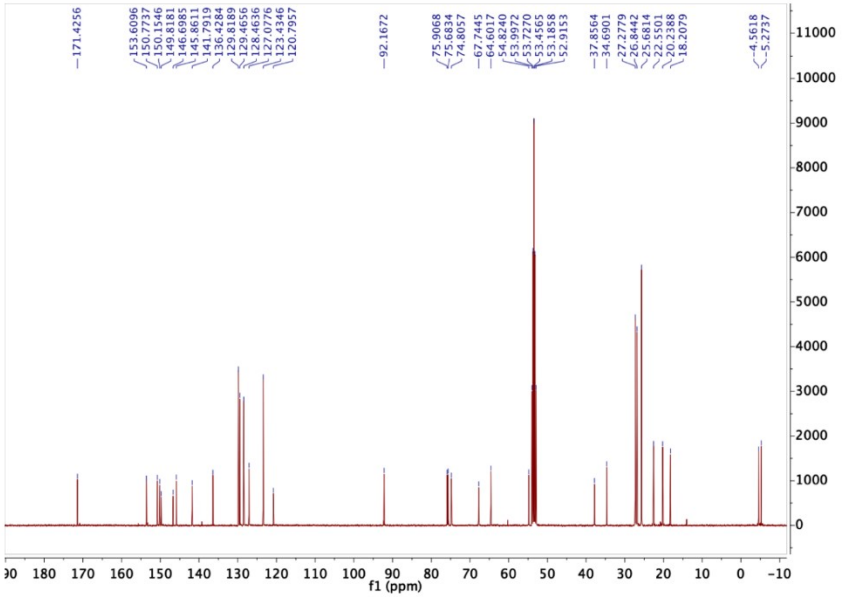


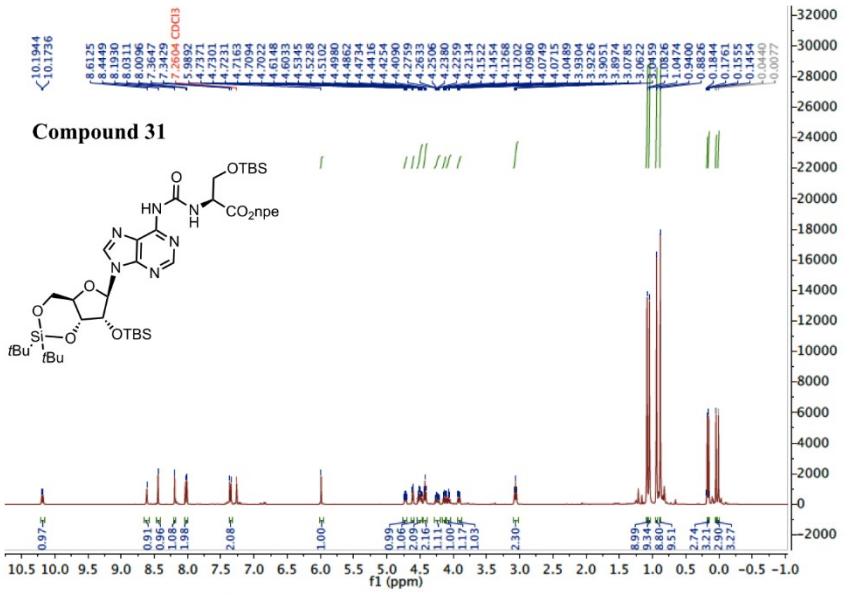


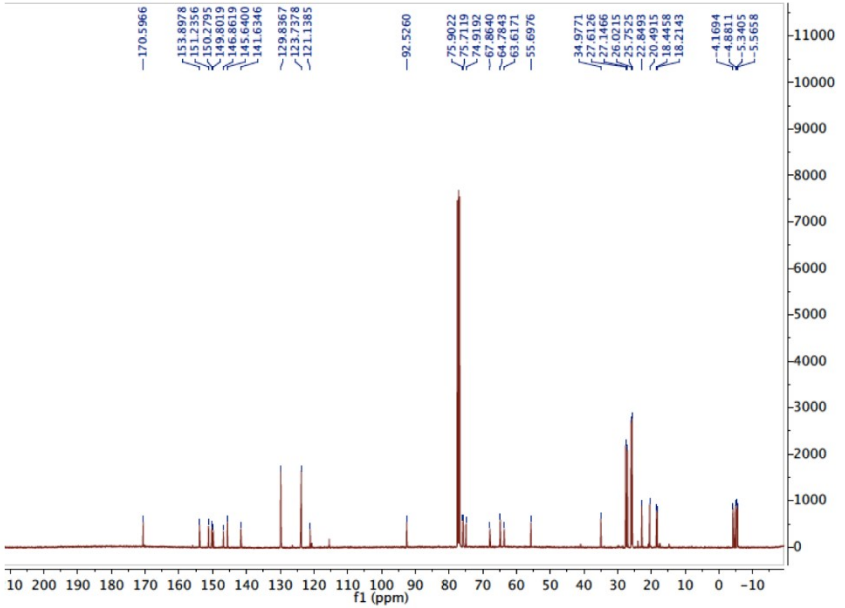

Compound 30a
CO2npe
OTBS
tBu
tBu
f1 (ppm)

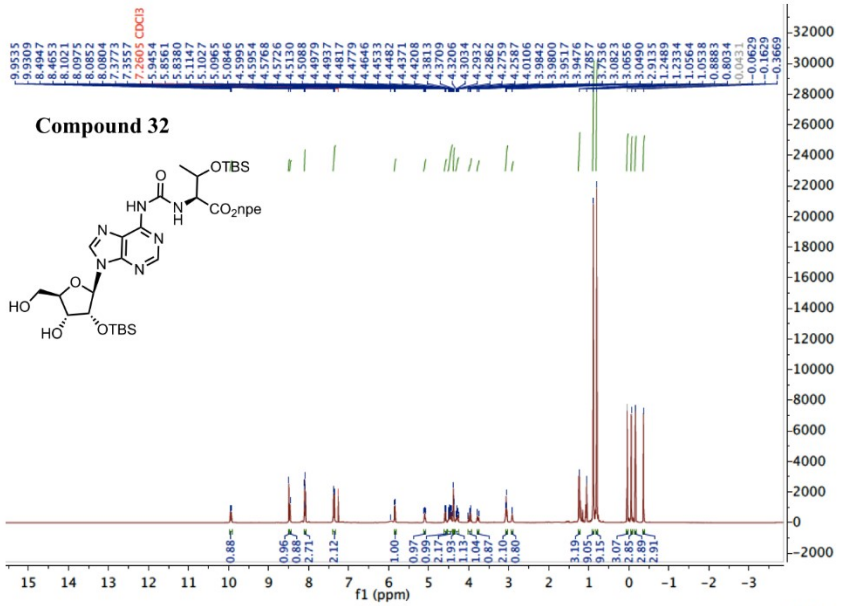

f1 (ppm)

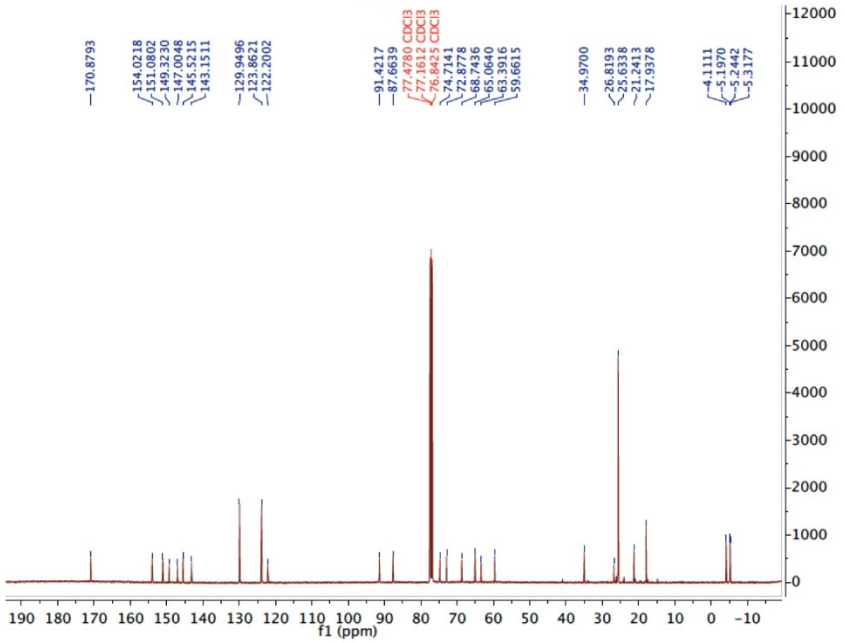

Compound 30b
CO2npe
OTBS
tBu
tBu
f1 (ppm)


f1 (ppm)

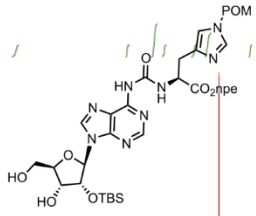

Compound 31
OTBS
HN
N
H
CO2npe
O
N
N
N
N
O
O
Si
O
OTBS
tBu
tBu
f1 (ppm)

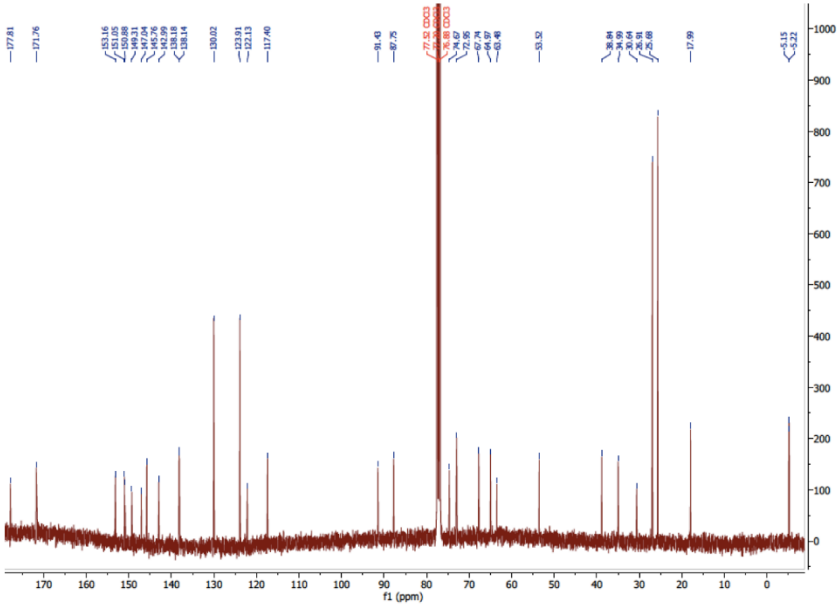

f1 (ppm)


Compound 32
OTBS
HN
N
H
CO2npe
O
N
N
N
N
HO
HO
OTBS
f1 (ppm)

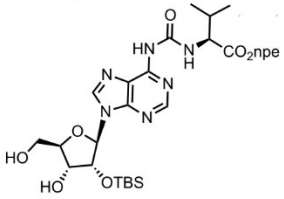

f1 (ppm)

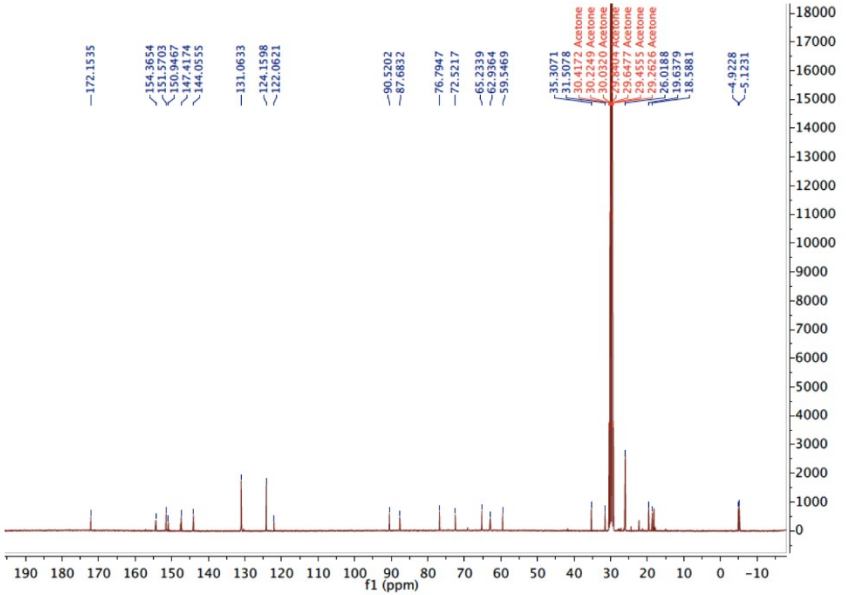
Compound 33

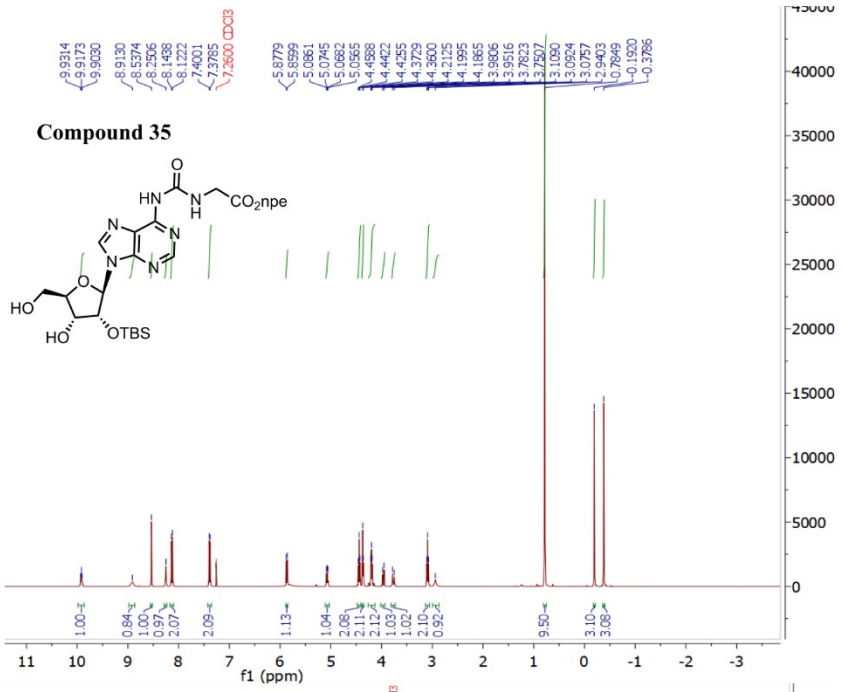


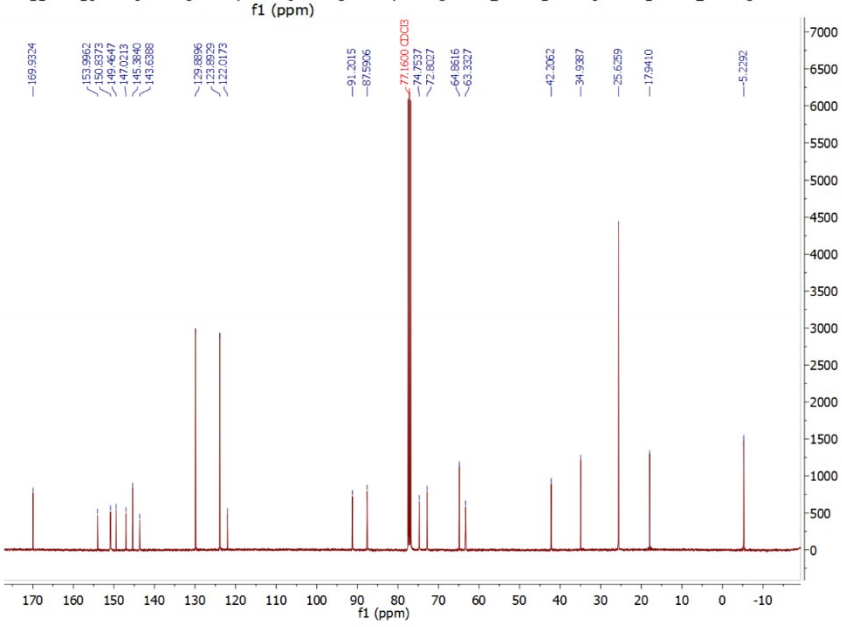


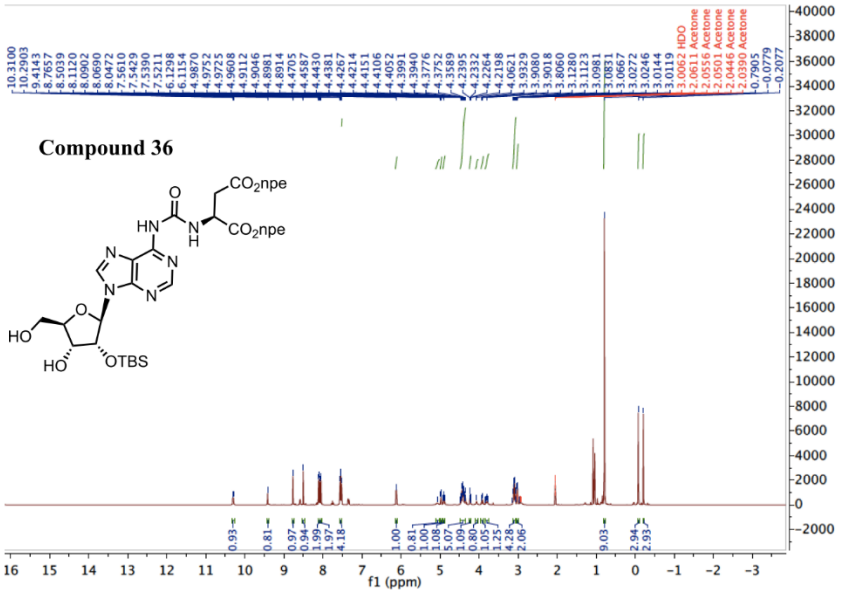
Compound 34

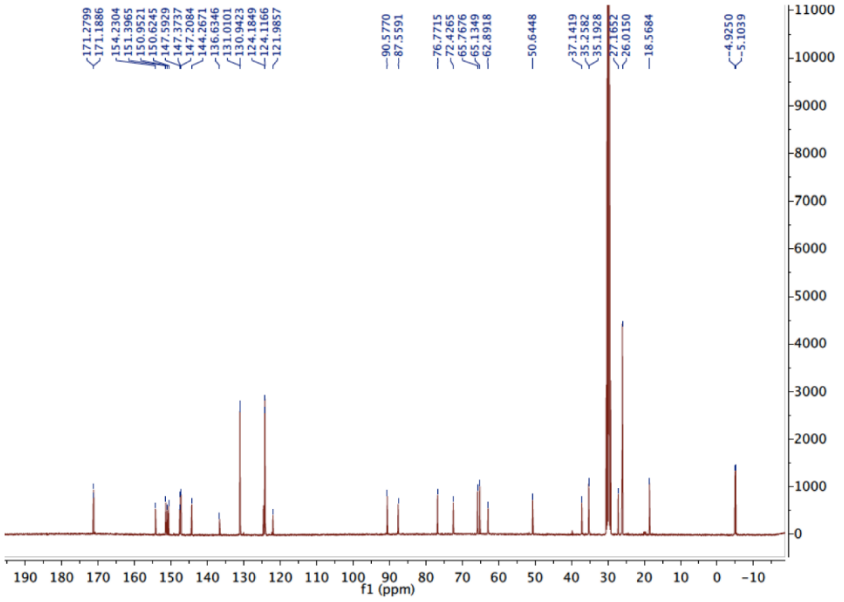


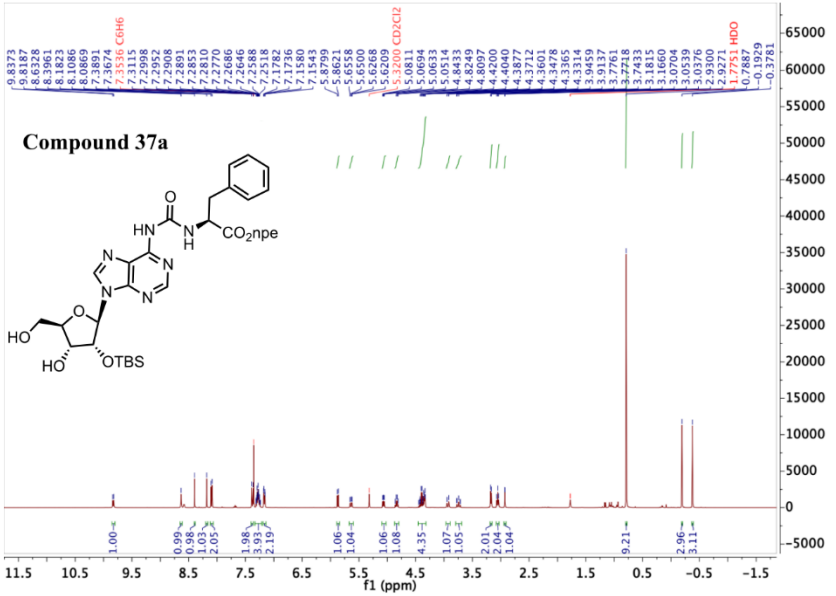


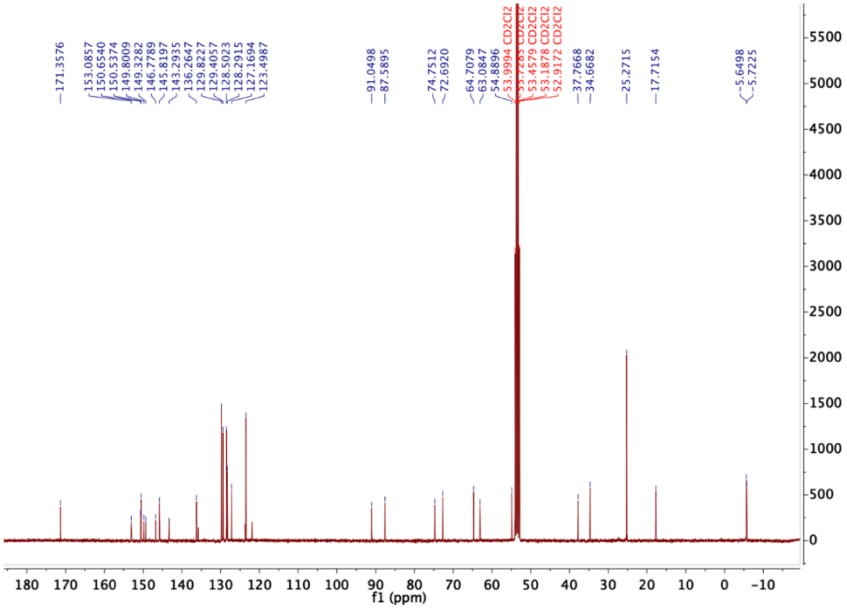

Compound 35
HO
HO
OTBS
HN
N
H
CO2npe
f1 (ppm)

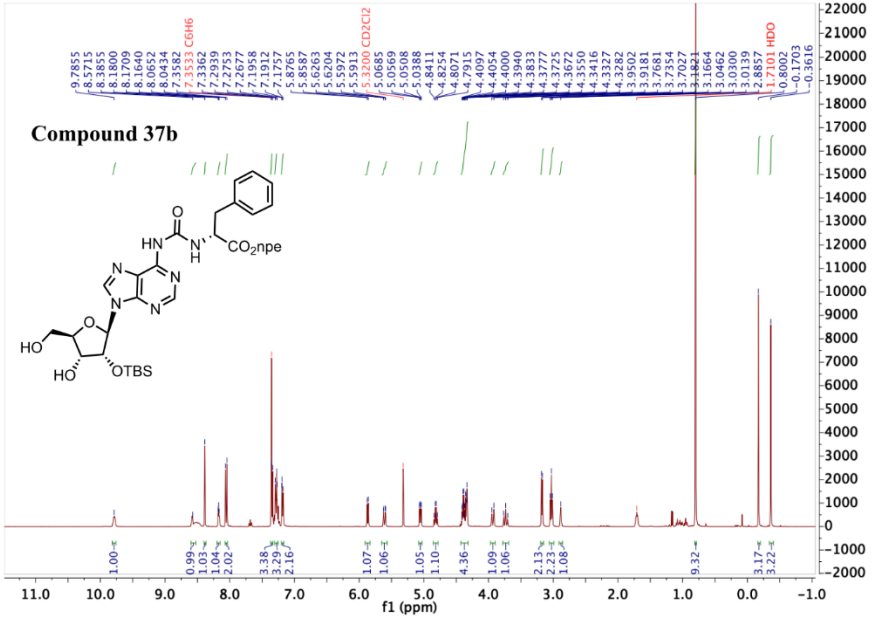

f1 (ppm)

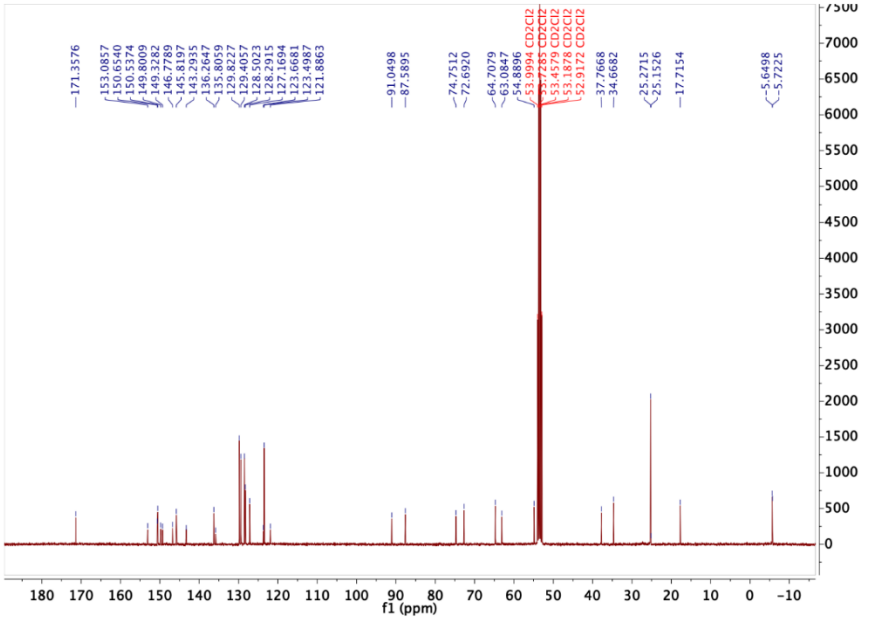

Compound 36
HO
HO
OTBS
HN
N
H
CO2npe
CO2npe
f1 (ppm)

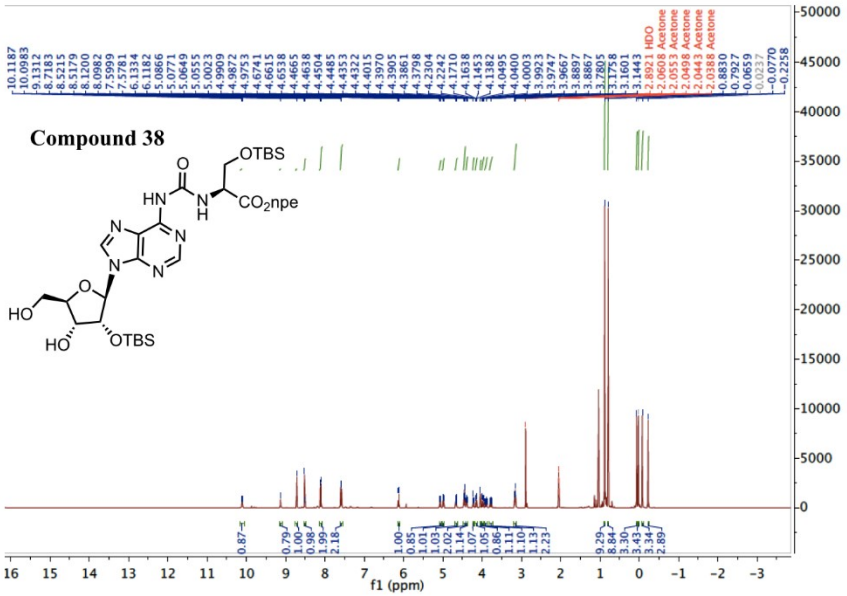

f1 (ppm)

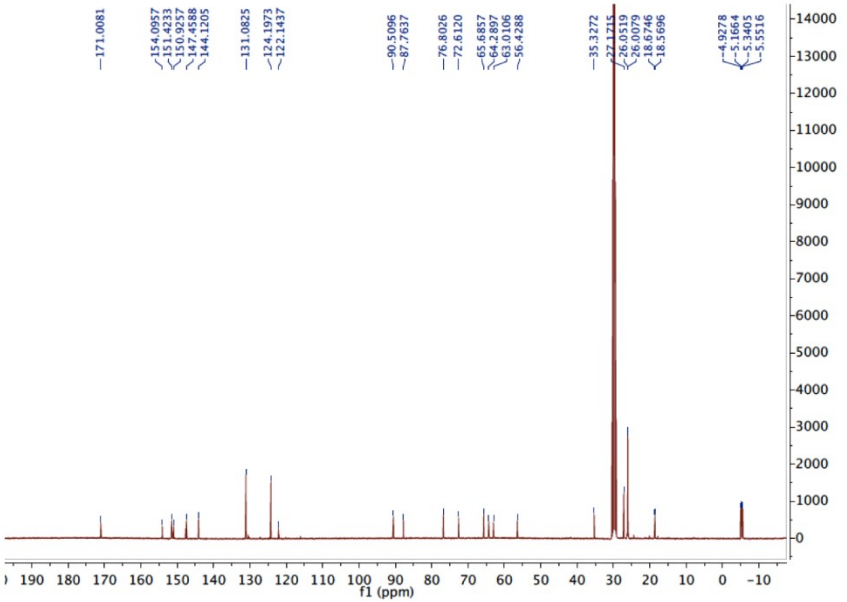

Compound 37a
HO
HO
OTBS
HN
N
H
CO2npe
f1 (ppm)

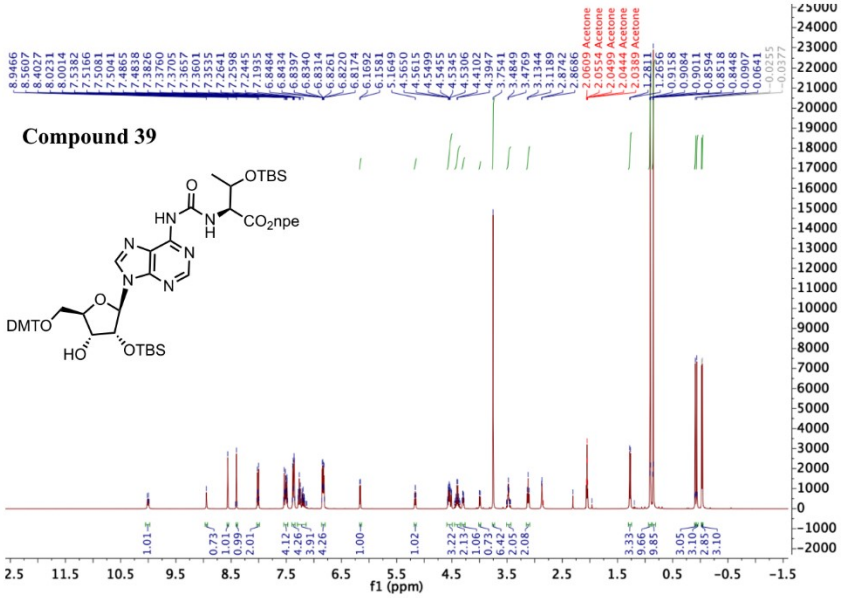

f1 (ppm)

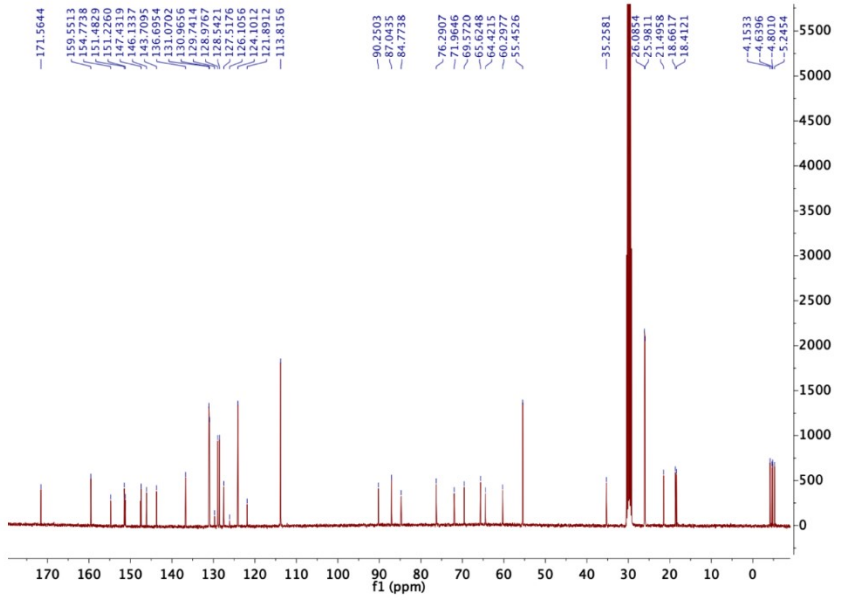

Compound 37b
HO
HO
OTBS
HN
N
H
CO2npe
f1 (ppm)

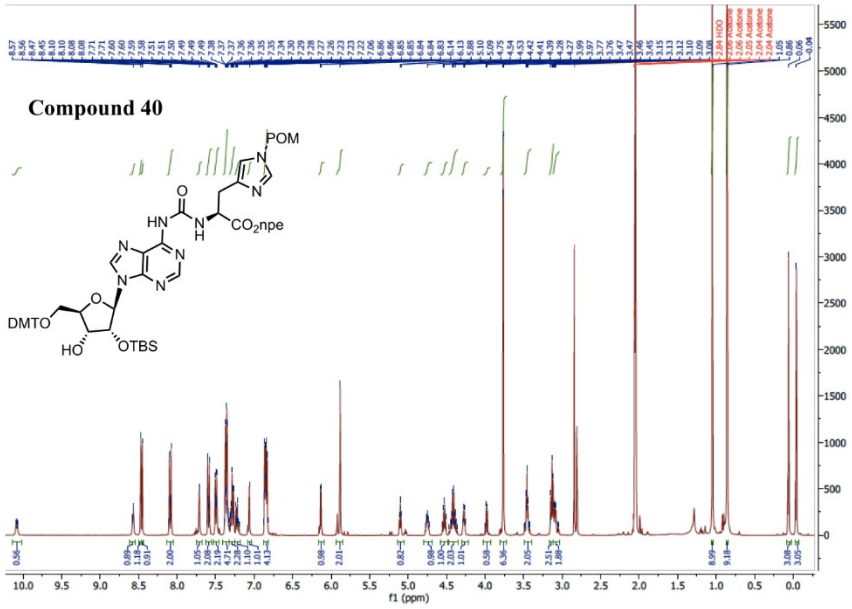

f1 (ppm)

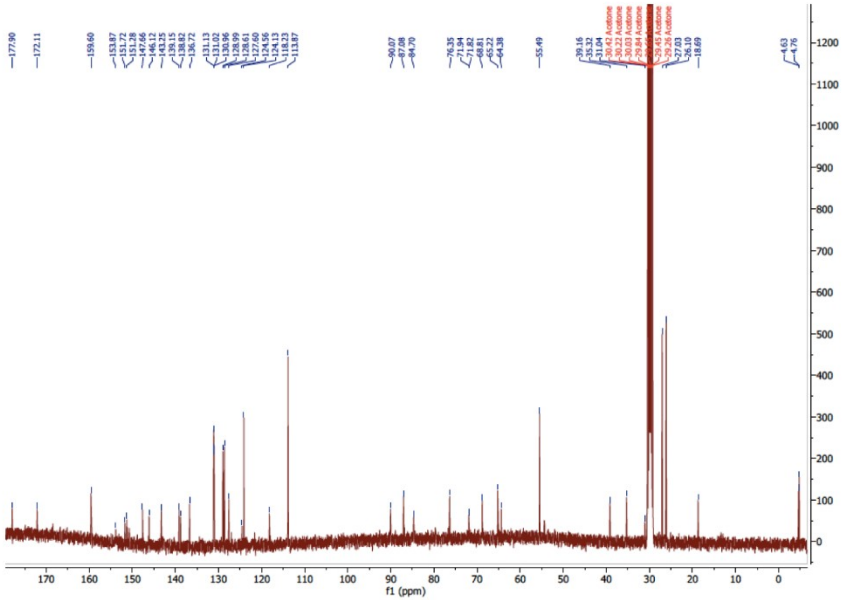

Compound 38
OTBS
HN
N
H
O
CO2npe
N
N
N
N
O
HO
HO
OTBS
f1 (ppm)

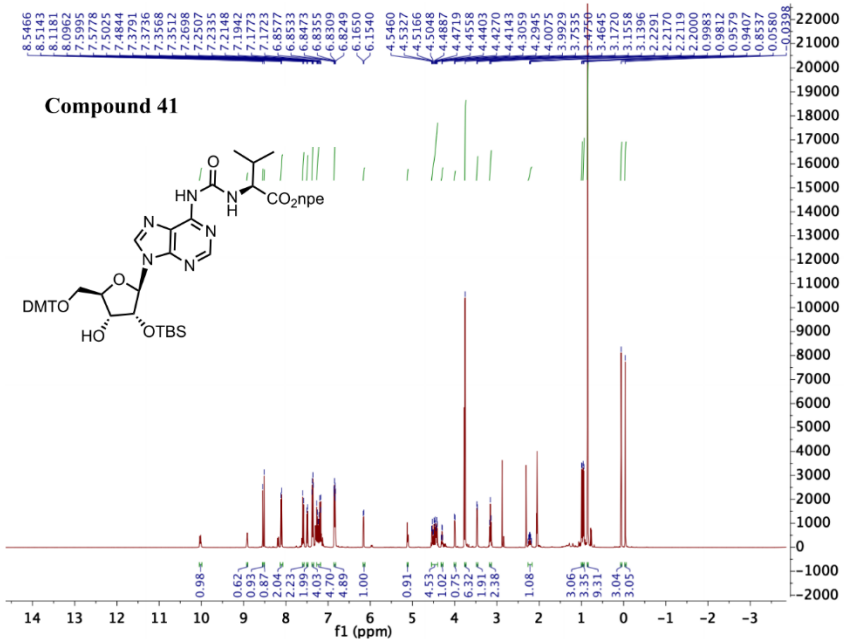

f1 (ppm)

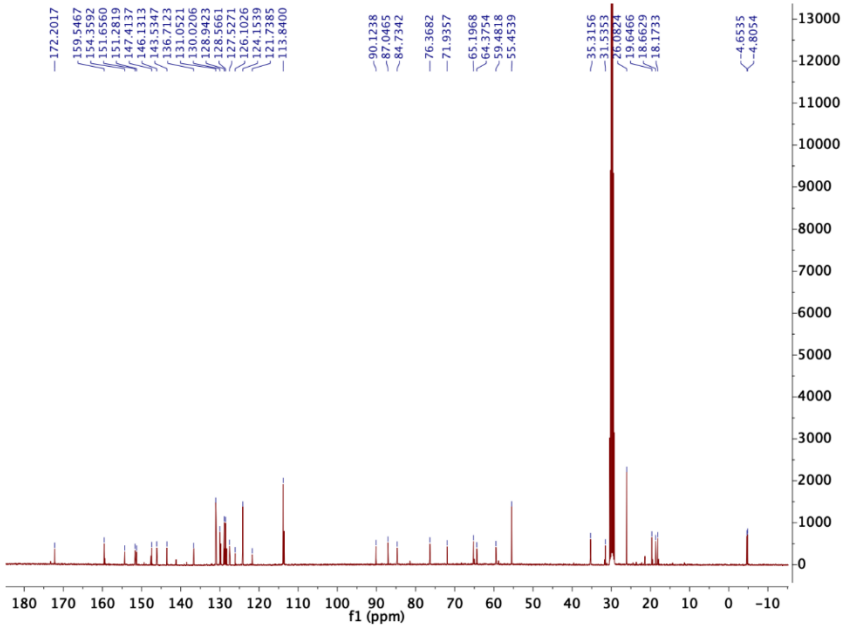

Compound 39
OTBS
HN
N
H
O
CO2npe
N
N
N
N
O
DMTO
HO
OTBS
f1 (ppm)

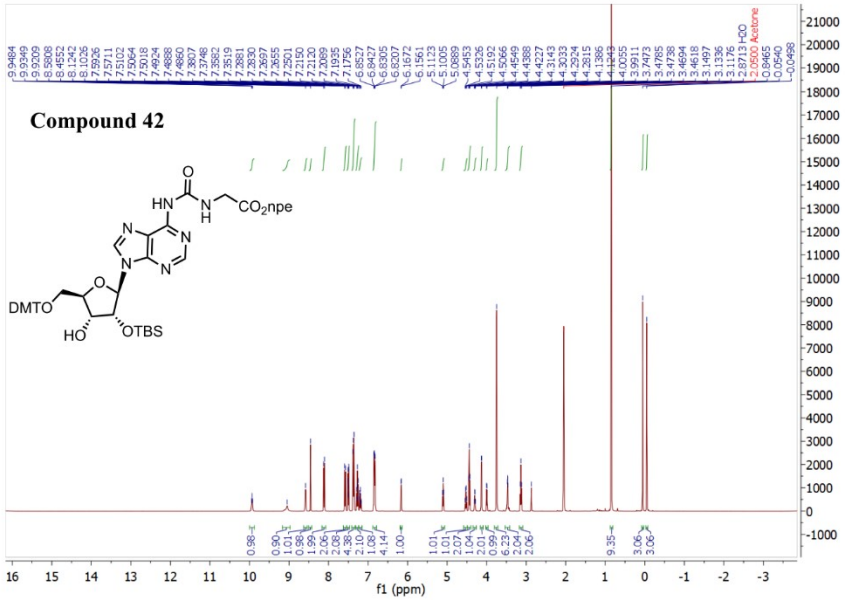

f1 (ppm)

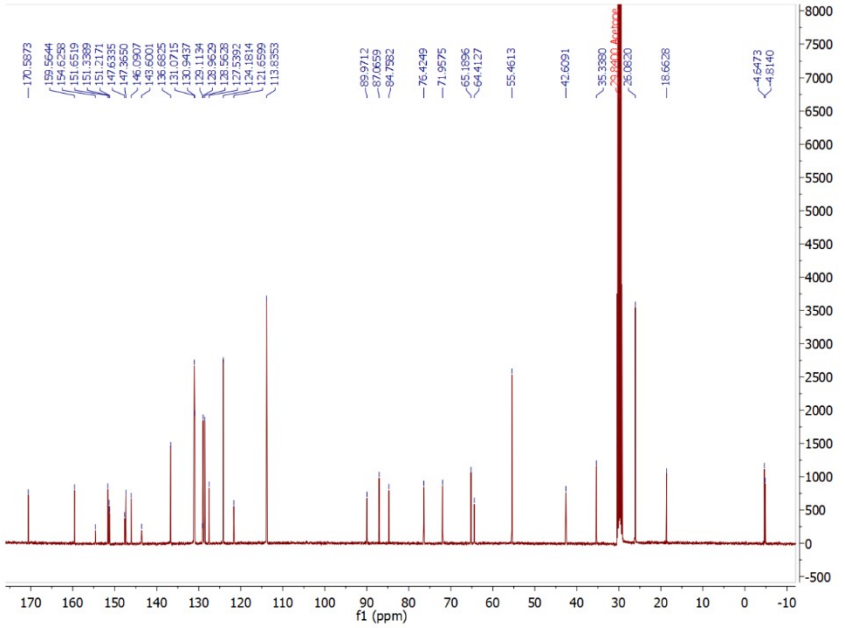

Compound 40
POM
HN
O
N
H
CO2npe
DMTO
HO
OTBS
f1 (ppm)

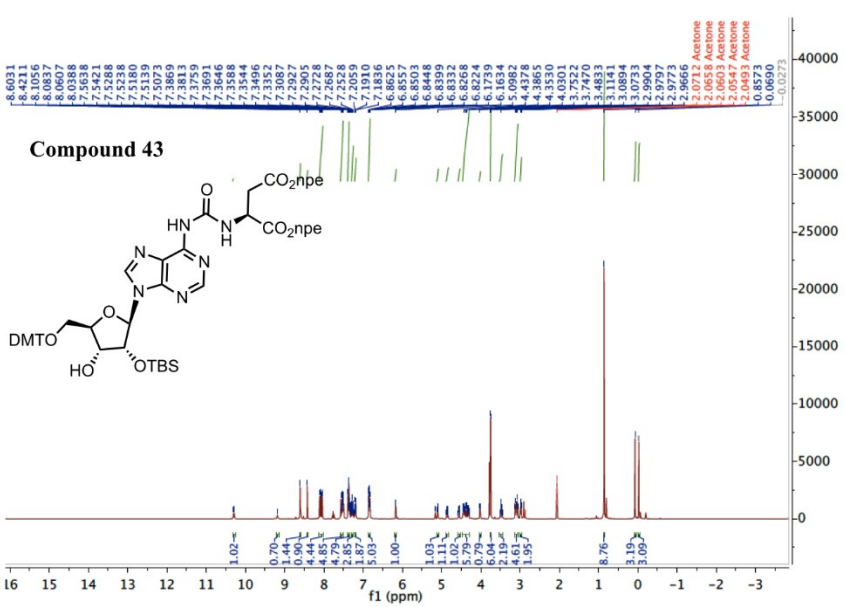

f1 (ppm)

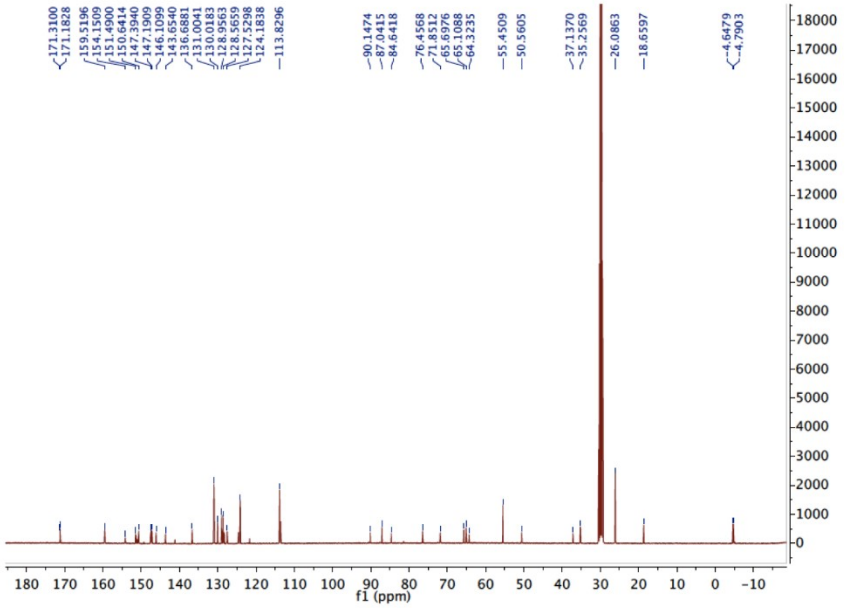

Compound 41
HN
O
N
H
CO2npe
DMTO
HO
OTBS
f1 (ppm)

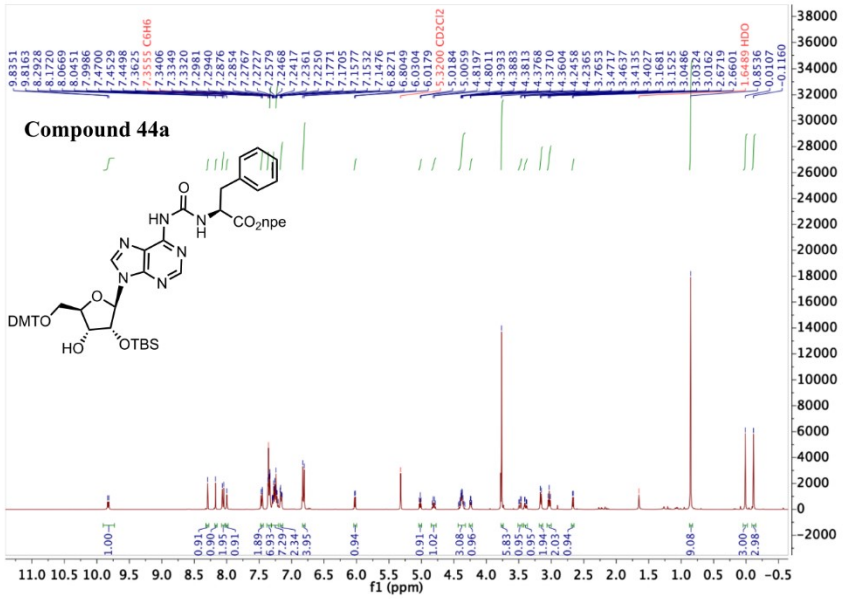

f1 (ppm)

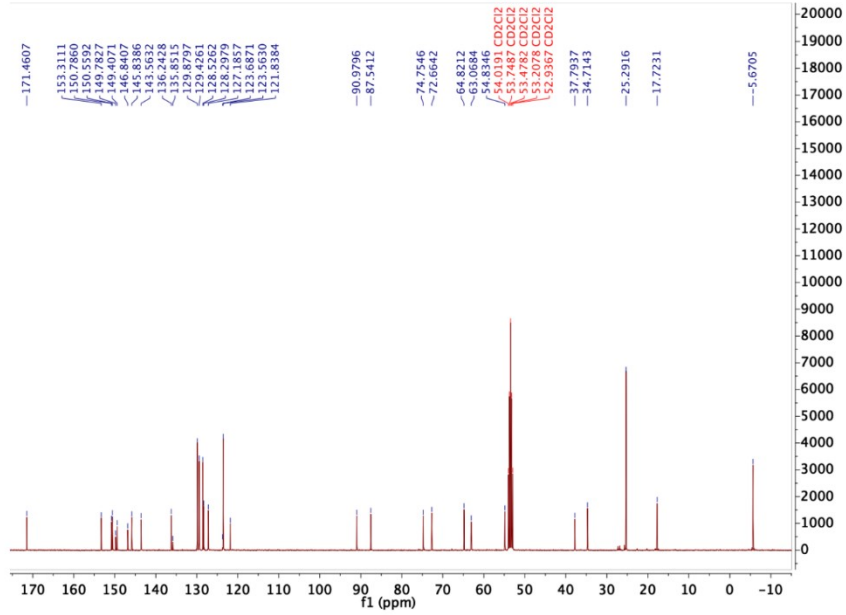

Compound 42
DMTO
HO
OTBS
HN
N
H
CO2npe

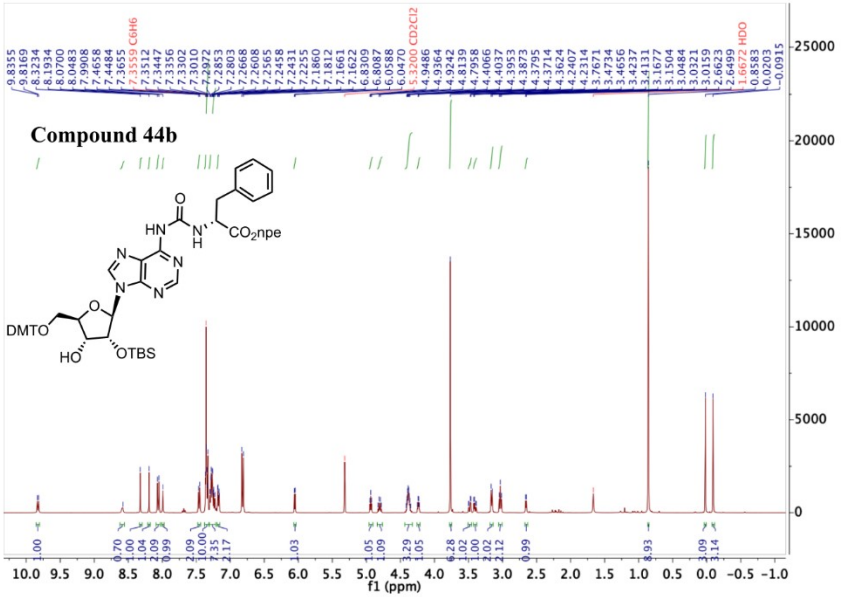


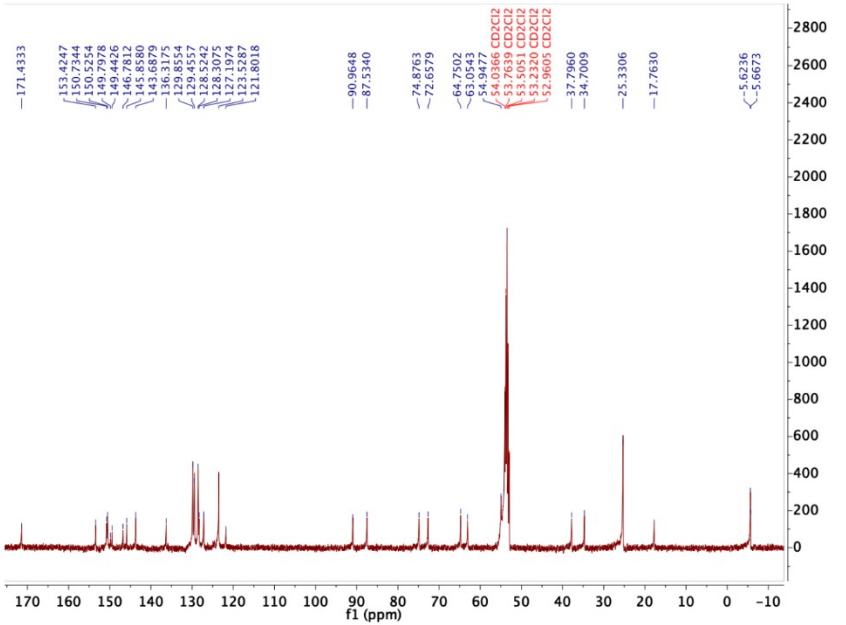

Compound 43
DMTO
HO
OTBS
HN
N
H
CO2npe
CO2npe



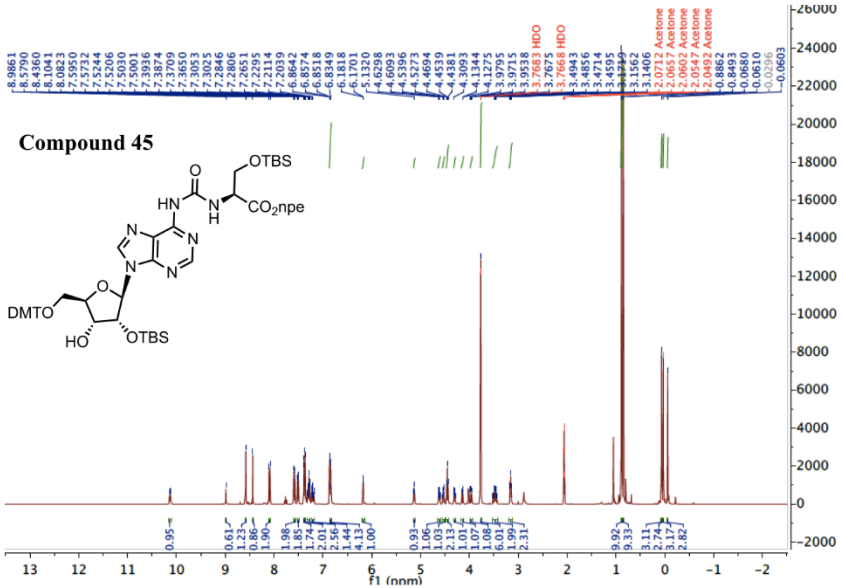

Compound 44a
DMTO
HO
OTBS
HN
N
H
CO2npe
f1 (ppm)

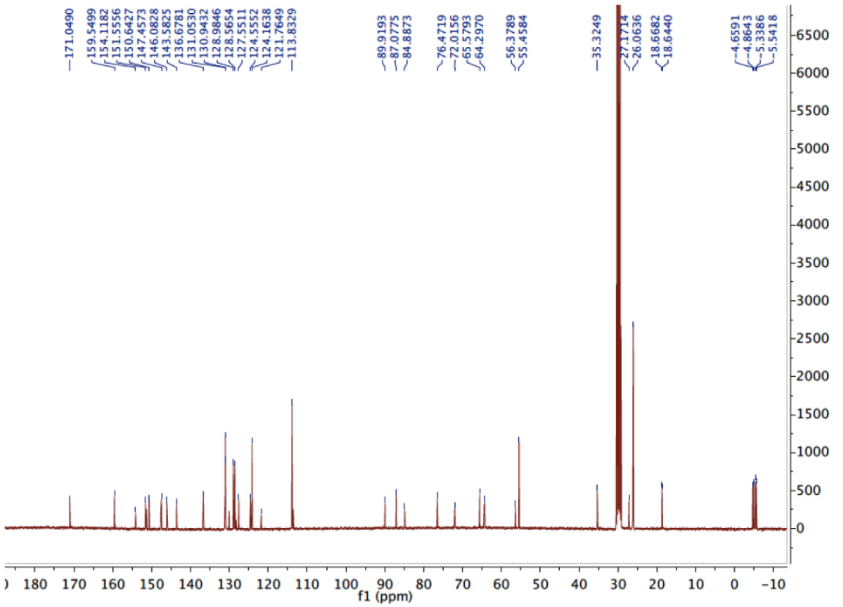

f1 (ppm)


Compound 44b
DMTO
HO
OTBS
HN
N
H
CO2npe
f1 (ppm)

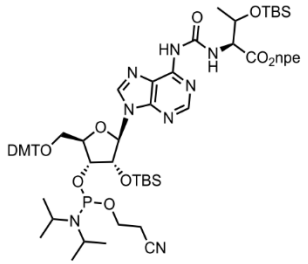

f1 (ppm)

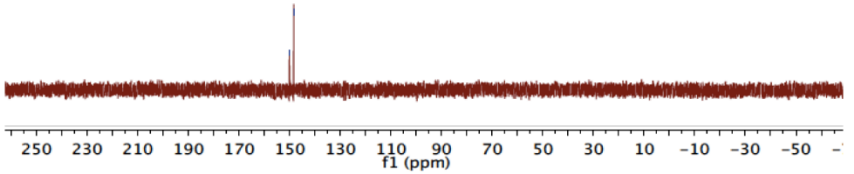
Compound 45



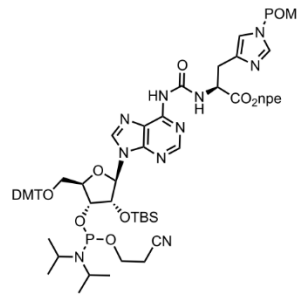


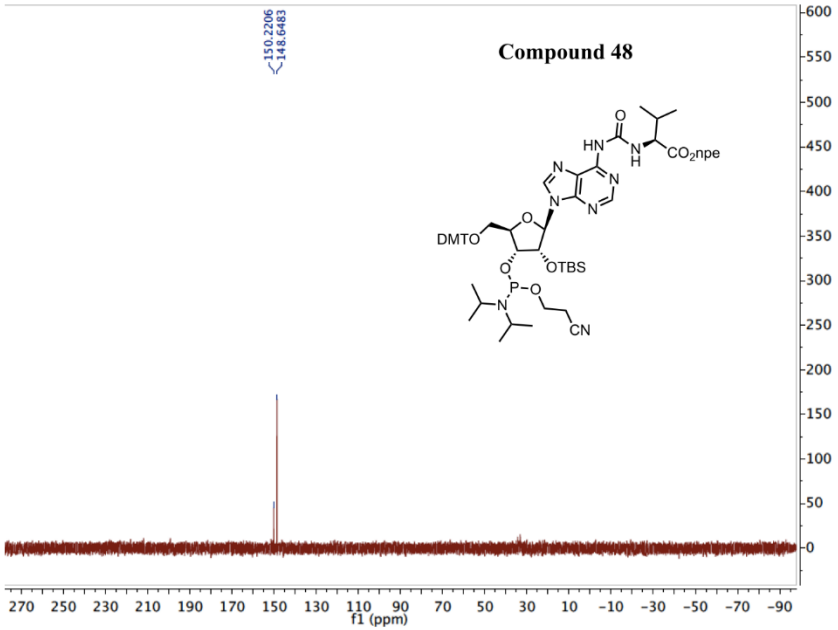
Compound 46

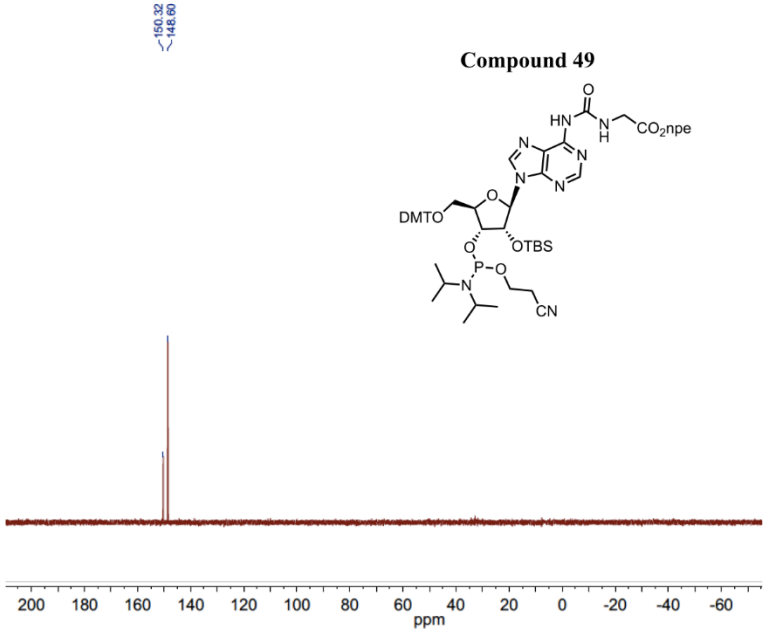


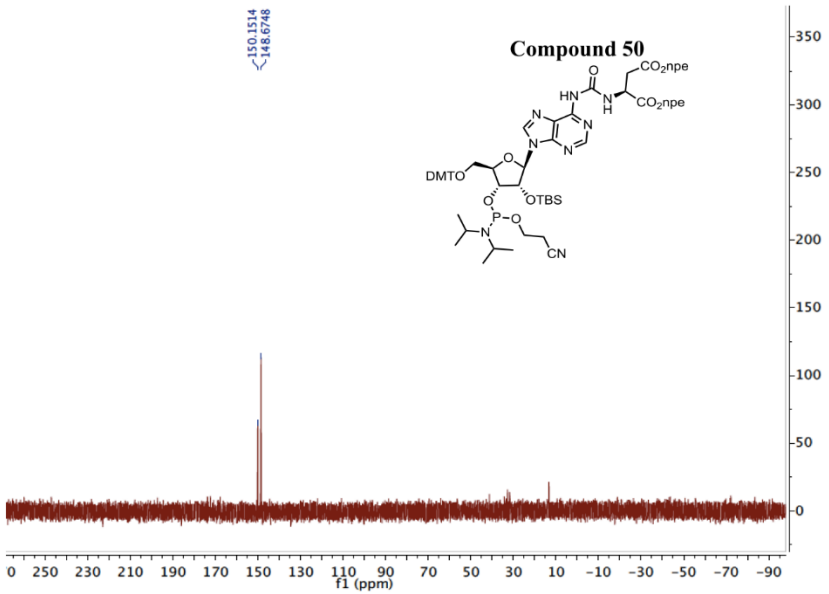


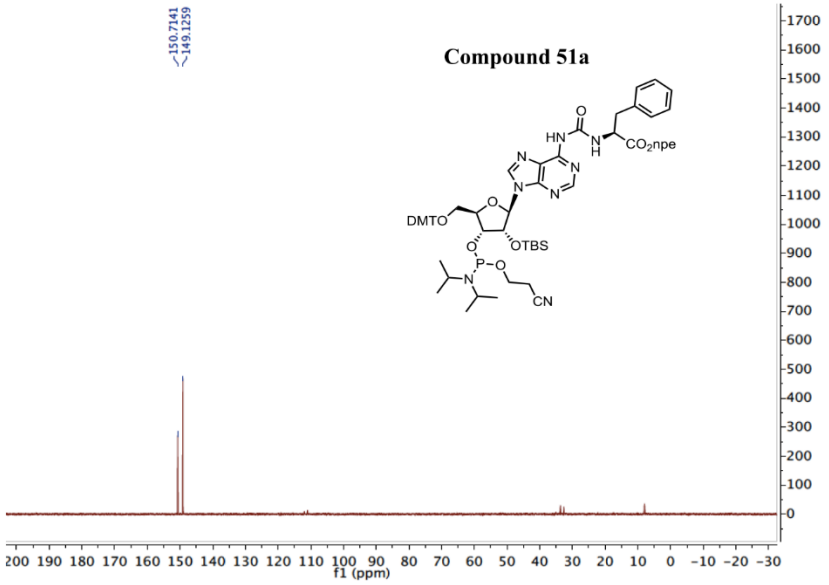
Compound 47

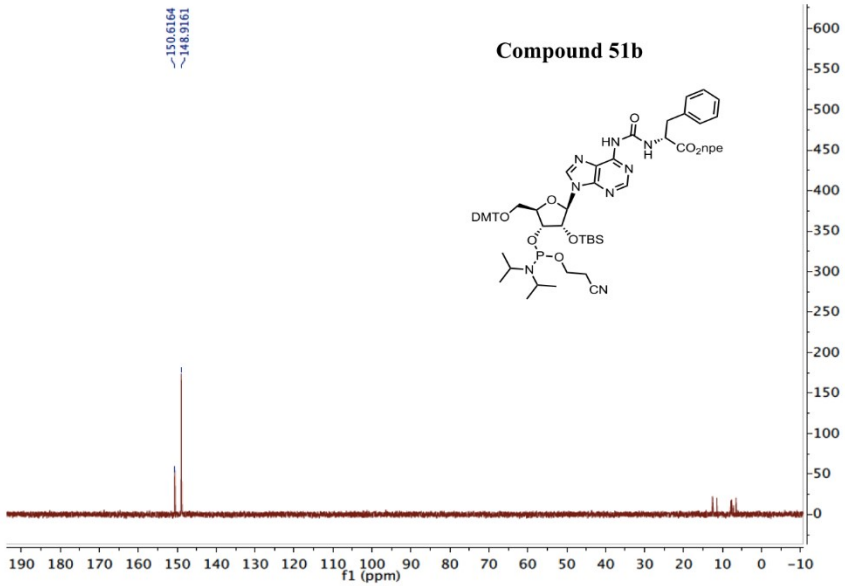


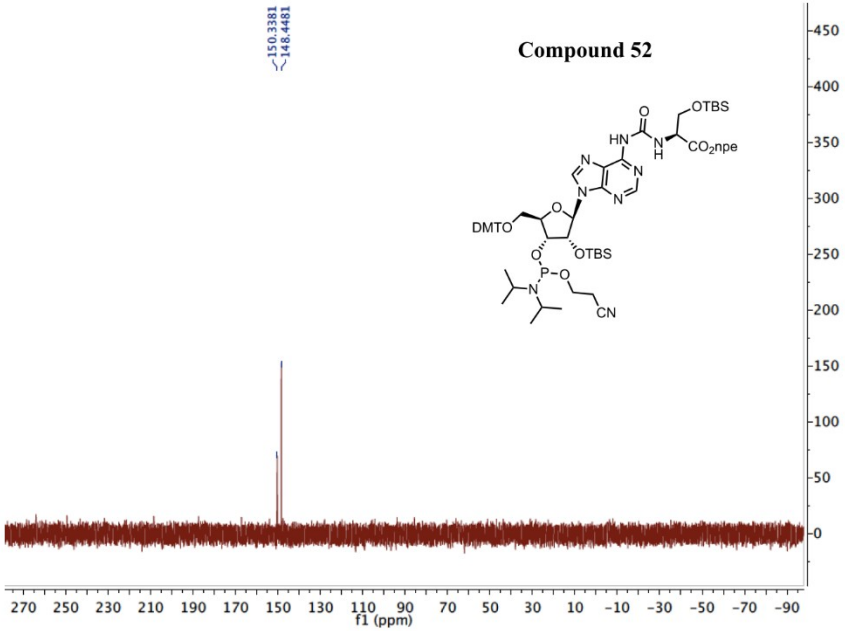

150.2206
148.6483
Compound 48
DMTO
OTBS
CN
f1 (ppm)

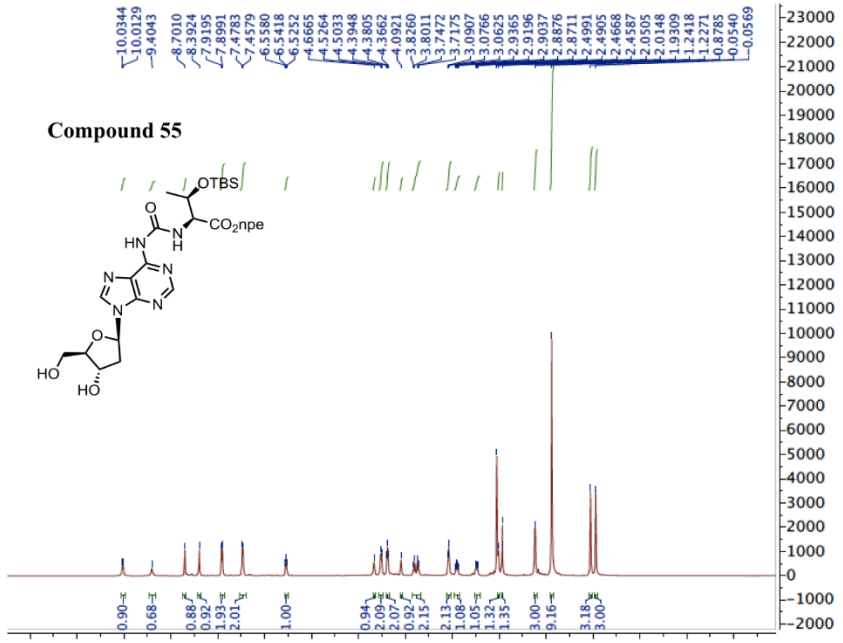

150.32
148.60
Compound 49
DMTO
OTBS
CN
ppm

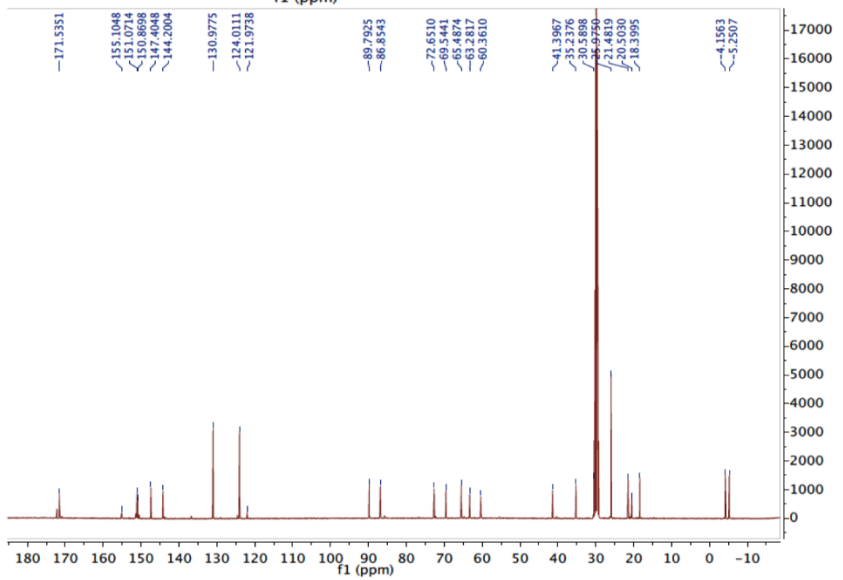

150.1514
148.6748
Compound 50
DMTO
OTBS
CN
f1 (ppm)

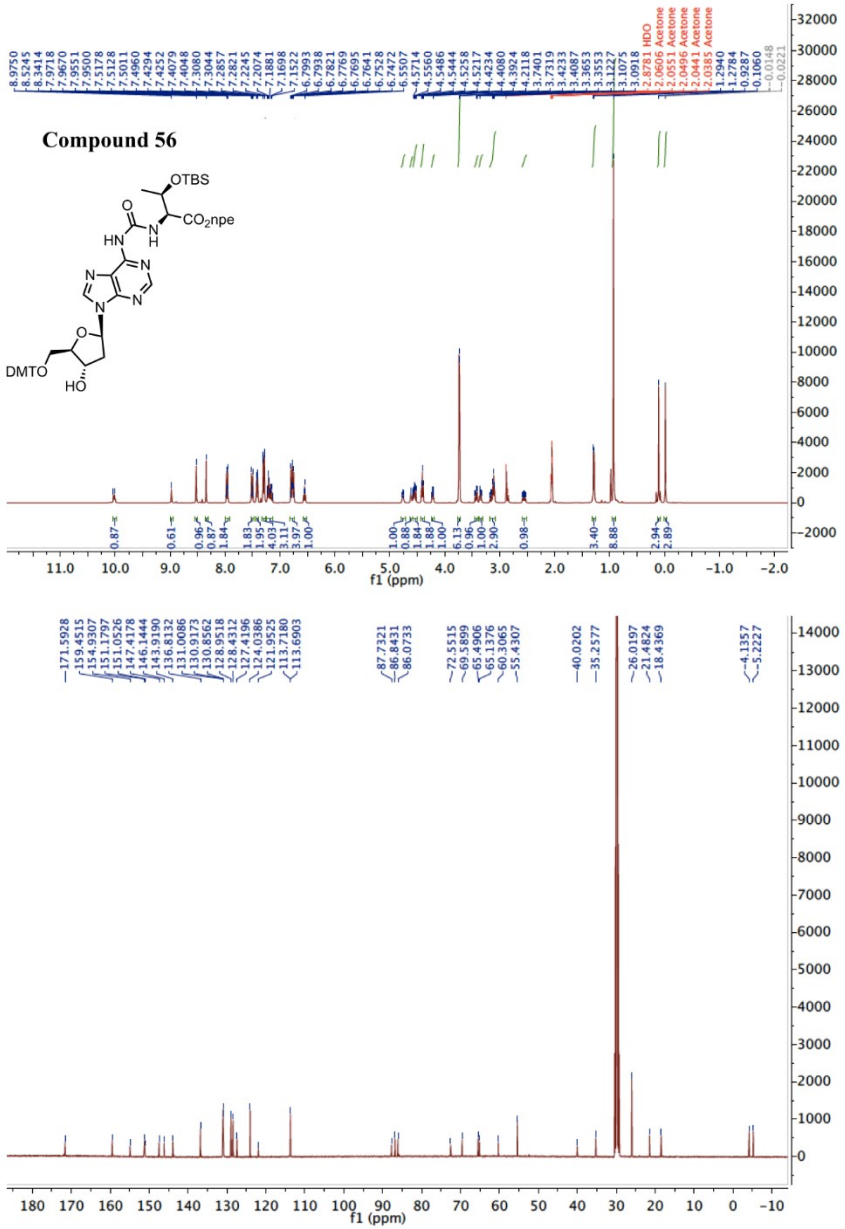

150.7141
149.1259
Compound 51a
DMTO
OTBS
CN
f1 (ppm)

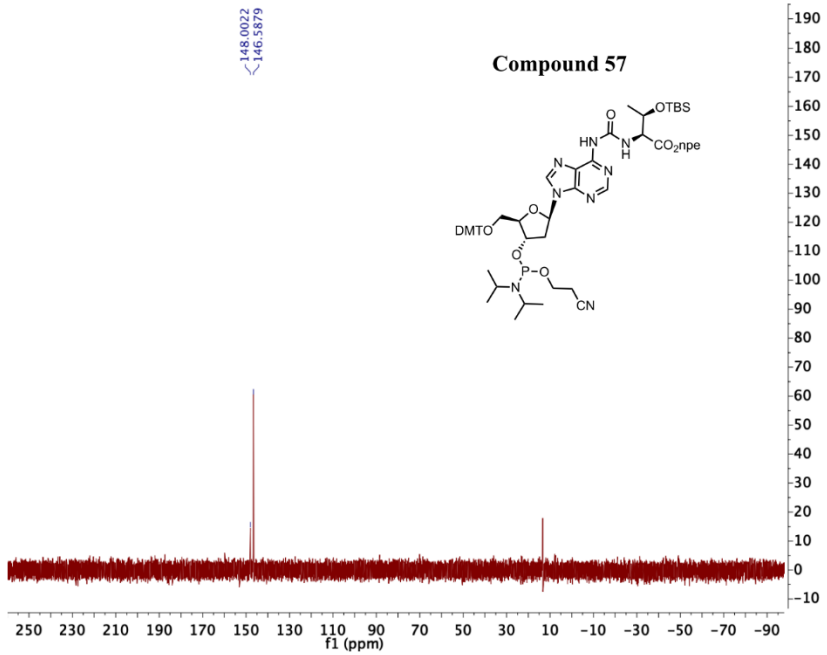

Compound 51b
150.6164
148.9161
DMTO
OTBS
CN
CO2npe
f1 (ppm)


Compound 52
150.3381
148.4481
DMTO
OTBS
OTBS
CN
CO2npe
f1 (ppm)


Compound 55
OTBS
CO2npe
HO
HO
f1 (ppm)


f1 (ppm)


Compound 56
OTBS
CO2npe
HN
DMTO
HO
f1 (ppm)


Compound 57
OTBS
CO2npe
HN
DMTO
CN
f1 (ppm)

6. References

[1] V. Boudou, J. Langridge, A. Van Aerschot, C. Hendrix, A. Millar, P. Weiss, P. Herdewijn, *Helvetica Chimica Acta* **2000**, *83*, 152 – 161.

[2] G. Lezczynska, P. Leonczak, A. Dziergowska, A. Malkiewicz, *Nucleosides, Nucleotides and Nucleic Acids* **2013**, *32*, 599 – 616.

[3] F. Himmelsbach, B. S. Schultz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59 – 72.

[4] V. Serebryany, L. Beigelman, *Tetrahedron Lett*. **2002**, *43*, 1983 – 1985.

[5] E. M. van der Wenden, J. K. von Frijtag Drabbe Künzel, R. A. A. Mathot, M. Danhof, A. P. IJzerman, W. Soudijn, *J. Med. Chem.* **1995**, *38*, 4000-4006.

Anhang II

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

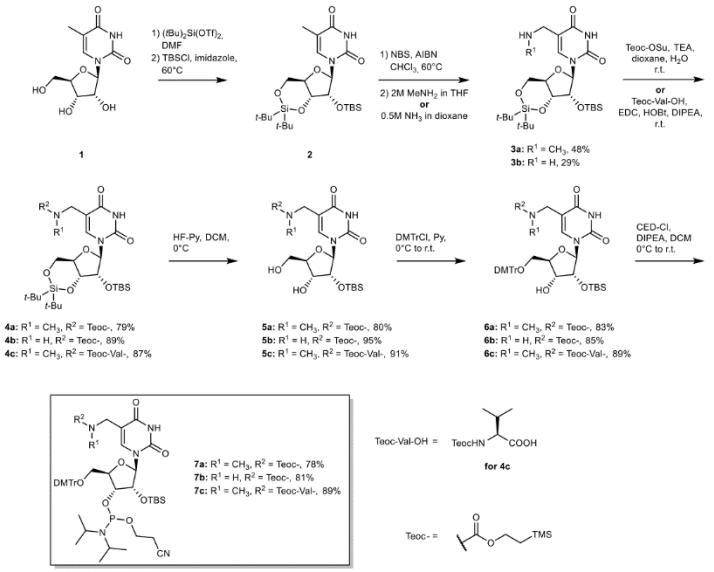
Table of Contents

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{^{1}H} NMR and ^{31}P{^{1}H} 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 $C_{24}H_{46}N_3O_6Si_2$ 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 $C_{31}H_{60}N_3O_6Si_3$ 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 $C_{30}H_{58}N_3O_6Si_3$ 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 $C_{44}H_{60}N_3O_{10}Si_2$ 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.

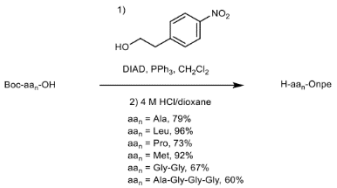
7a: Yield: 78%; R_f = 0.17 (2:1 *i*-Hexane/EtOAc); $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.3, 150.2, 148.8, 148.3; HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{53}H_{77}N_5O_{11}PSi_2$ 1046.4901, Found 1046.4896.

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 $C_{58}H_{86}N_6O_{12}PSi_2$ 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 0ºC 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 0ºC. 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.

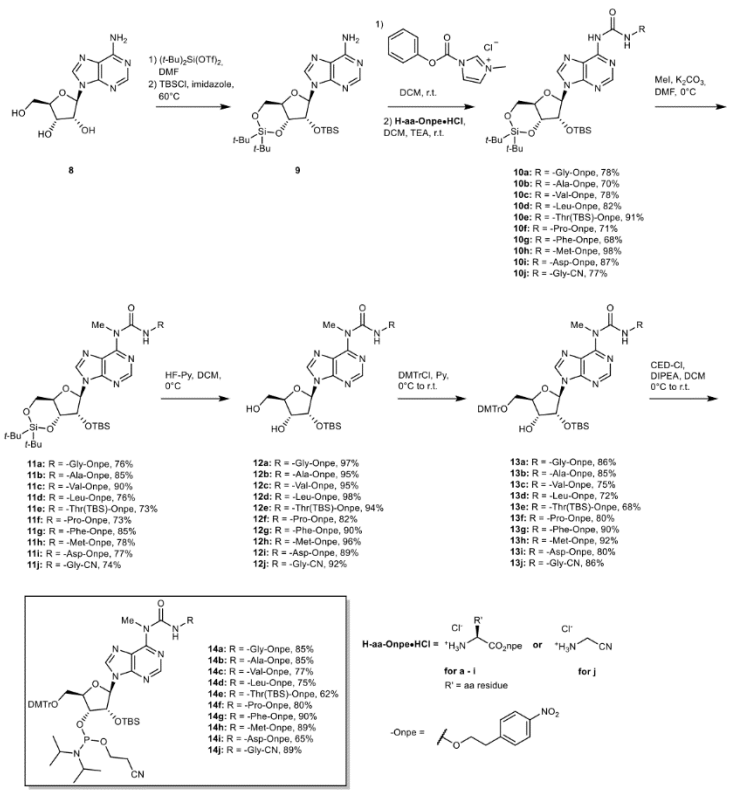
H-Met-Onpe•HCl: Yield: 92% over two steps; IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2852 (w), 1756 (m), 1743 (m), 1598 (w), 1567 (w), 1509 (s), 1347 (s), 1279 (m), 1256 (w), 1230 (w), 1206 (m), 1194 (m), 1148 (w), 1109 (w), 1066 (m), 1000 (w), 856 (m), 827 (m), 793 (w), 769 (m), 744 (s), 694 (m); ^1H NMR (400 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 0ºC 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 0ºC. 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



8

9

10a: R = -Gly-Onpe, 78%
10b: R = -Ala-Onpe, 70%
10c: R = -Val-Onpe, 78%
10d: R = -Leu-Onpe, 82%
10e: R = -Thr(TBS)-Onpe, 91%
10f: R = -Pro-Onpe, 71%
10g: R = -Phe-Onpe, 68%
10h: R = -Met-Onpe, 98%
10i: R = -Asp-Onpe, 87%
10j: R = -Gly-CN, 77%

11a: R = -Gly-Onpe, 76%
11b: R = -Ala-Onpe, 85%
11c: R = -Val-Onpe, 90%
11d: R = -Leu-Onpe, 76%
11e: R = -Thr(TBS)-Onpe, 73%
11f: R = -Pro-Onpe, 73%
11g: R = -Phe-Onpe, 85%
11h: R = -Met-Onpe, 78%
11i: R = -Asp-Onpe, 77%
11j: R = -Gly-CN, 74%

12a: R = -Gly-Onpe, 97%
12b: R = -Ala-Onpe, 95%
12c: R = -Val-Onpe, 95%
12d: R = -Leu-Onpe, 98%
12e: R = -Thr(TBS)-Onpe, 94%
12f: R = -Pro-Onpe, 82%
12g: R = -Phe-Onpe, 90%
12h: R = -Met-Onpe, 96%
12i: R = -Asp-Onpe, 89%
12j: R = -Gly-CN, 92%

13a: R = -Gly-Onpe, 86%
13b: R = -Ala-Onpe, 85%
13c: R = -Val-Onpe, 75%
13d: R = -Leu-Onpe, 72%
13e: R = -Thr(TBS)-Onpe, 68%
13f: R = -Pro-Onpe, 80%
13g: R = -Phe-Onpe, 90%
13h: R = -Met-Onpe, 92%
13i: R = -Asp-Onpe, 80%
13j: R = -Gly-CN, 86%

14a: R = -Gly-Onpe, 85%
14b: R = -Ala-Onpe, 85%
14c: R = -Val-Onpe, 77%
14d: R = -Leu-Onpe, 75%
14e: R = -Thr(TBS)-Onpe, 62%
14f: R = -Pro-Onpe, 80%
14g: R = -Phe-Onpe, 90%
14h: R = -Met-Onpe, 89%
14i: R = -Asp-Onpe, 65%
14j: R = -Gly-CN, 89%

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 $C_{27}H_{46}N_7O_5Si_2$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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 $C_{28}H_{48}N_7O_5Si_2$ 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); $^{13}C\{^1H\}$ 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 $C_{28}H_{40}N_7O_9Si$ 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); $^{13}C\{^1H\}$ 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); $^{13}C\{^1H\}$ 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 $C_{31}H_{46}N_7O_9Si$ 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$

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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); ^{1}H 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); $^{13}C\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm): 150.1, 148.7; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{58}H_{75}N_9O_{12}PSi$ 1148.5037; Found 1148.5052.

14b: Yield: 85%; R_f = 0.50 (1:1 *i*-Hexane/EtOAc); $^{31}P\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ 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.

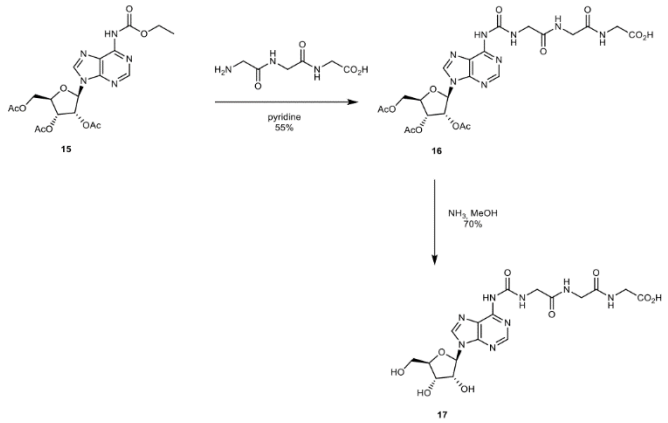
14g: Yield: 90%; R_f = 0.30 (5:3 *i*-Hexane/EtOAc); $^{31}P\{^{1}H\}$ NMR (162 MHz, acetone-d_6, 298 K) δ (ppm) 150.7, 149.1; HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{65}H_{81}N_9O_{12}PSi$ 1238.5506; Found 1238.5530.

14h: Yield: 89%; R_f = 0.30 (6:4 DCM/EtOAc); $^{31}P\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ 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); $^{31}P\{^{1}H\}$ 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



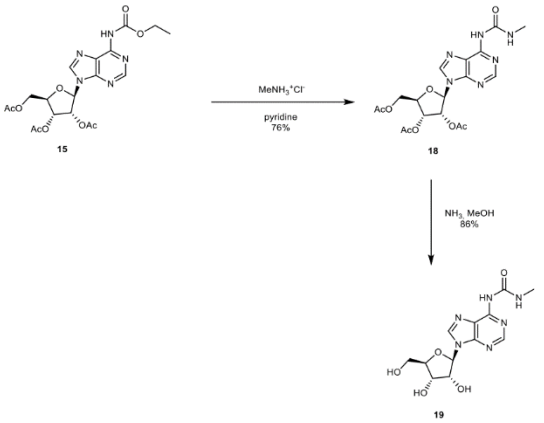
Scheme S4. Synthesis of N^6-triglycinylcarbamoyl adenosine **17**.

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

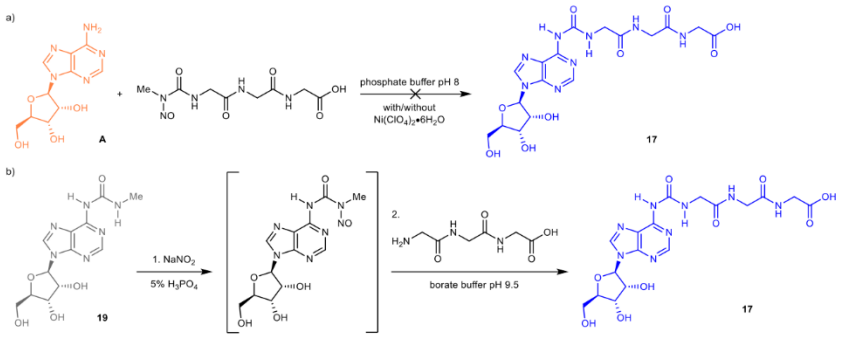


Scheme S5. Synthesis of N^6-methylurea adenosine **19**.

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 $Ni(ClO_4)_2•6H_2O$ (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.

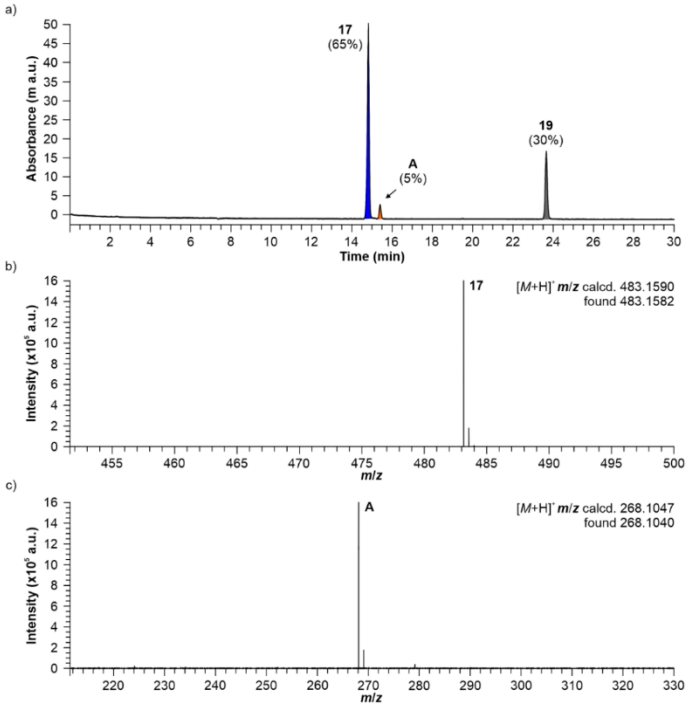
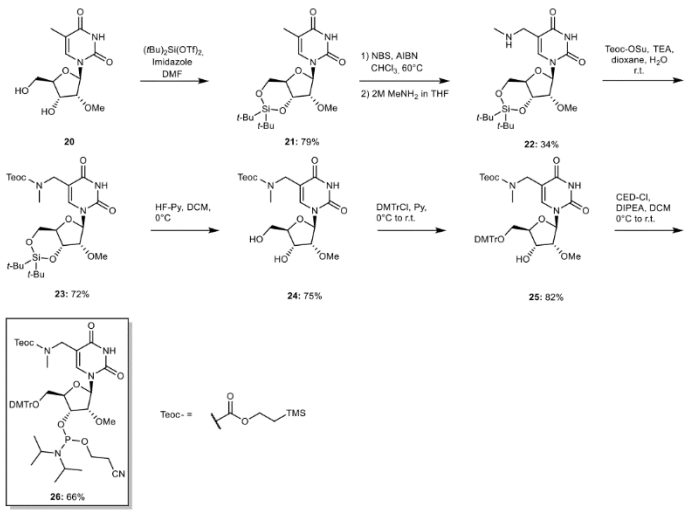


Figure S1. a) HPL-chromatogram of the reaction crude using the **Method B** shown in Scheme S6; mass spectra of the chromatographic peaks observed at: b) 14.8 and c) 15.4 min. The mass spectra confirmed the formation of the compounds **17** and **A**.

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^{5}U **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{^{1}H} 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 $C_{19}H_{33}N_2O_6Si$ 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{^{1}H} 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 $C_{20}H_{36}N_3O_6Si$ 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{^{1}H} 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{^{1}H} 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.

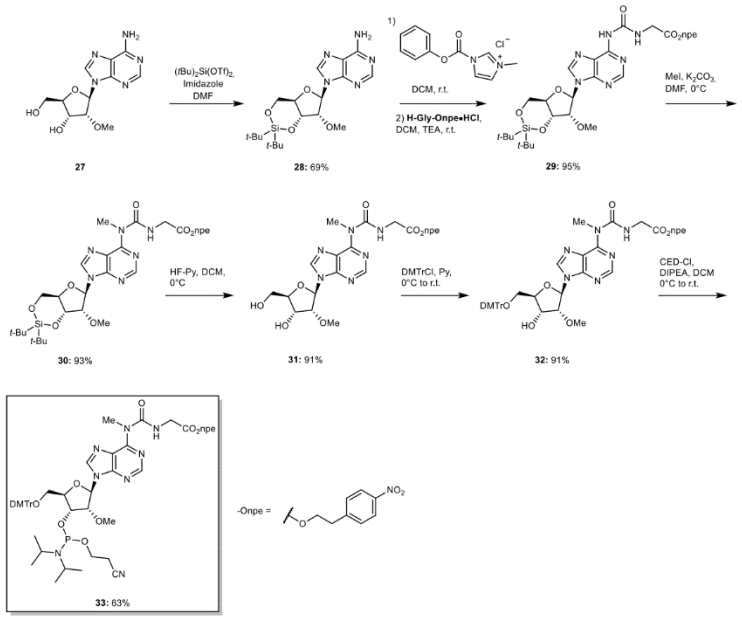
25: Yield: 82%; R_f = 0.34 (1:1 DCM/EtOAc); IR (ATR) $\tilde{\nu}$ (cm^{-1}): 2953 (w), 1694 (m), 1607 (w), 1508 (m), 1461 (m), 1397 (w), 1344 (w), 1298 (w), 1245 (s), 1175 (m), 1166 (m), 1063 (m), 1032 (s), 962 (w), 832 (s), 756 (w), 726 (w); For major rotamer: ^{1}H NMR (400 MHz, acetone-d_6, 298 K) 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{^{1}H} 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{^{1}H} 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 $C_{30}H_{42}N_7O_8Si$ 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 $C_{31}H_{44}N_7O_8Si$ 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 $C_{53}H_{63}N_8O_{12}P$ 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

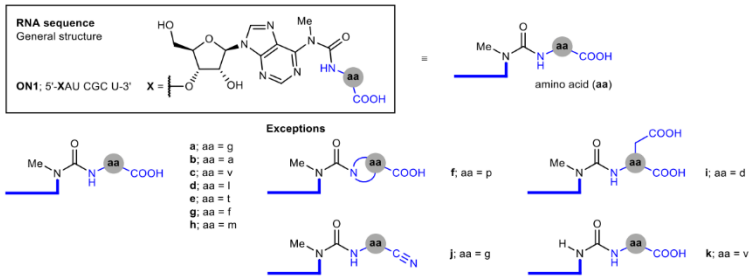


Figure S2. RNA sequence and general structure of amino acid-modified carbamoyl adenosine derivatives.

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

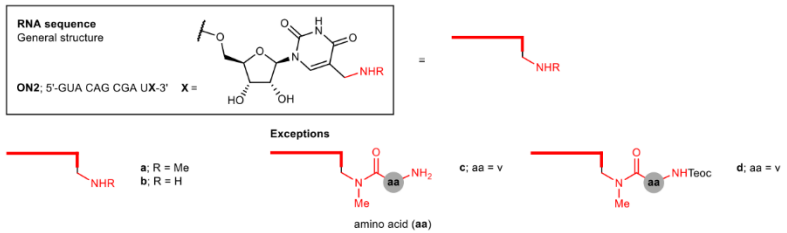


Figure S3. RNA sequence and general structure of (methyl)aminomethyl uridine derivatives.

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 Um**X**-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 **A**GC U-3'; **X** = m^6g^6A (one mismatch marked in red)

ON1s: 5'-**X**A**G** C**C**C 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 $[M\text{-}H]^-$	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**

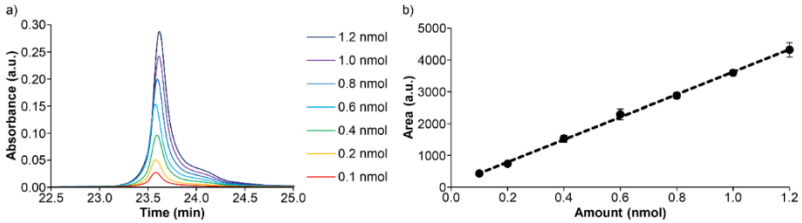


Figure S4. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON1**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent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

Calibration curve of **CON2**

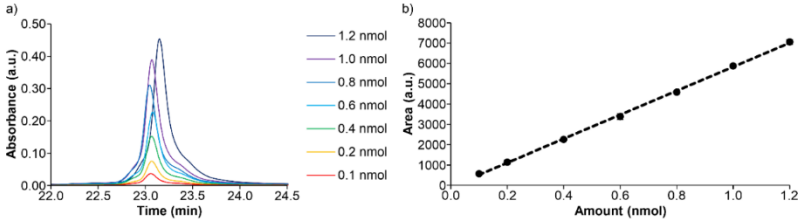


Figure S5. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON2**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent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

Calibration curve of **CON3**

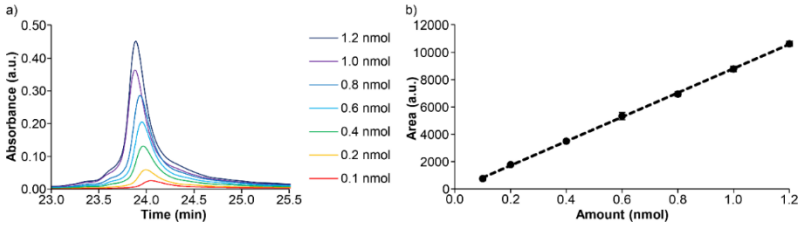


Figure S6. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON3**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent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

Calibration curve of **ON3a**

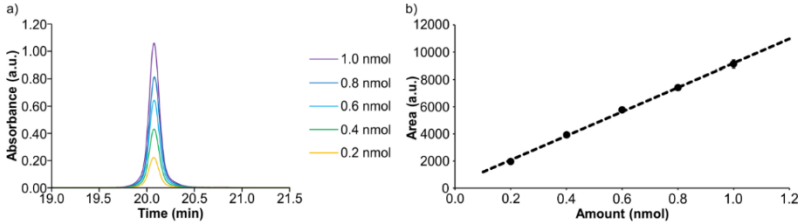


Figure S7. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **ON3a**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent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

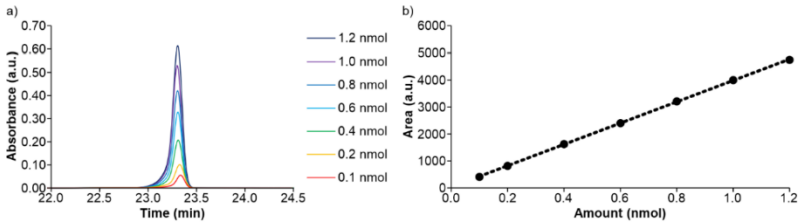


Figure S8. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON4**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent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

Calibration curve of **CON5**

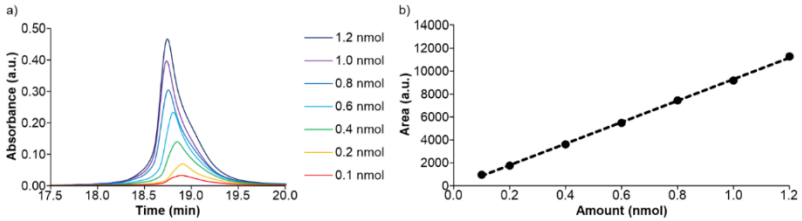


Figure S9. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON5**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent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

Calibration curve of **CON6**

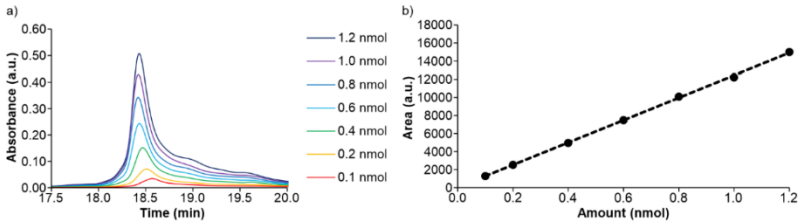


Figure S10. a) Selected region of the HPL-chromatograms upon the injection of incremental amounts (nmol) and b) chromatographic area (a.u.) *vs.* amount (nmol) of **CON6**. In b) the line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent 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 ($M^{-1}\cdot cm^{-1}$)
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).

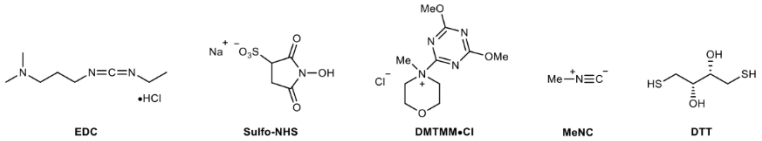


Figure S11. Activators of carboxylic acid and nitrile groups.

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

6.1 Control experiments

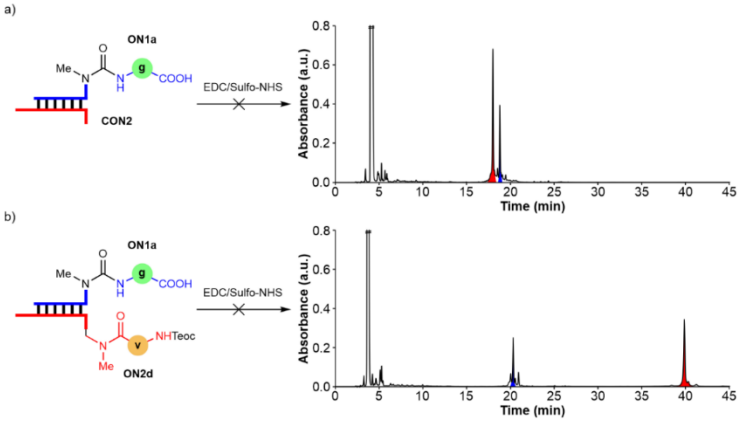


Figure S12. HPL-chromatograms of the reactions of **ON1a**; **X** = m^6g^6A with: a) **CON2** (complementary canonical ON) and b) **ON2d**; **X** = Teoc-ymnm^{5}U in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

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'-ymnm^{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^6g^6A) and ON2a (mnm^5U)

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

MES buffer at pH 6 (adjusted with NaOH)

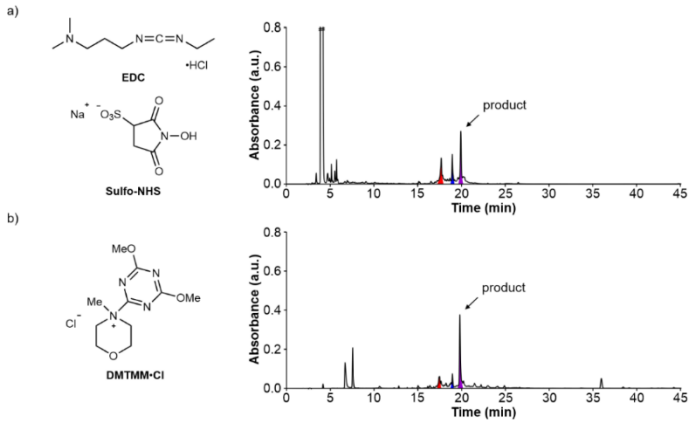


Figure S13. HPL-chromatograms of the reactions of **ON1a**; **X** = m^6g^6A with **ON2a** using: a) EDC/Sulfo-NHS and b) DMTMM•Cl as activators.

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**.

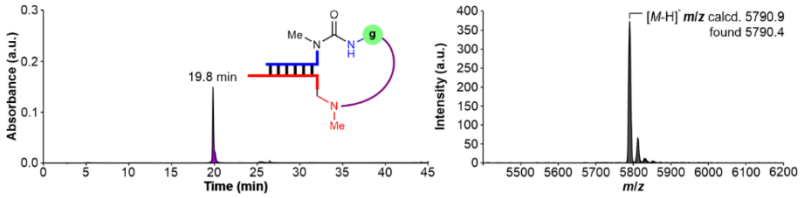


Figure S14. Left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product **ON3a**.

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.

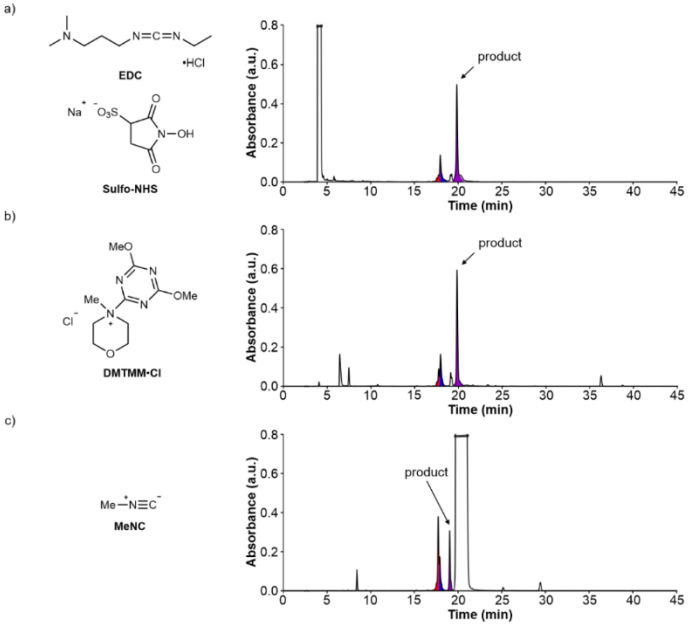


Figure S15. HPL-chromatograms of the reactions of **ON1a**; **X** = m^5g^5A with **ON2b** using: a) EDC/Sulfo-NHS; b) DMTMM•Cl and c) MeNC as activators. MES buffer (100 mM) at pH 6 in a) and b). DCI buffer (50 mM) at pH 6 in c).

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**.

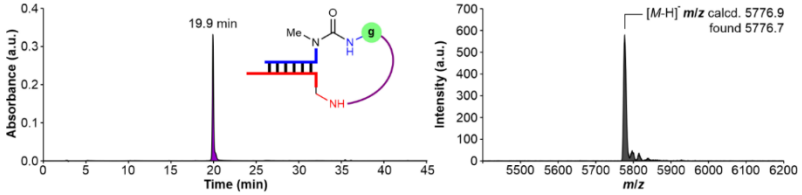


Figure S16. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product **ON3b**.

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

Scheme S11. Coupling of **ON1a**; **X** = m^{8}g^{6}A with **ON2c**. The formed peptide bond is marked in purple.

Buffer at pH 6

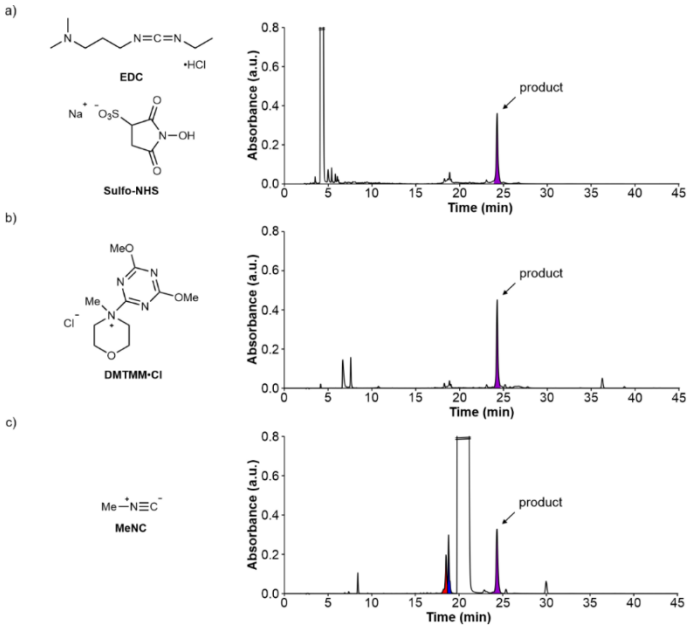


Figure S17. HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2c** using: a) EDC/Sulfo-NHS; b) DMTMM•Cl and c) MeNC as activators. MES buffer (100 mM) at pH 6 in a) and b). DCI buffer (50 mM) at pH 6 in c).

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**.

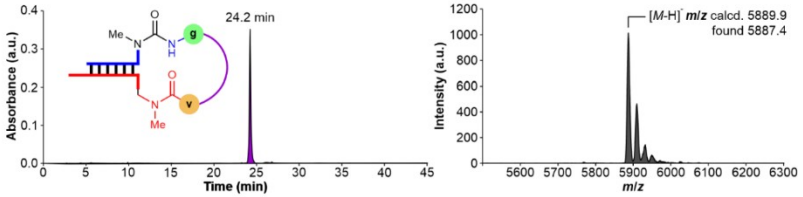


Figure S18. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product **ON3c**.

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

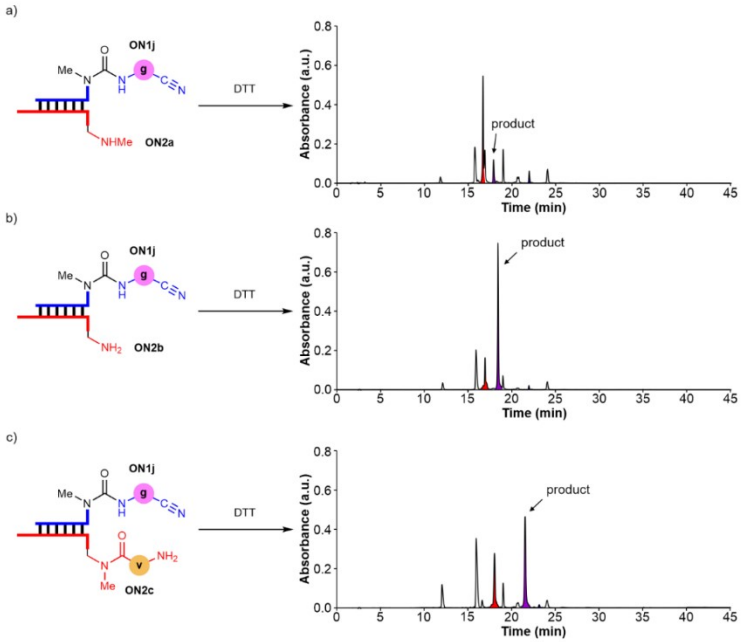


Figure S19. HPL-chromatograms of the reactions of **ON1j**; **X** = m^{6}g^{6}A (amino nitrile) with: a) **ON2a**; **X** = mnm^{5}U; b) **ON2b**; **X** = nm^{5}U and c) **ON2c**; **X** = vmnm^{5}U in boric acid buffer at pH 8 using DTT as activator.

Table S13. Results obtained in the coupling reactions of **ON1j**; **X** = m^{6}g^{6}A (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}g^{6}A (amino nitrile)	**ON2a**; **X** = mnm^{5}U	12±1
	ON2b; **X** = nm^{5}U	65±2
	ON2c; **X** = vmnm^{5}U	42±1

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

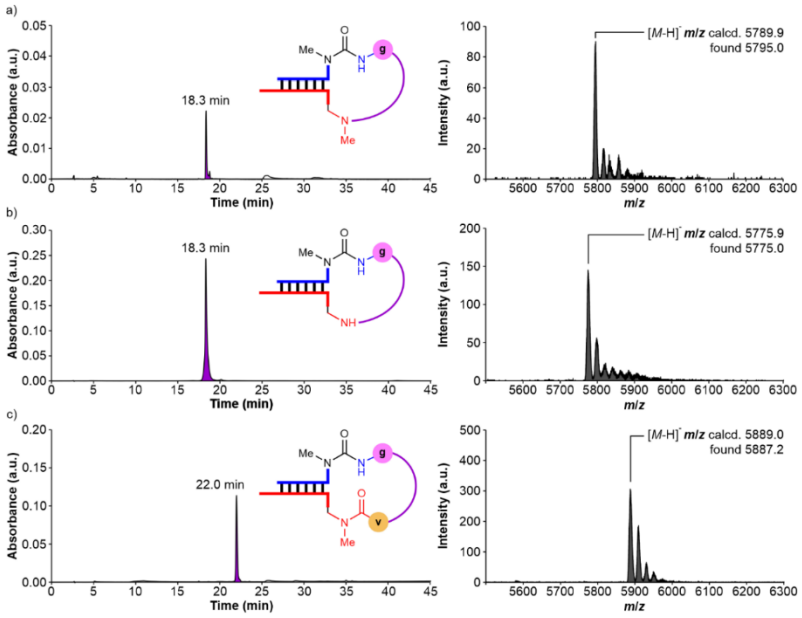


Figure S20. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON1j**; **X** = m^{6}g^{6}A (amino nitrile) with: a) **ON2a**; **X** = mnm^{5}U; b) **ON2b**; **X** = nm^{5}U and c) **ON2c**; **X** = vmnm^{5}U.

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

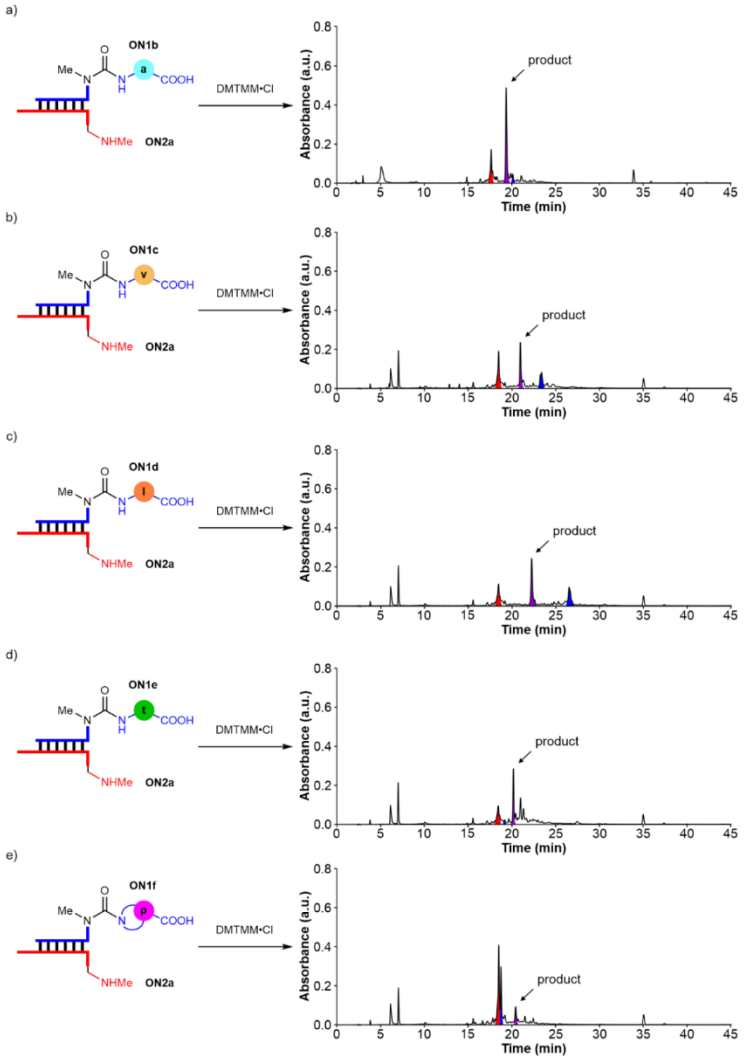


Figure S21. HPL-chromatograms of the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1b**; **X** = m^{6}a^{6}A; b) **ON1c**; **X** = m^{6}v^{6}A; c) **ON1d**; **X** = m^{6}i^{6}A; d) **ON1e**; **X** = m^{6}t^{6}A and e) **ON1f**; **X** = m^{6}p^{6}A in MES buffer at pH 6 using DMTMM•Cl as activator.

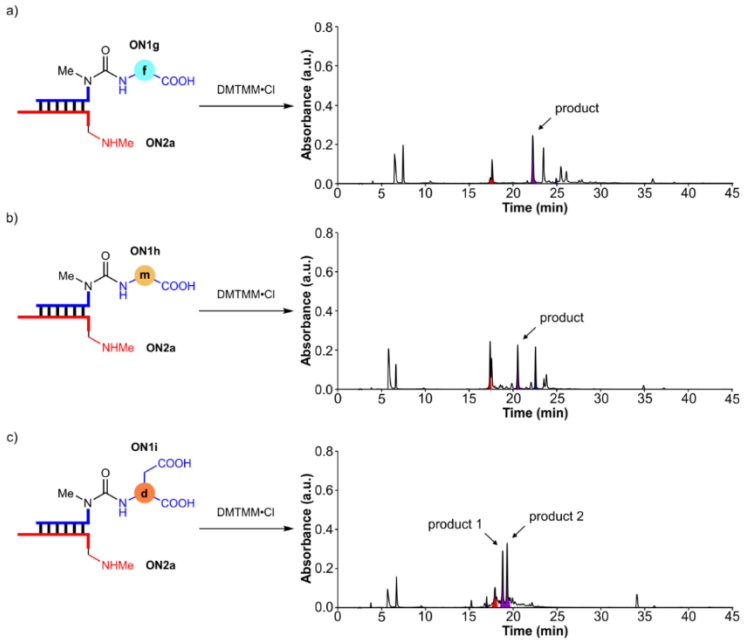


Figure S22. HPL-chromatograms of the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1g**; **X** = m^{6}f^{6}A; b) **ON1h**; **X** = m^{6}m^{6}A and c) **ON1i**; **X** = m^{6}d^{6}A in MES buffer at pH 6 using DMTMM•Cl as activator. For **ON1i**, the two peaks corresponded to the products of the 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}l^{6}A		27±1
ON1e; **X** = m^{6}t^{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.

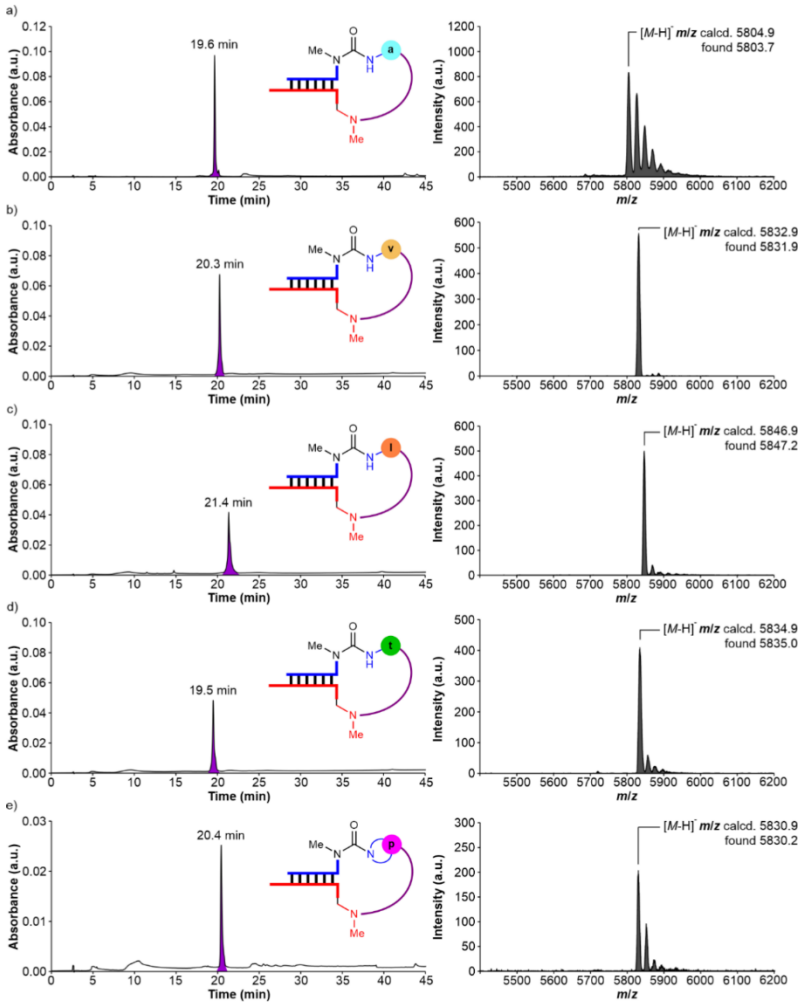


Figure S23. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1b**; **X** = m^{6}a^{6}A; b) **ON1c**; **X** = m^{6}v^{6}A; c) **ON1d**; **X** = m^{6}l^{6}A; d) **ON1e**; **X** = m^{6}t^{6}A and e) **ON1f**; **X** = m^{6}p^{6}A.

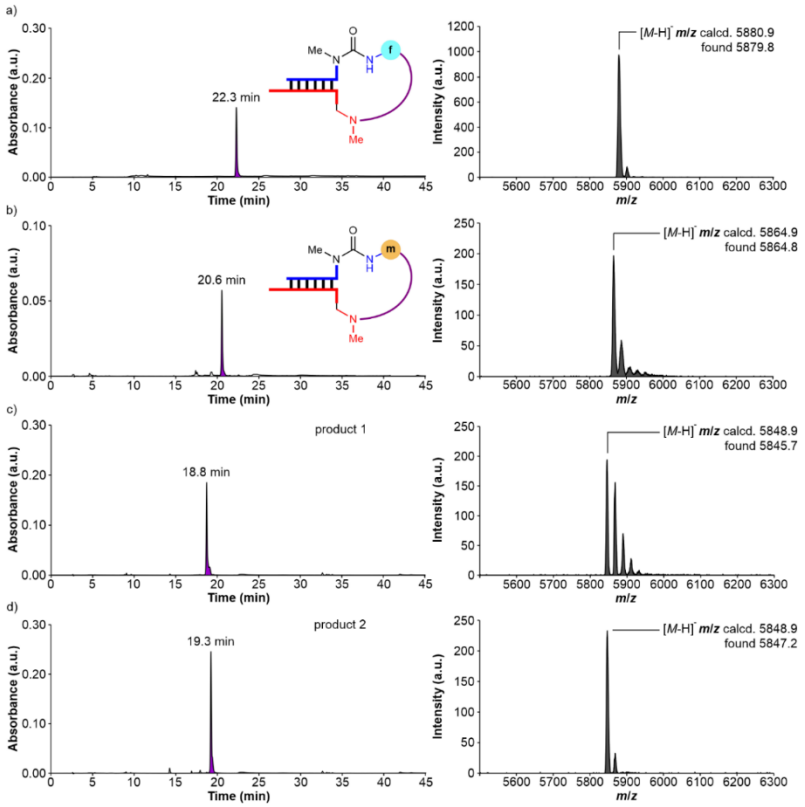


Figure S24. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2a**; **X** = mnm^{5}U with: a) **ON1g**; **X** = m^{6}t^{6}A; b) **ON1h**; **X** = m^{6}m^{6}A; c) **ON1i**; **X** = m^{6}d^{6}A and d) **ON1i**; **X** = m^{6}d^{6}A.

6.7 Coupling reactions of ON1b-i (m^{6}aa^{6}A) with ON2c

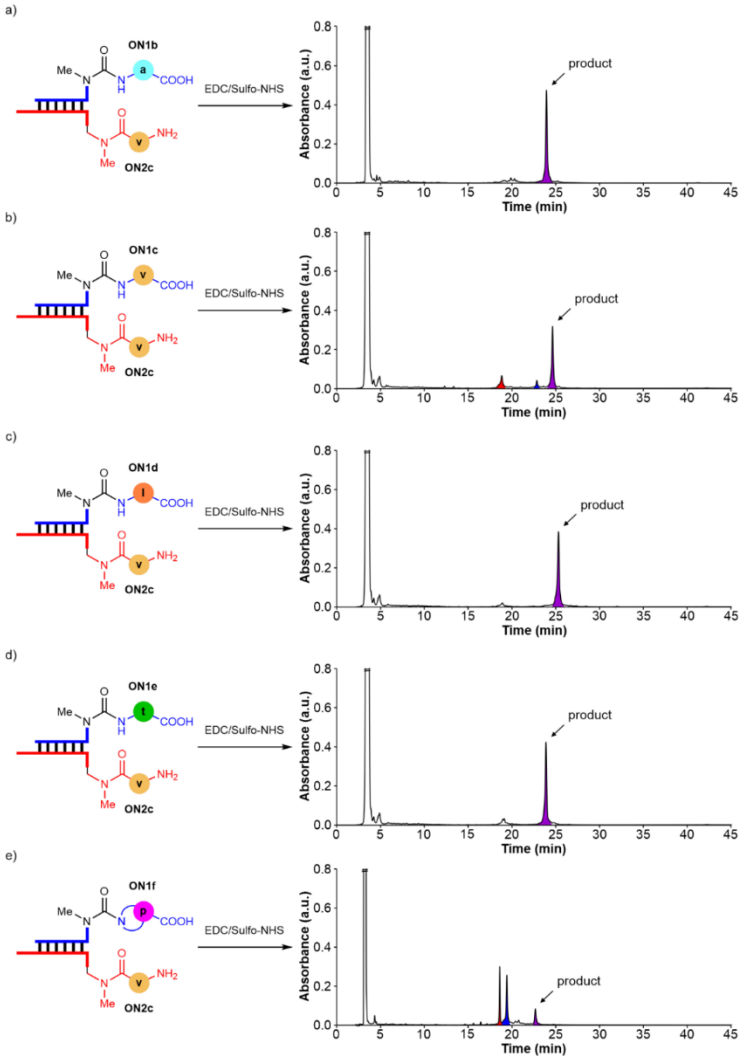


Figure S25. HPL-chromatograms of the reactions of **ON2c**; **X** = ymnm^{5}U with: a) **ON1b**; **X** = m^{6}a^{6}A; b) **ON1c**; **X** = m^{6}y^{6}A; c) **ON1d**; **X** = m^{6}f^{6}A; d) **ON1e**; **X** = m^{6}t^{6}A and e) **ON1f**; **X** = m^{6}p^{6}A in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

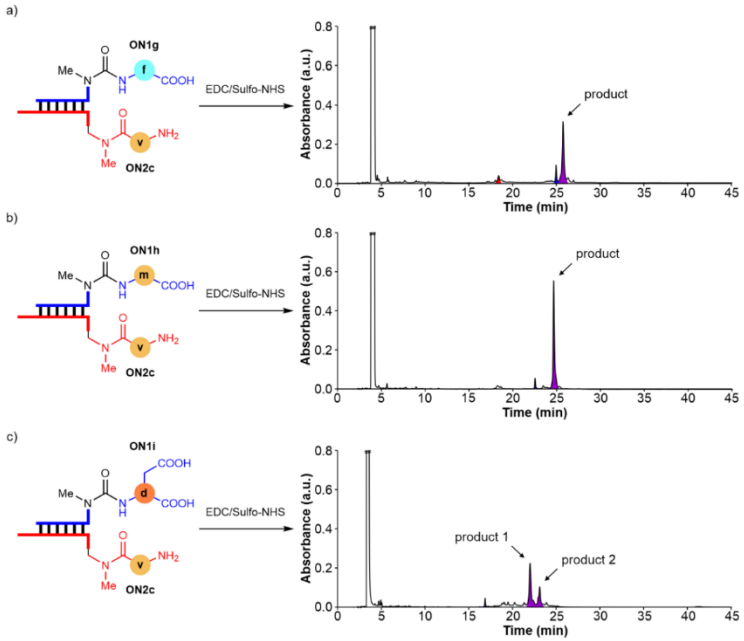


Figure S26. HPL-chromatograms of the reactions of **ON2c**; **X** = ymnm^{5}U with: a) **ON1g**; **X** = m^{6}f^{6}A; b) **ON1h**; **X** = m^{6}m^{6}A and c) **ON1i**; **X** = m^{6}d^{6}A in MES buffer at pH 6 using EDC/Sulfo-NHS as activator. For **ON1i**, the two peaks corresponded to the products 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}aa^{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}a^{6}A	**ON2c**; **X** = ymnm^{5}U	76±2
ON1c; **X** = m^{6}v^{6}A		54±1
ON1d; **X** = m^{6}l^{6}A		77±1
ON1e; **X** = m^{6}t^{6}A		77±1
ON1f; **X** = m^{6}p^{6}A		18±4 (55±5)[b]
ON1g; **X** = m^{6}f^{6}A		50±1
ON1h; **X** = m^{6}m^{6}A		70±2
ON1i; **X** = m^{6}d^{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.

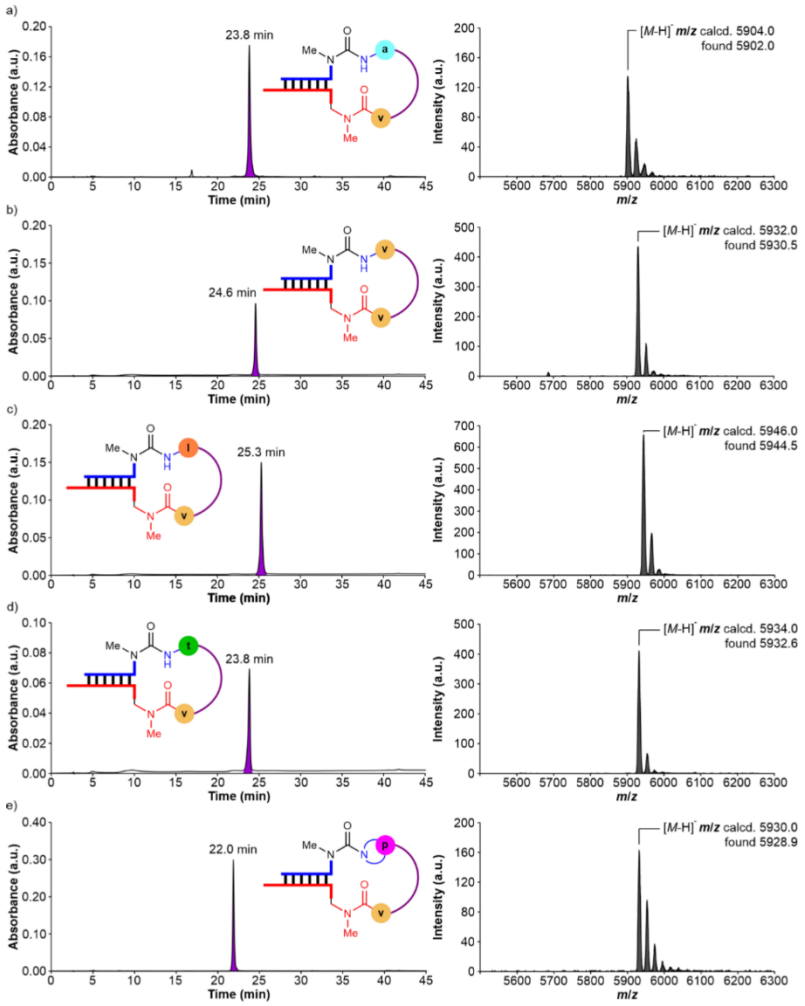


Figure S27. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2c**; **X** = ymnm^{5}U with: a) **ON1b**; **X** = m^{6}a^{6}A; b) **ON1c**; **X** = m^{6}v^{6}A; c) **ON1d**; **X** = m^{6}l^{6}A; d) **ON1e**; **X** = m^{6}t^{6}A and e) **ON1f**; **X** = m^{6}p^{6}A.

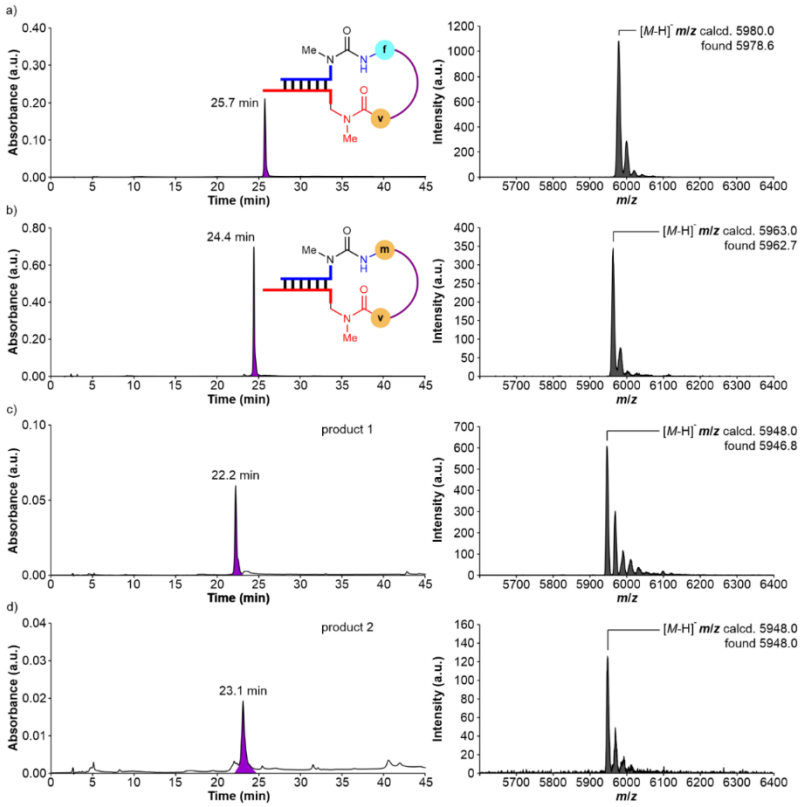


Figure S28. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2c**; **X** = vmnm^{5}U with: a) **ON1g**; **X** = m^{5}f^{6}A; b) **ON1h**; **X** = m^{5}m^{5}A; c) **ON1i**; **X** = m^{6}d^{6}A and d) **ON1i**; **X** = m^{6}d^{6}A.

7. Synthesized peptide-oligonucleotides using solid support beads

7.1 Donor peptide-oligonucleotides with a complementary sequence

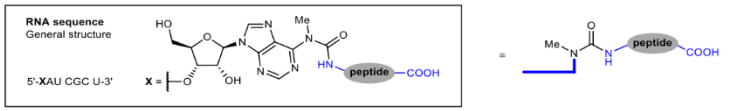


Figure S29. RNA sequence and general structure of peptide-modified carbamoyl adenosine derivatives.

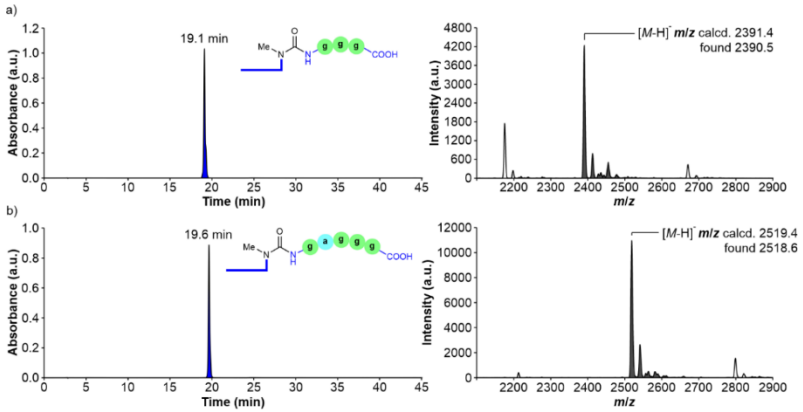


Figure S30. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the synthesized peptide-oligonucleotides: a) 5'-m^6(ggg)^{6}A-RNA-3' and b) 5'-m^6(gaggg)^{6}A-RNA-3'.

7.2 Acceptor peptide-oligonucleotides with a complementary sequence

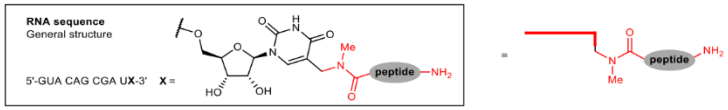


Figure S31. RNA sequence and general structure of peptide-modified methylaminomethyl uridine derivatives.

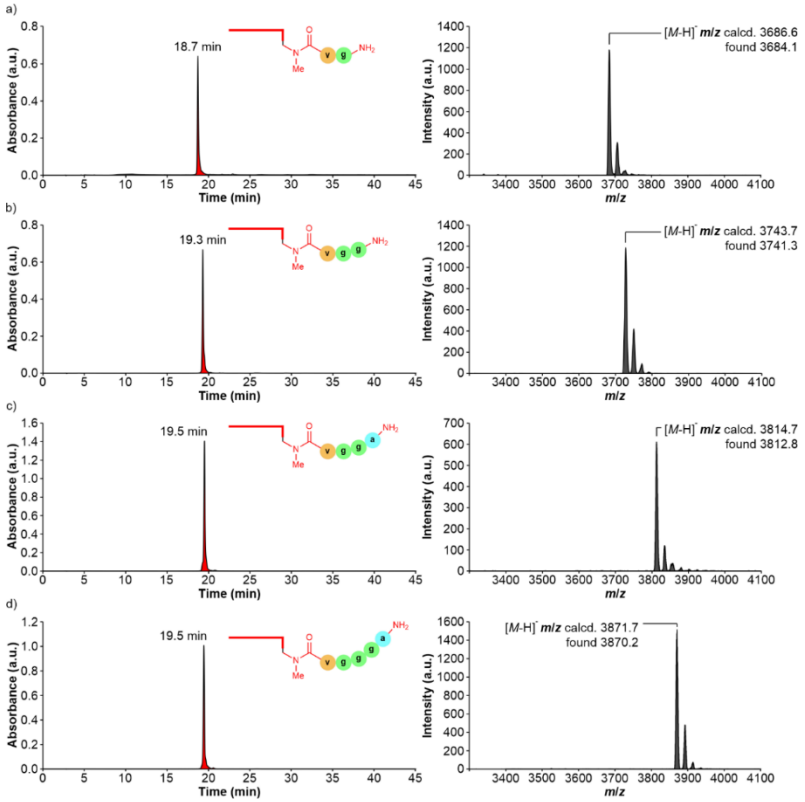


Figure S32. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the synthesized peptide-oligonucleotides: a) 3'-gvmnm^{5}U-RNA-5'; b) 3'-ggvmnm^{5}U-RNA-5'; c) 3'-aggvmnm^{5}U-RNA-5' and d) 3'-agggvmnm^{5}U-RNA-5'.

5'-GmUmAm CmAmGm CmGmAm Um**X**-3'; **X** = aggmnm^{5}U

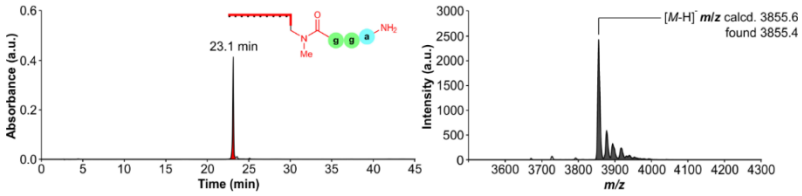


Figure S33. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the synthesized peptide-oligonucleotide 3'-aggmnm^{5}U-RNA-5' containing 2'-OMe nucleosides.

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

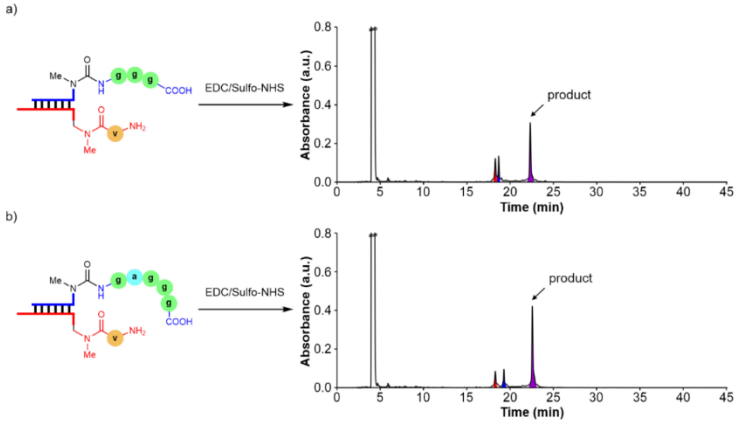


Figure S34. HPL-chromatograms of the reactions of **ON2c**; **X** = ymnm^{5}U with: a) 5'-m^6(ggg)^{6}A-RNA-3' and b) 5'-m^6(gaggg)^{6}A-RNA-3' in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S16. Results obtained in the coupling reactions of **ON2c**; **X** = vmnm^{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(ggg)^{6}A-RNA-3'	**ON2c**; **X** = vmnm^{5}U	35±1
5'-m^6(gaggg)^{6}A-RNA-3'		43±1

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

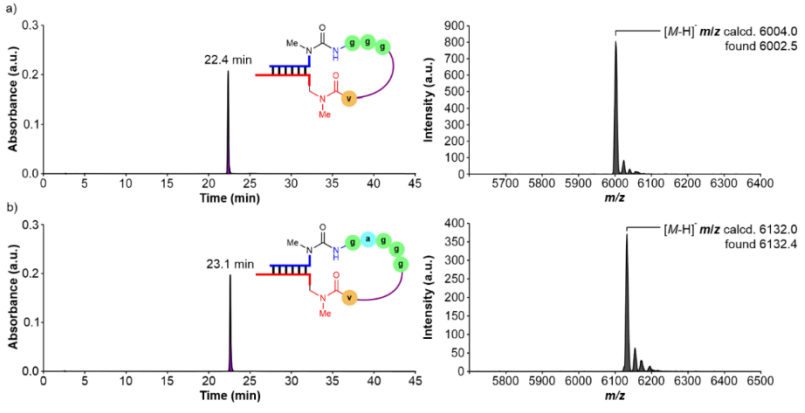


Figure S35. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON2c**; **X** = vmnm^{5}U with: a) 5'-m^6(ggg)^{6}A-RNA-3' and b) 5'-m^6(gaggg)^{6}A-RNA-3'.

8.2 Coupling reactions of ON1a (m^{6}g^{6}A) with acceptor peptide-oligonucleotides

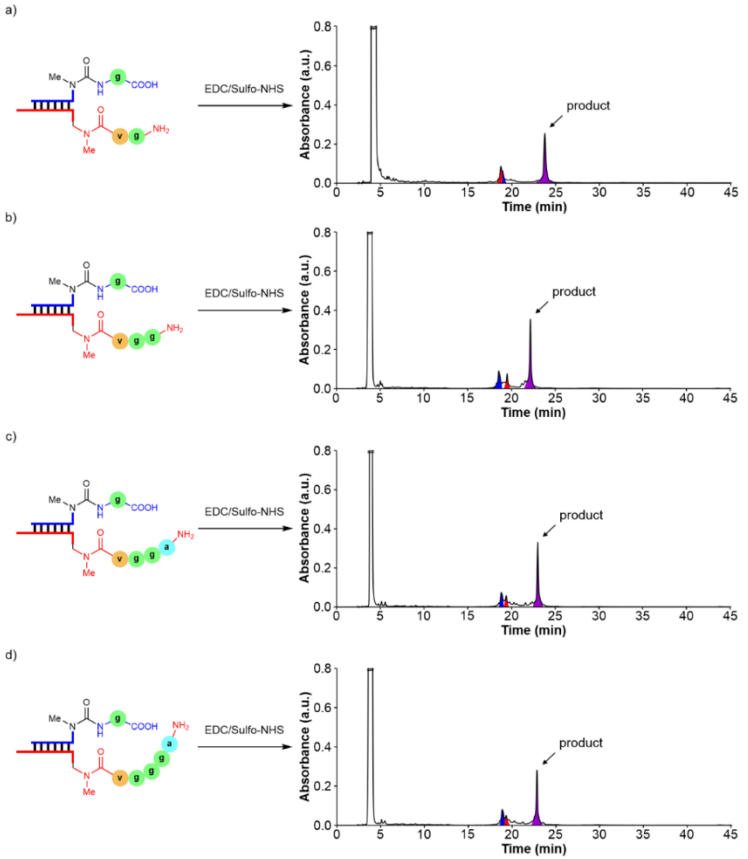


Figure S36. HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g^{6}A with: a) 3'-gvmnm^{5}U-RNA-5'; b) 3'-ggvmnm^{5}U-RNA-5'; c) 3'-aggvmnm^{5}U-RNA-5' and d) 3'-agggvmnm^{5}U-RNA-5' in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S17. Results obtained in the coupling reactions of **ON1a**; **X** = m^{6}g^{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}g^{6}A	3'-gvmnm^{5}U-RNA-5'	51±1
	3'-ggvmnm^{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.

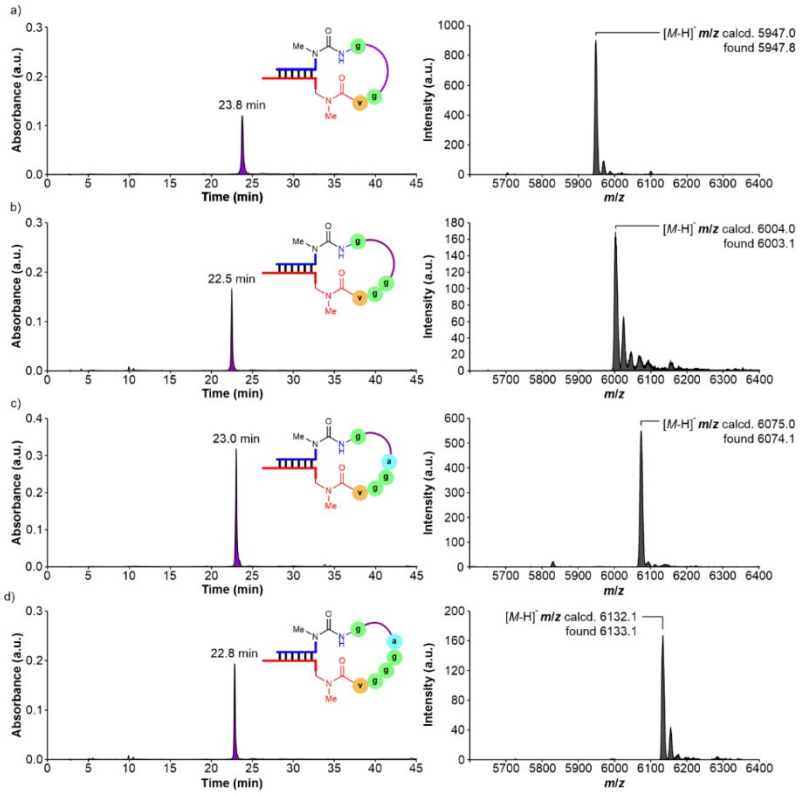


Figure S37. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of **ON1a**; **X** = m^6g^6A with: a) 3'-ggvmnm^{5}U-RNA-5'; b) 3'-ggvmnm^{5}U-RNA-5'; c) 3'-aggvmnm^{5}U-RNA-5' and d) 3'-agggvmnm^{5}U-RNA-5'.

8.3 Coupling reactions of donor and acceptor peptide-oligonucleotides

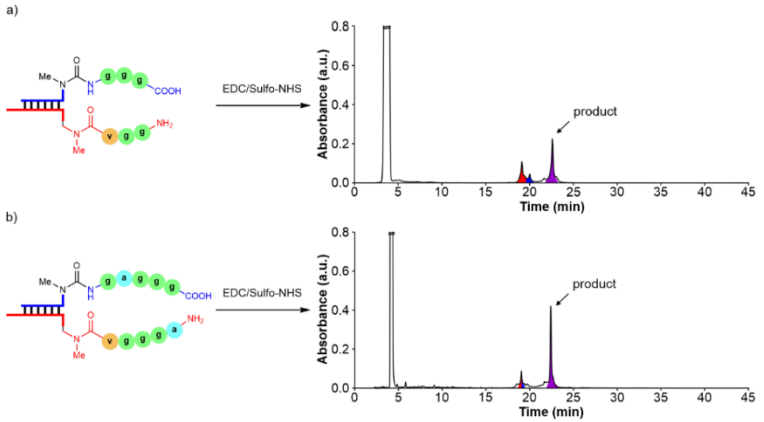


Figure S38. HPL-chromatograms of the reactions of: a) 5'-m^6(ggg)^{6}A-RNA-3' with 3'-ggvmnm^{5}U-RNA-5' and b) 5'-m^6(gaggg)^{6}A-RNA-3' with 3'-agggvmnm^{5}U-RNA-5' in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

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**.

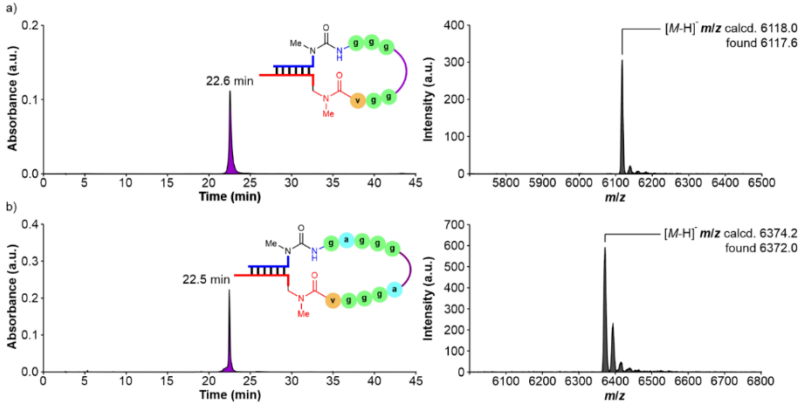


Figure S39. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of: a) 5'-m^6(ggg)^{6}A-RNA-3' with 3'-ggvmnm^{5}U-RNA-5' and b) 5'-m^6(gaggg)^{6}A-RNA-3' with 3'-agggvmnm^{5}U-RNA-5'.

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.

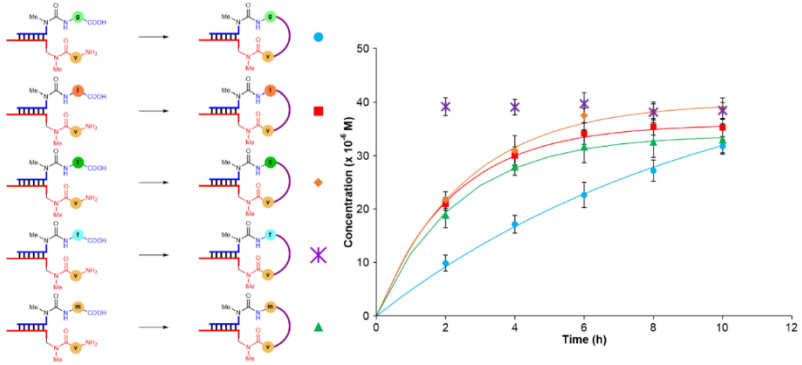


Figure S40. Concentration of the product (M) *vs.* time (h) in selected peptide coupling reactions using DMTMM•Cl as activator. Lines show fit of the data to the theoretical kinetic model. Error bars are the standard deviations.

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** = $m^6\underline{t}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	0.39±0.04	2.50 × 10^{-12}
ON1g; **X** = $m^6\underline{i}^6A$		>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.

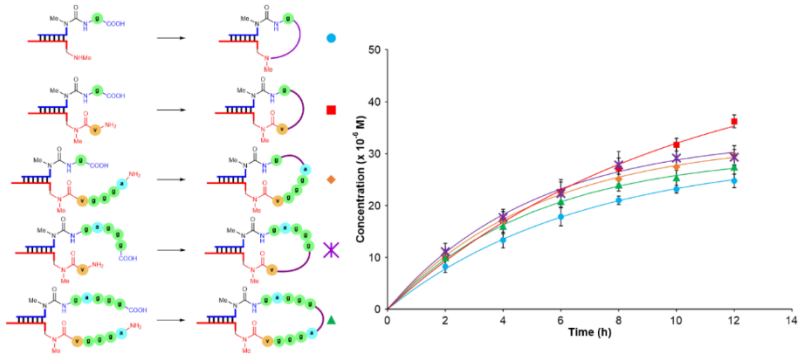


Figure S41. Concentration of the product (M) *vs.* time (h) in selected peptide coupling reactions using DMTMM•Cl as activator. Lines show fit of the data to the theoretical kinetic model. Error bars are the standard deviations.

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.

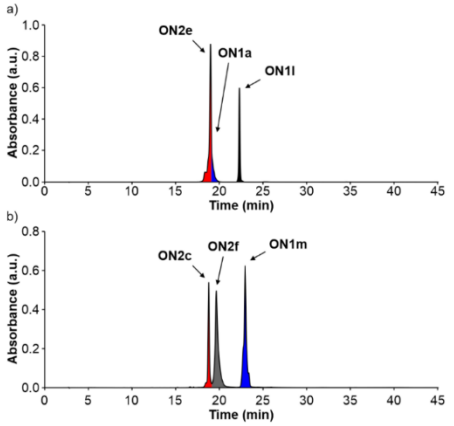


Figure S42. HPL-chromatograms of equimolar mixtures of: a) **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1l**; **X** = $m^6\underline{v}^6A$ and **ON2e**; $\mathbf{X^1}$ = $gmnm^5U$ and $\mathbf{X^2}$ = nm^5U, and b) **ON1m**; $\mathbf{X^1}$ = $m^6\underline{v}^6A$ and $\mathbf{X^2}$ = $m^6\underline{g}^6A$, **ON2c**; **X** = $\underline{v}mnm^5U$ and **ON2f**; **X** = $\underline{v}mnm^5U$.

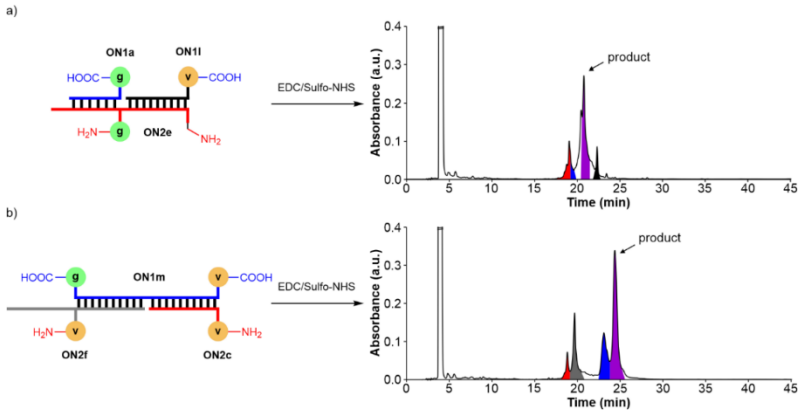


Figure S43. HPL-chromatograms of the reactions of: a) **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1l**; **X** = $m^6\underline{v}^6A$ and **ON2e**; **X**¹ = $\underline{g}mnm^5U$ and **X**² = nm^5U and b) **ON1m**; **X**¹ = $m^5\underline{v}^6A$ and **X**² = $m^6\underline{g}^6A$, **ON2c**; **X** = $\underline{v}mnm^5U$ and **ON2f**; **X** = $\underline{v}mnm^5U$ 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** = $m^6\underline{g}^6A$ **ON1l**; **X** = $m^5\underline{v}^6A$	**ON2e**; **X**¹ = $\underline{g}mnm^5U$ and **X**² = nm^5U	35±2 (29±1)[b]
ON1m; **X**¹ = $m^5\underline{v}^6A$ and **X**² = $m^6\underline{g}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$ **ON2f**; **X** = $\underline{v}mnm^5U$	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.

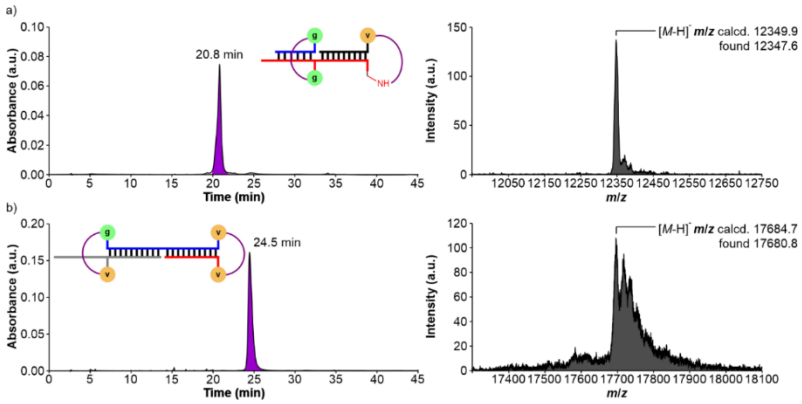


Figure S44. left) HPL-chromatograms and right) MALDI-TOF mass spectra (negative mode) of the isolated products from the reactions of: a) **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1l**; **X** = $m^5\underline{v}^6A$ and **ON2e**; **X**¹ = $\underline{g}mnm^5U$ and **X**² = nm^5U and b) **ON1m**; **X**¹ = $m^5\underline{v}^6A$ and **X**² = $m^6\underline{g}^6A$, **ON2c**; **X** = $\underline{v}mnm^5U$ and **ON2f**; **X** = $\underline{v}mnm^5U$.

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.

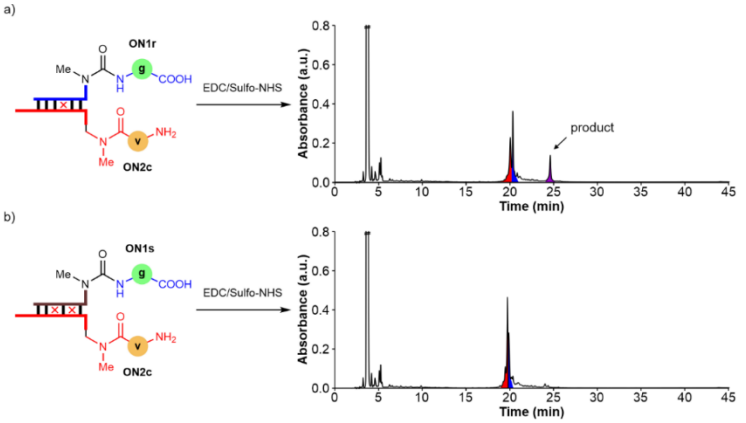


Figure S45. HPL-chromatograms of the reactions of **ON2c**; **X** = $\underline{v}mnm^5U$ with: a) **ON1r**; **X** = $m^6\underline{g}^6A$ and b) **ON1s**; **X** = $m^6\underline{g}^6A$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

Table S22. Results obtained in the coupling reactions of **ON2c**; **X** = $\underline{v}mnm^5U$ with **ON1r**; **X** = $m^6\underline{g}^6A$ or **ON1s**; **X** = $m^6\underline{g}^6A$ using EDC/Sulfo-NHS as activator.

Donor strand	Acceptor strand	Yield (%)[a]
ON1r; **X** = $m^6\underline{g}^6A$	**ON2c**; **X** = $\underline{v}mnm^5U$	~14 (~35)[b]
ON1s; **X** = $m^6\underline{g}^6A$		< 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.

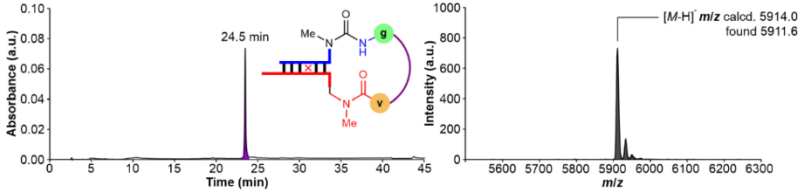


Figure S46. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product from the reaction of **ON2c**; **X** = $\underline{v}mnm^5U$ with **ON1r**; **X** = $m^6\underline{g}^6A$.

Scheme S12. Annealing and coupling reaction of **ON1d**; **X** = $m^6\underline{l}^6A$, **ON1s**; **X** = $m^6\underline{g}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$. The formed peptide bonds are marked in purple.

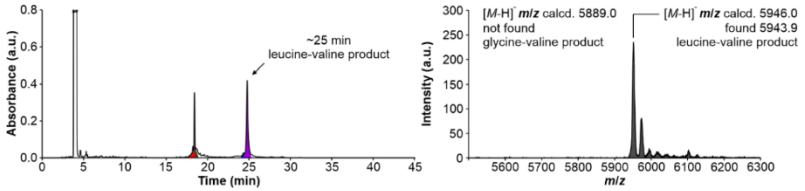


Figure S47. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the reaction of **ON1d**; **X** = $m^6\underline{i}^6A$, **ON1s**; **X** = $m^6\underline{g}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

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 $\underline{iv}$-peptide (%)[a]	Yield of $\underline{gv}$-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.

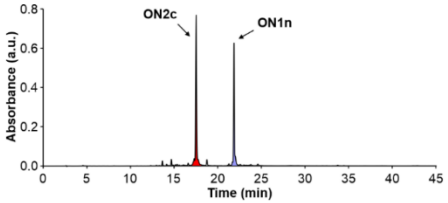


Figure S48. HPL-chromatogram of an equimolar mixture of **ON1n**; **X** = $m^6\underline{v}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$.

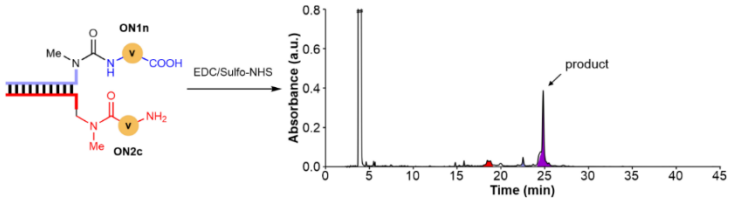


Figure S49. HPL-chromatogram of the reaction of **ON1n**; **X** = $m^6\underline{v}^6A$ with **ON2c**; **X** = $\underline{v}mnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

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).

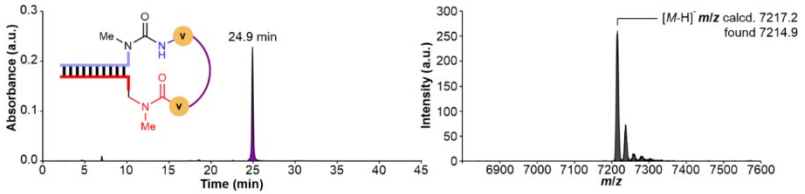
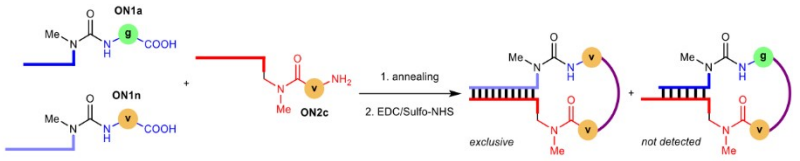


Figure S50. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product from the reaction of **ON1n**; **X** = $m^6\underline{v}^6A$ with **ON2c**; **X** = $\underline{v}mnm^5U$.



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.

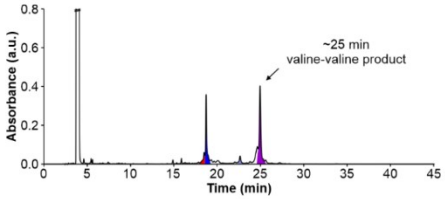


Figure S51. HPL-chromatogram of the reaction of **ON1a**; **X** = $m^6\underline{g}^6A$, **ON1n**; **X** = $m^6\underline{v}^6A$ and **ON2c**; **X** = $\underline{v}mnm^5U$ in MES buffer at pH 6 using EDC/Sulfo-NHS as activator.

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 $\underline{vv}$-peptide (%)[a]	Yield of $\underline{gv}$-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).

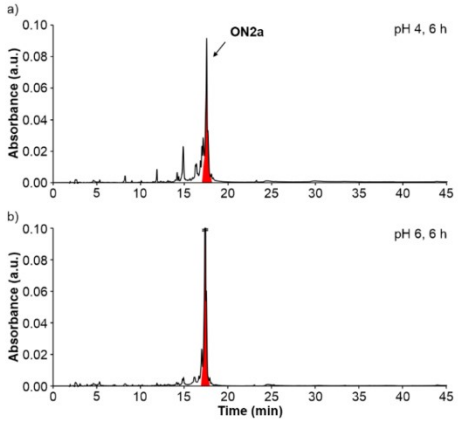


Figure S52. HPL-chromatograms of the stability of **ON2a**; **X** = mnm^{5}U in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 6 h at 90°C.

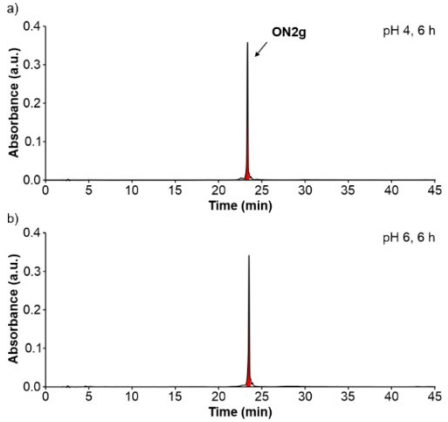


Figure S53. HPL-chromatograms of the stability of **ON2g**; **X** = mnm^{5}U in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 6 h at 90°C.

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

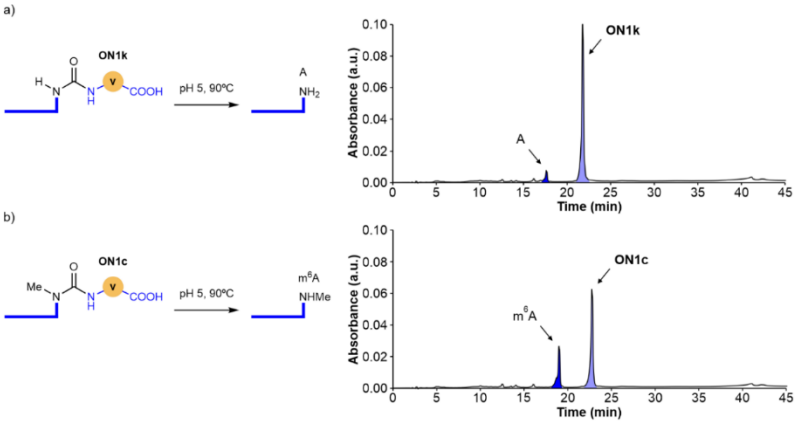


Figure S54. HPL-chromatograms of the cleavage reactions of: a) **ON1k**; **X** = v^{6}A and b) **ON1c**; **X** = m^{6}v^{6}A in acetate buffer at pH 5 after 12 h at 90°C.

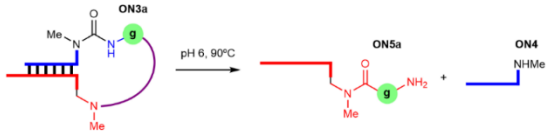
These experiments indicated that the urea cleavage reaction of the unmethylated aa^{6}A-RNA donor strand **ON1k** was slower than that of the methylated version, m^{6}aa^{6}A-RNA **ON1c**.

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^6g^6A coupled with mnm^5U)



Scheme S14. Cleavage of urea in **ON3a**. The peptide bond is marked in purple.

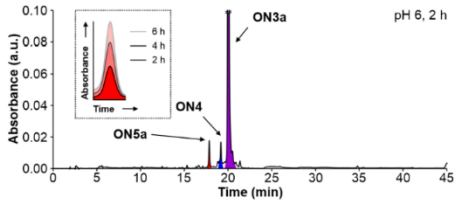


Figure S55. HPL-chromatogram of the cleavage reaction of **ON3a** in MES buffer at pH 6 after 2 h at 90°C. Inset shows the selected region of the HPL-chromatograms after 2, 4 and 6 h.

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^6A)**	**ON5a ($gmnm^5U$)**
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.

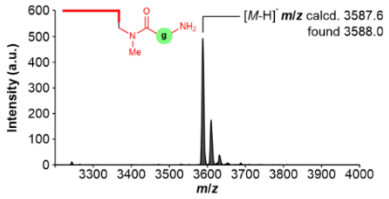


Figure S56. MALDI-TOF mass spectrum (negative mode) of the isolated **ON5a** ($gmnm^5U$).

Additional experiments at pH 4 and pH 6

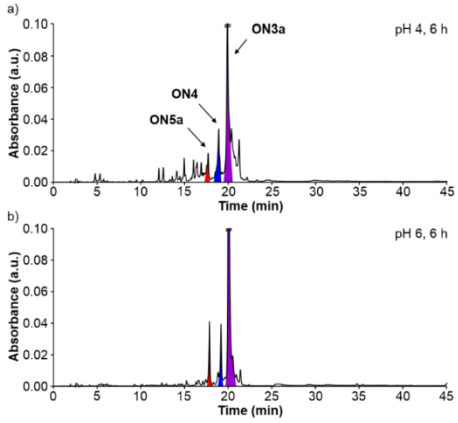


Figure S57. HPL-chromatograms of the cleavage reactions of **ON3a** in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 6 h at 90°C.

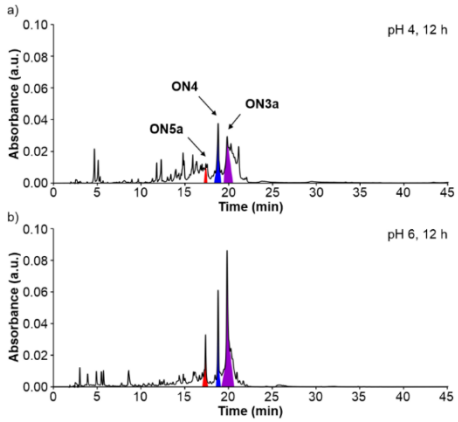


Figure S58. HPL-chromatograms of the cleavage reactions of **ON3a** in: a) acetate buffer at pH 4 and b) MES buffer at pH 6 after 12 h at 90°C.

Table S29. Results obtained in the cleavage reaction of **ON3a** (average of, at least, two experiments).[a]

		Average Amount ± Error (%)
pH	**Time (h)**	**ON5a ($gmnm^5U$)**
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^6g^6A coupled with $vmnm^5U$)

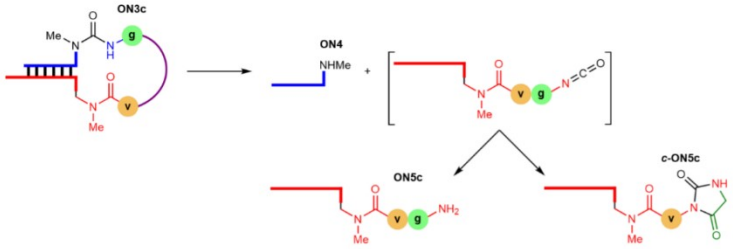


Figure S59. Cleavage of urea in **ON3c**. The peptide bond is marked in purple.

Cleavage reactions at 60°C

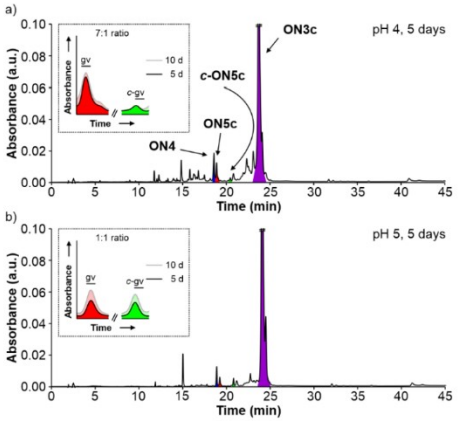


Figure S60. HPL-chromatograms of the cleavage reactions of **ON3c** in acetate buffer at: a) pH 4 and b) pH 5 after 5 days at 60°C. Inset shows the selected region of the HPL-chromatograms after 5 and 10 days.

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^6A)**	**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

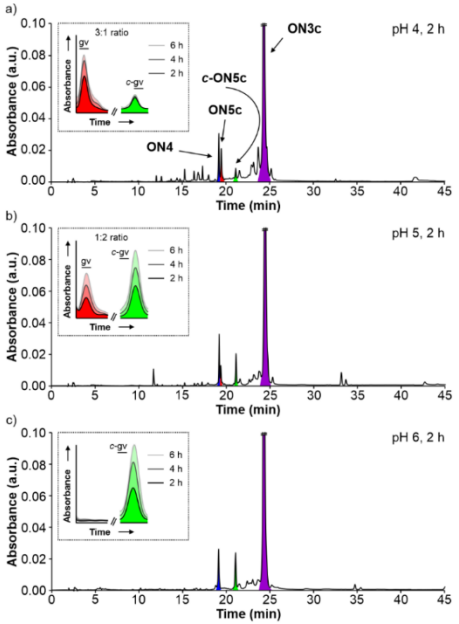


Figure S61. HPL-chromatograms of the cleavage reactions of **ON3c** in: a) acetate buffer at pH 4; b) acetate buffer at pH 5 and c) MES buffer at pH 6 after 2 h at 90°C. Inset shows the selected region of the HPL-chromatograms after 2, 4 and 6 h.

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^6A)**	**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.

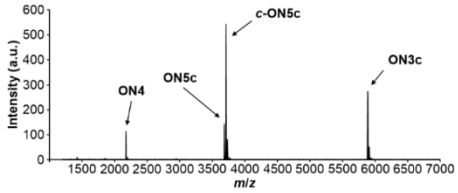


Figure S62. MALDI-TOF mass spectrum (negative mode) of the cleavage reaction of **ON3c** in acetate buffer at pH 5 after 2 h at 90°C. A similar MALDI-TOF mass spectrum was obtained at pH 4. The indicated peaks correspond to the [*M*-H]$^-$ ions.

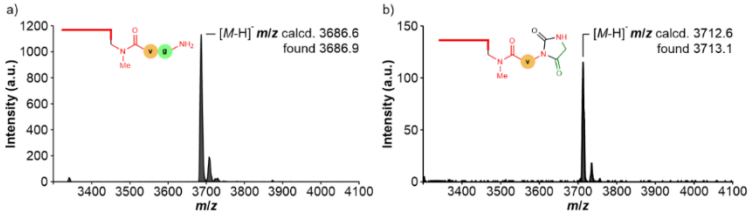
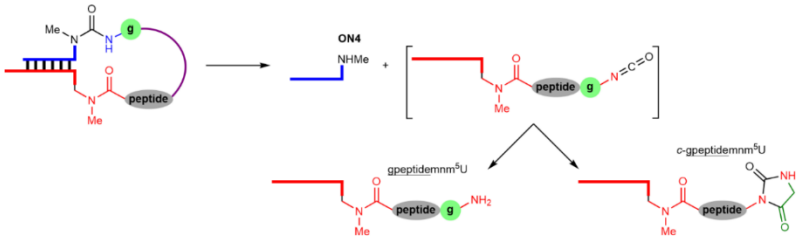


Figure S63. MALDI-TOF mass spectrum (negative mode) of the isolated: a) **ON5c** (gvmnm^{5}U) and b) ***c*-ON5c** (*c*-gvmnm^{5}U).

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.

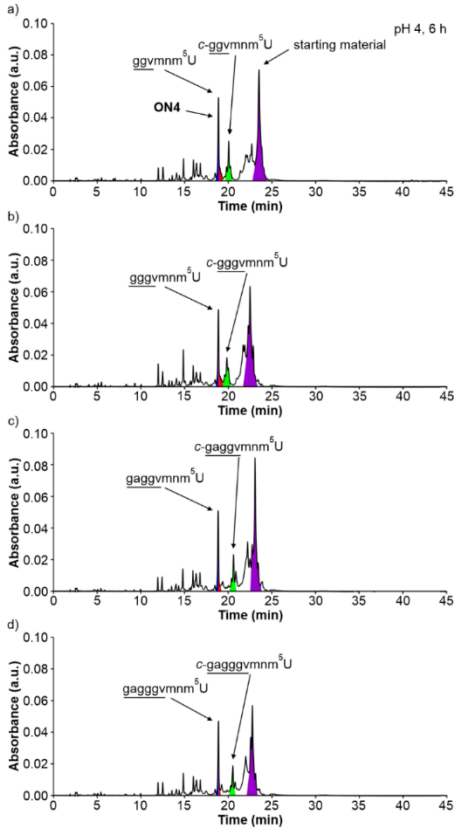


Figure S64. HPL-chromatograms of the cleavage reactions of peptide-oligonucleotides (Section 8.2) in acetate buffer at pH 4 to give: a) ggvmnm^{5}U; b) gggvmnm^{5}U; c) gaggvmnm^{5}U and d) gagggvmnm^{5}U oligonucleotides, together with hydantoin side products, after 6 h at 90°C.

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.

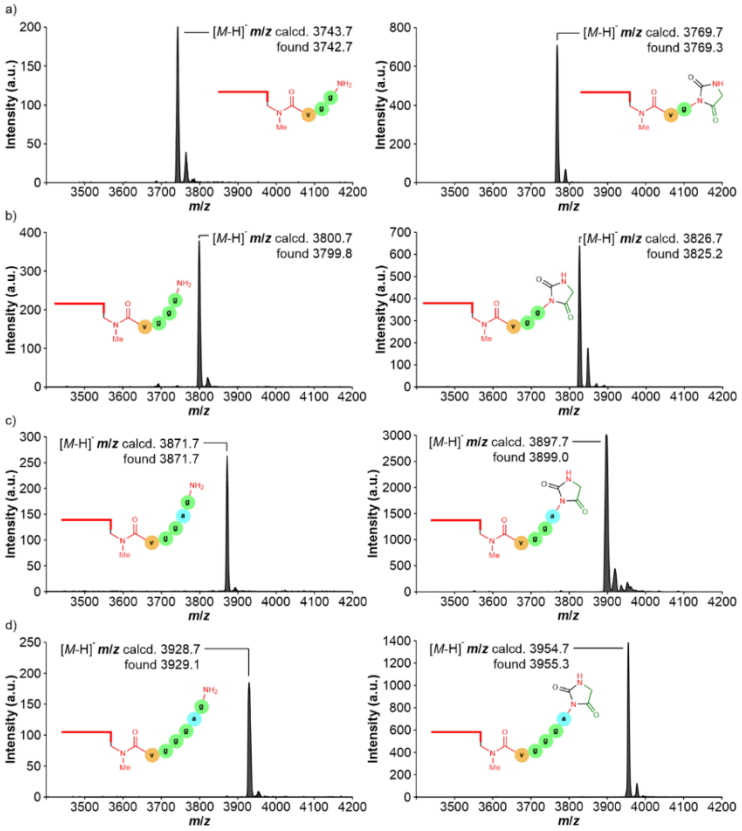


Figure S65. MALDI-TOF mass spectra (negative mode) of the isolated: a) ggvmnm^{5}U; b) gggvmnm^{5}U; c) gaggvmnm^{5}U and d) gagggvmnm^{5}U oligonucleotides (left) and hydantoin side products (right). Note that the analyzed 3'-H_2N-peptidemnm^{5}U-RNA-5' samples (left) contained the m^{6}A product (*m/z* region not shown).

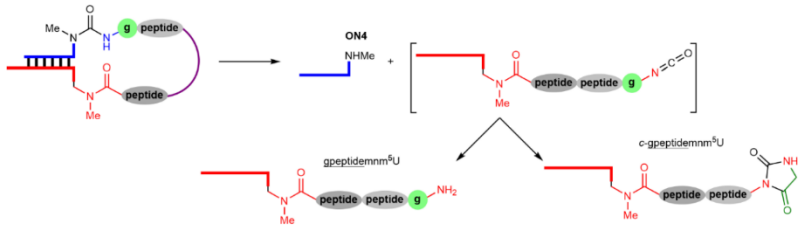


Figure S66. Cleavage of urea in gpeptide-peptidemnm^{5}U-oligonucleotides. The peptide bond is marked in purple.

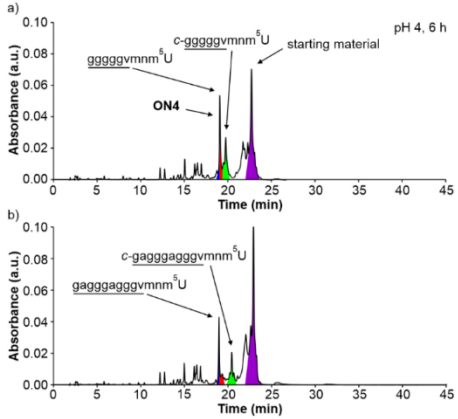


Figure S67. HPL-chromatograms of the cleavage reactions of peptide-oligonucleotides (Section 8.3) in acetate buffer at pH 4 to give: a) gggggvmnm^{5}U and b) gagggagggvmnm^{5}U oligonucleotides, together with hydantoin side products, after 6 h at 90°C.

The 3'-H_2N-peptidemnm^{5}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.

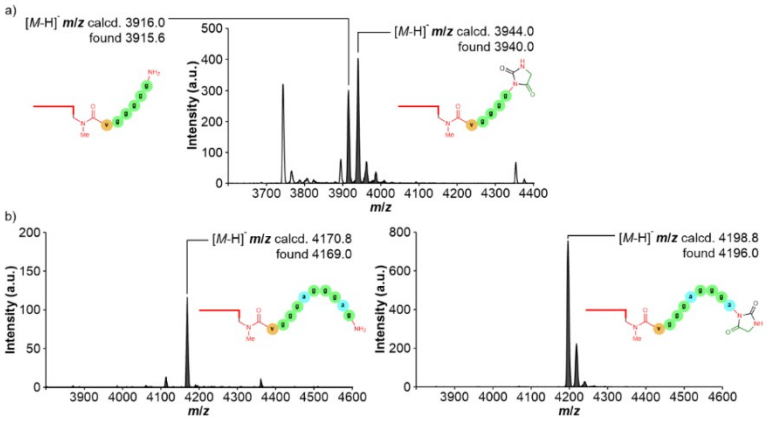


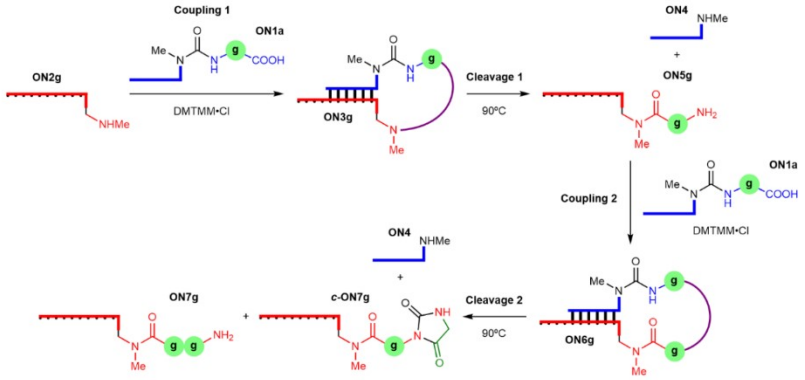
Figure S68. MALDI-TOF mass spectra (negative mode) of the isolated: a) ggggg**v**mnm^{5}U and b) gagggagggvmnm^{5}U. Hydantoin side products are also shown. Note that the analyzed 3'-H_2N-peptidemnm^{5}U-RNA-5' samples contained the m^{6}A product (*m/z* region not shown).

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.

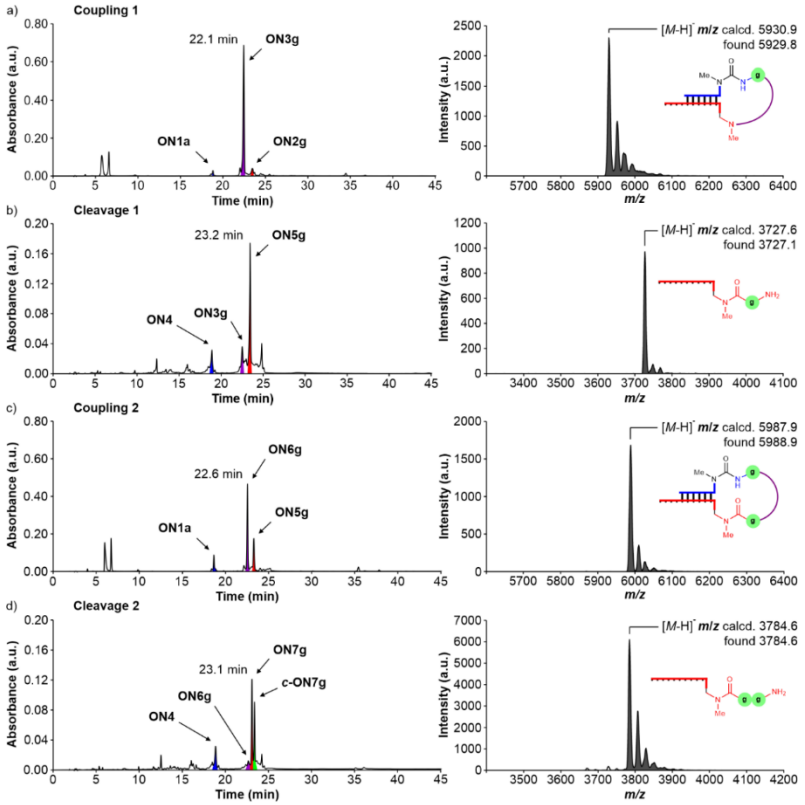


Figure S69. left) HPL-chromatograms of the reactions of **ON1a**; **X** = m^{6}g^{6}A with **ON2g**: a) coupling 1; b) cleavage 1; c) coupling 2 and d) cleavage 2. The product of each step was separated by HPLC and added into the next reaction. right) MALDI-TOF mass spectra (negative mode) of the isolated 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.

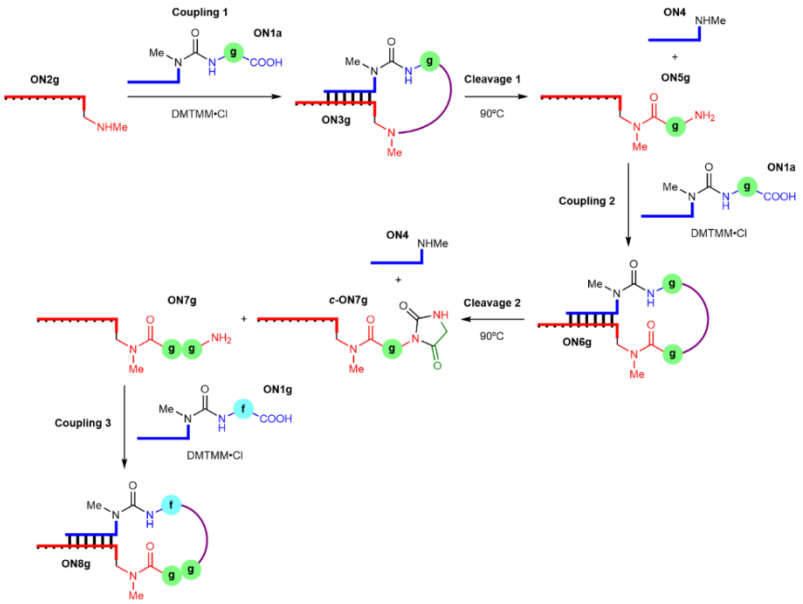


Figure S70. Coupling and cleavage of **ON1a**; **X** = m^6g^6A and **ON1g**; **X** = m^6f^6A with **ON2g**. The formed peptide bond is marked in purple.

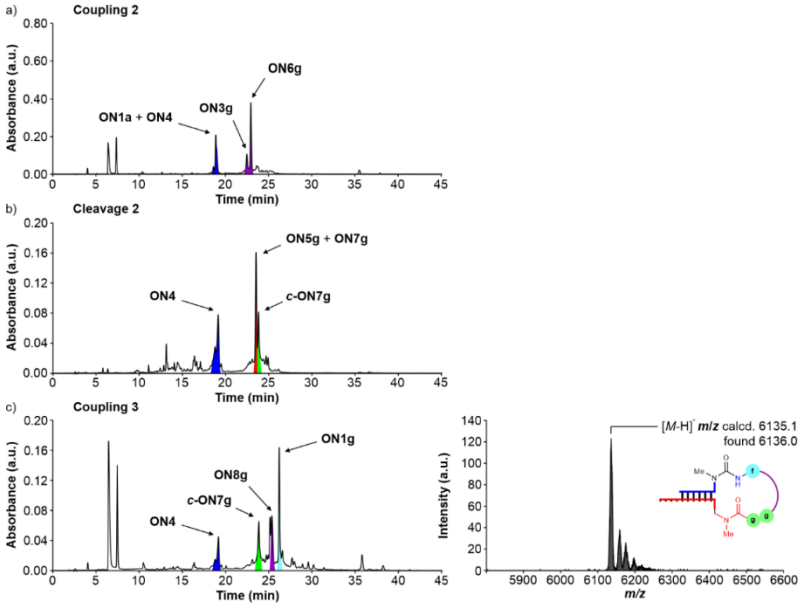


Figure S71. HPL-chromatograms of the one pot reaction of **ON1a**; **X** = m^6g^6A and **ON1g**; **X** = m^6f^6A with **ON2g**: a) coupling 2; b) cleavage 2 and c) coupling 3.

Table S35. Results obtained in the one pot reaction of **ON1a**; **X** = m^6g^6A and **ON1g**; **X** = m^6f^6A 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

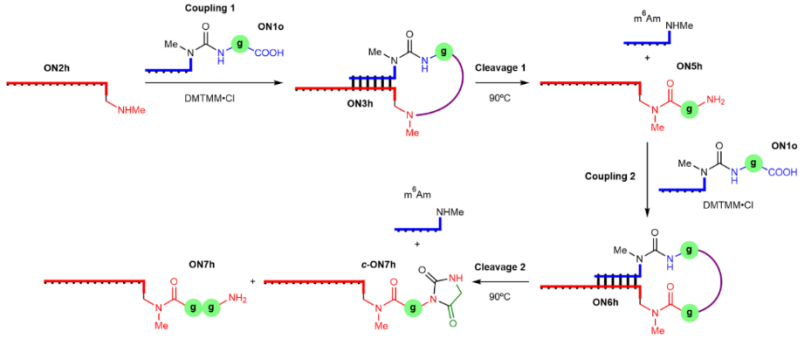


Figure S72. Coupling and cleavage of **ON1o**; **X** = m^6g^6Am with **ON2h**. The formed peptide bond is marked in purple.

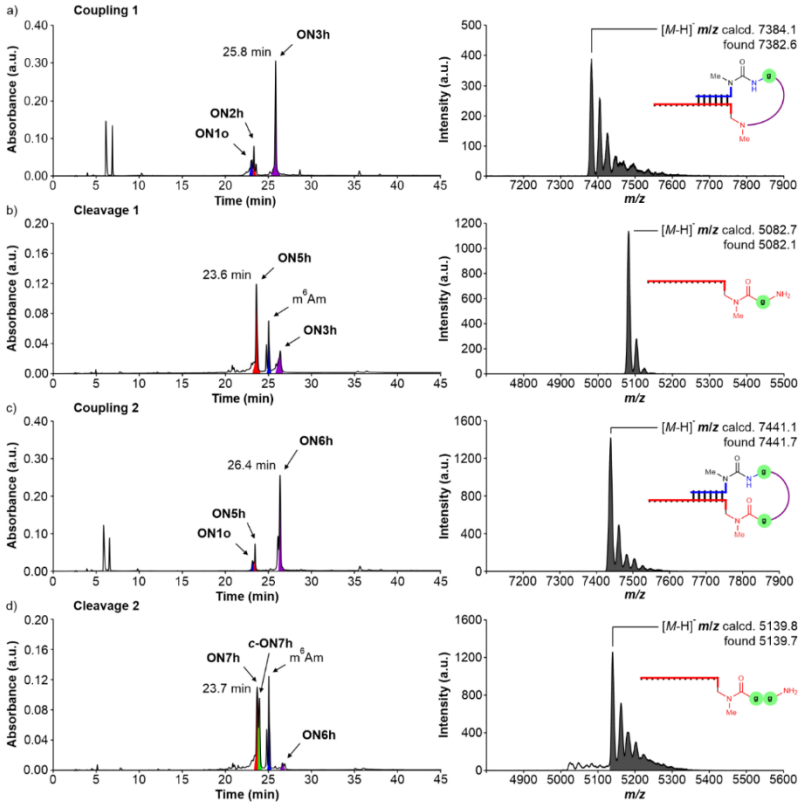


Figure S73. left) HPL-chromatograms of the reactions of **ON1o**; **X** = m^{6}g^{6}Am with **ON2h**: a) coupling 1; b) cleavage 1; c) coupling 2 and d) cleavage 2. The product of each step was separated by HPLC and added into the next reaction. right) MALDI-TOF mass spectra (negative mode) of the isolated 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

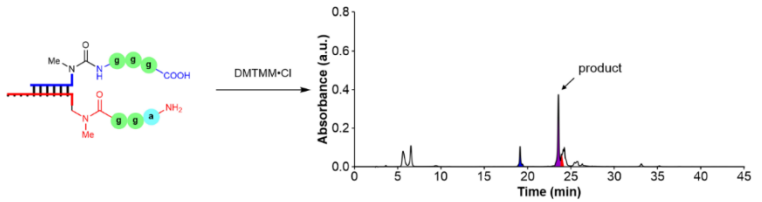


Figure S74. HPL-chromatogram of the reaction of 5'-m^6(ggg)^{6}A-RNA-3' with 3'-aggmnm^{5}U-RNA-5' containing 2'-OMe nucleosides in MES buffer at pH 6 using DMTMM•Cl as activator.

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**.

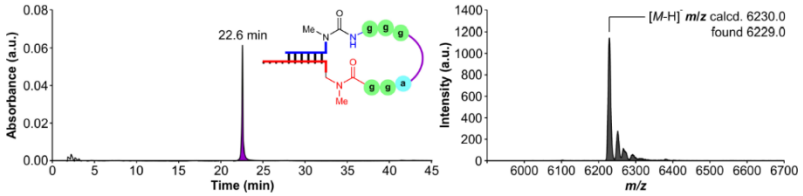
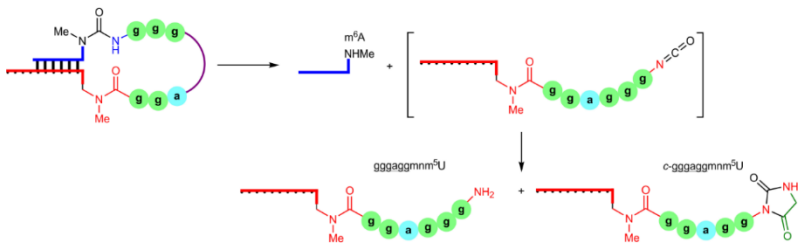


Figure S75. left) HPL-chromatogram and right) MALDI-TOF mass spectrum (negative mode) of the isolated product from the reaction of 5'-m^6(ggg)^{6}A-RNA-3' with 3'-aggmnm^{5}U-RNA-5' containing 2'-OMe nucleosides.



Scheme S17. Cleavage of urea in peptide-peptidemnm^{5}U-oligonucleotide. The peptide bond is marked in purple.

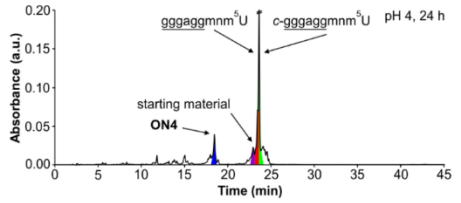


Figure S76. HPL-chromatogram of the cleavage reaction of peptide-oligonucleotide in acetate buffer at pH 4 to give gggaggmnm^{5}U and c-gggaggmnm^{5}U oligonucleotides after 24 h at 90°C.

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**.

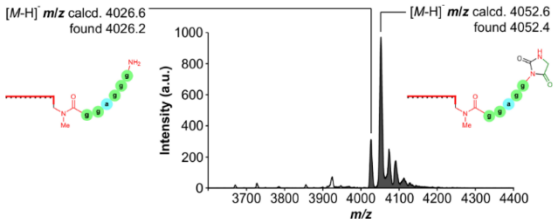


Figure S77. MALDI-TOF mass spectra (negative mode) of the isolated gggaggmnm^{5}U and hydantoin side product.

15.4 Coupling reactions between ON2g and donor oligonucleotides of different length

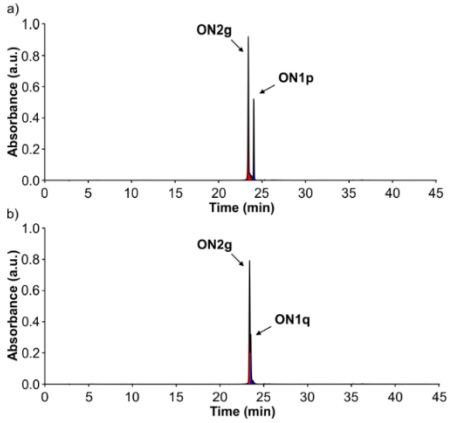
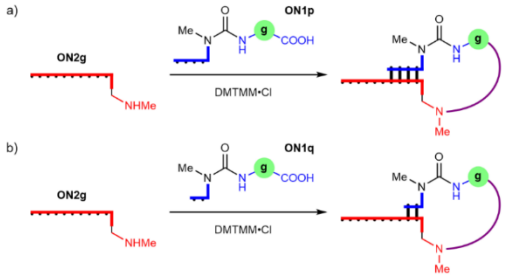


Figure S78. HPL-chromatograms of an equimolar mixture of **ON2g**; **X** = mnm^{5}U with: a) **ON1p**; **X** = m^{6}g^{6}Am and b) **ON1q**; **X** = m^{6}g^{6}Am.



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.

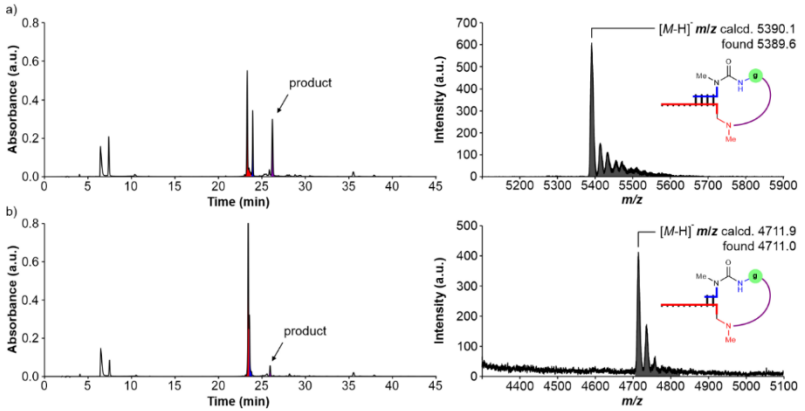


Figure S79. left) HPL-chromatograms of the reactions of: a) **ON1p**; **X** = m^{6}g^{6}Am and b) **ON1q**; **X** = m^{6}g^{6}Am with **ON2g**; **X** = mnm^{5}U in MES buffer at pH 6 using DMTMM•Cl as activator. The reaction b) was carried out at 0°C using 1 M NaCl. right) MALDI-TOF mass spectra (negative 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

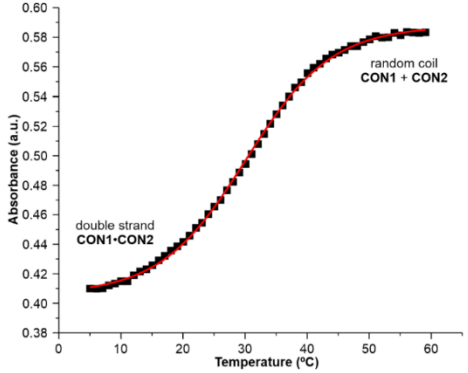


Figure S80. Melting curve of **CON1** and **CON2**. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.1°C.

16.2 Melting temperatures of double strands from donor and acceptor oligonucleotides

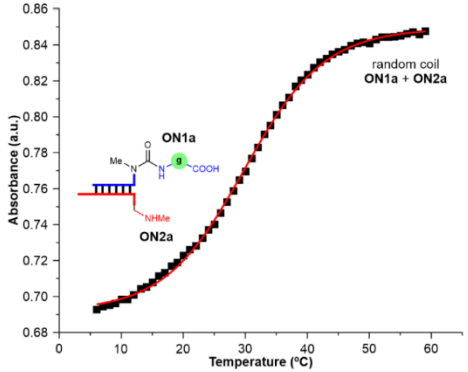


Figure S81. Melting curve of **ON1a**; **X** = m^{6}g^{6}A and **ON2a**; **X** = mnm^{5}U. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.4°C.

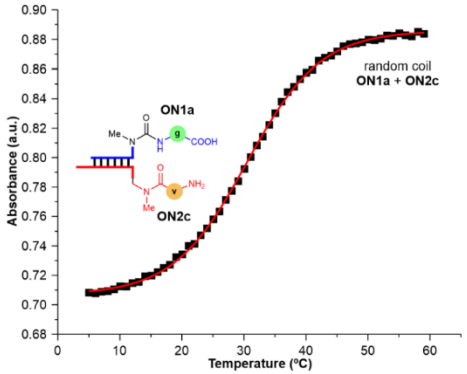


Figure S82. Melting curve of **ON1a**; **X** = m^{6}g^{6}A and **ON2c**; **X** = vmnm^{5}U. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.5°C.

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**.

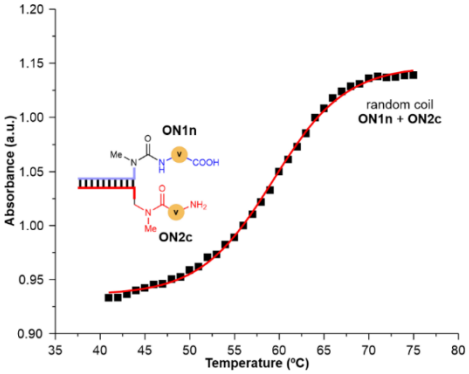


Figure S83. Melting curve of **ON1n**; **X** = m^{6}v^{6}A and **ON2c**; **X** = vmnm^{5}U. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 59.2°C.

16.3 Melting temperatures of double strands from donor and acceptor peptide-oligonucleotides

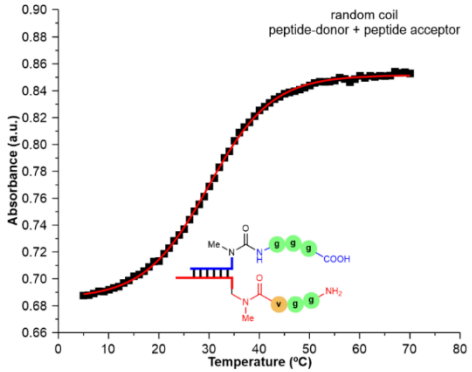


Figure S84. Melting curve of 5'-m^6(ggg)^{6}A-RNA-3' with 3'-ggvmnm^{5}U-RNA-5'. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 30.0°C.

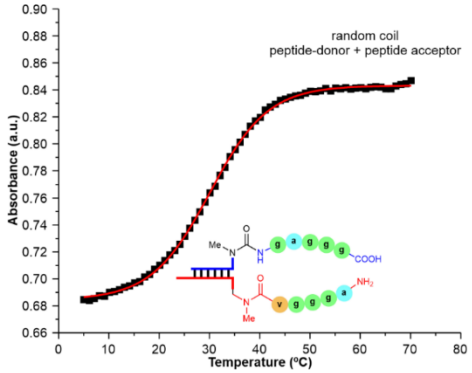


Figure S85. Melting curve of 5'-m^6(gaggg)^{6}A-RNA-3' with 3'-agggvmnm^{5}U-RNA-5'. Line shows the fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function. T_m = 29.8°C.

16.4 Melting temperatures of selected cyclic peptide products

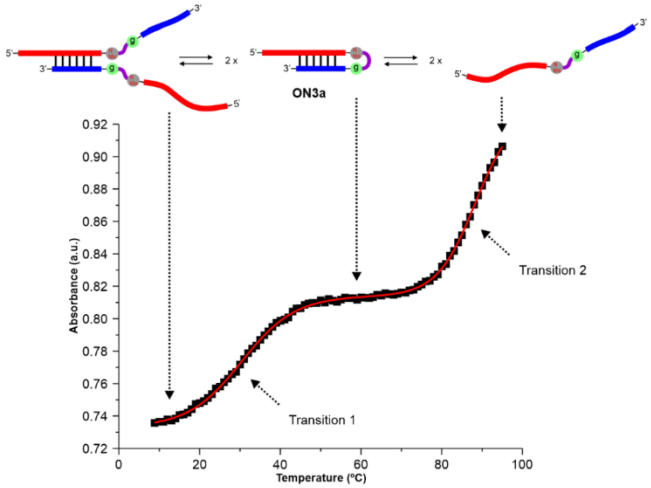


Figure S86. Melting curve of **ON3a**. Line shows the fit of the data to a three-state melting model using a double-sigmoidal Boltzmann function. T_{m1} = 30.8°C and T_{m2} = 87.5 °C. Top panel shows representation of the three states involved in the two transitions.

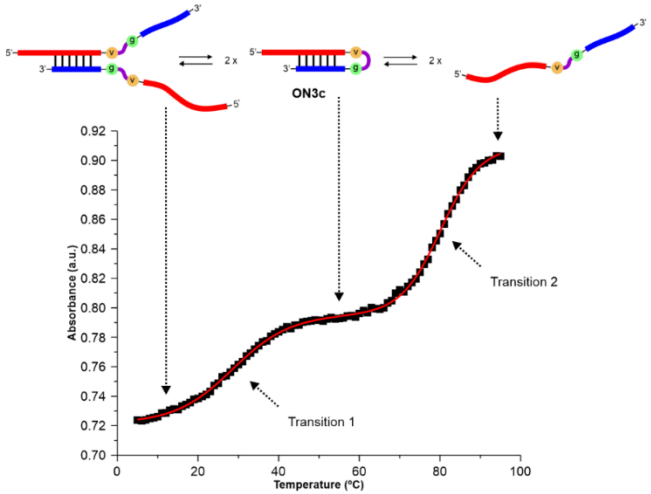


Figure S87. Melting curve of **ON3c**. Line shows the fit of the data to a three-state melting model using a double-sigmoidal Boltzmann function. T_{m1} = 28.4°C and T_{m2} = 80.1°C. Top panel shows representation of the three states involved in the two transitions.

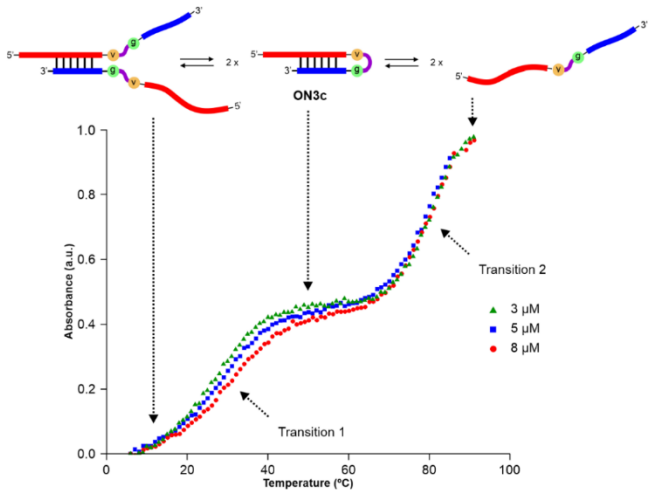
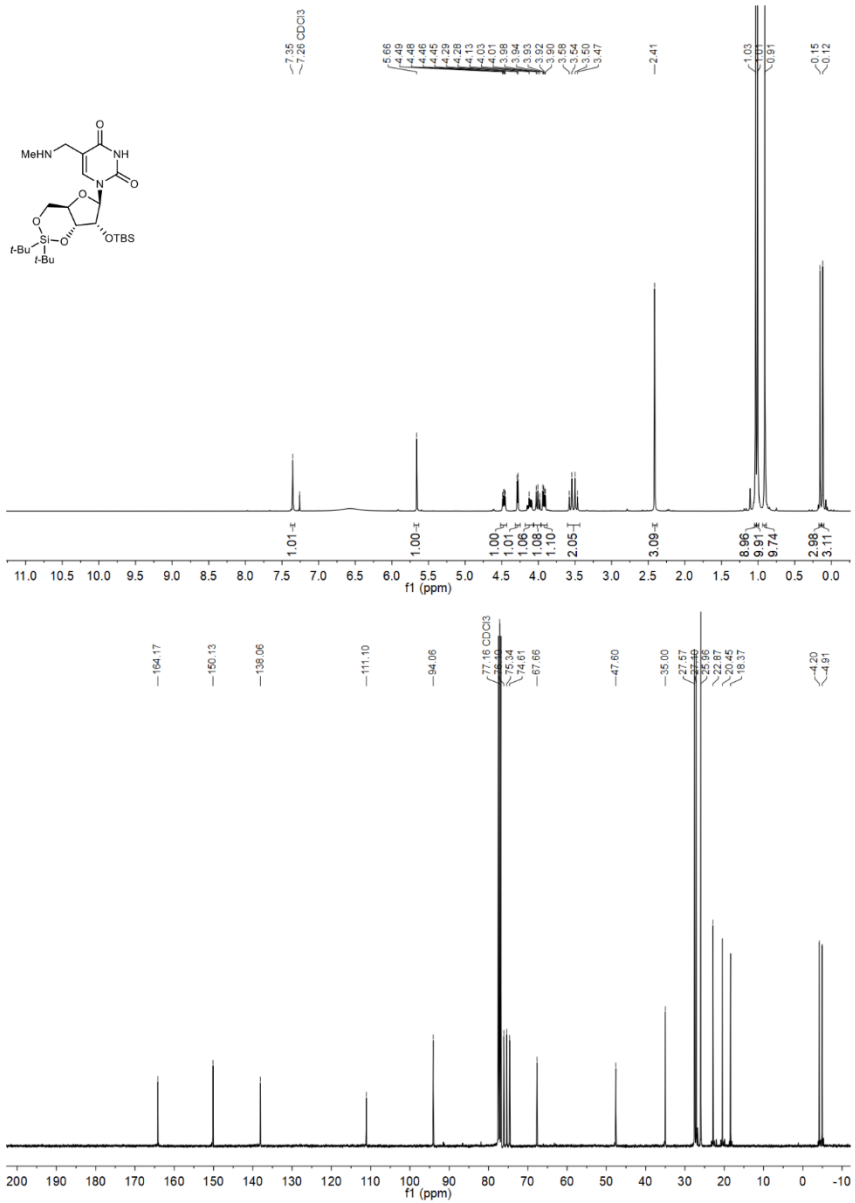


Figure S88. Normalized melting curves of **ON3c** at 3, 5 and 8 μM concentration. Top panel shows representation of the three states involved in the two transitions.

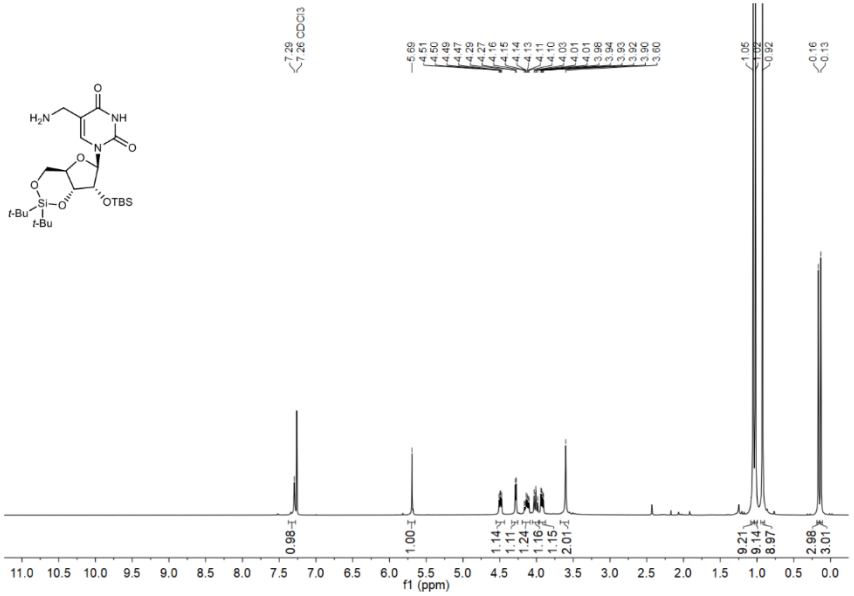
We observed a two-step melting profile in the experiments performed with the RNA oligonucleotides **ON3a** and **ON3c** (**Figure S86**, **Figure S87** and **Figure S88**). At low temperature (transition 1), the double strand (duplex) is transformed into the hairpin. At high temperature (transition 2), 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]

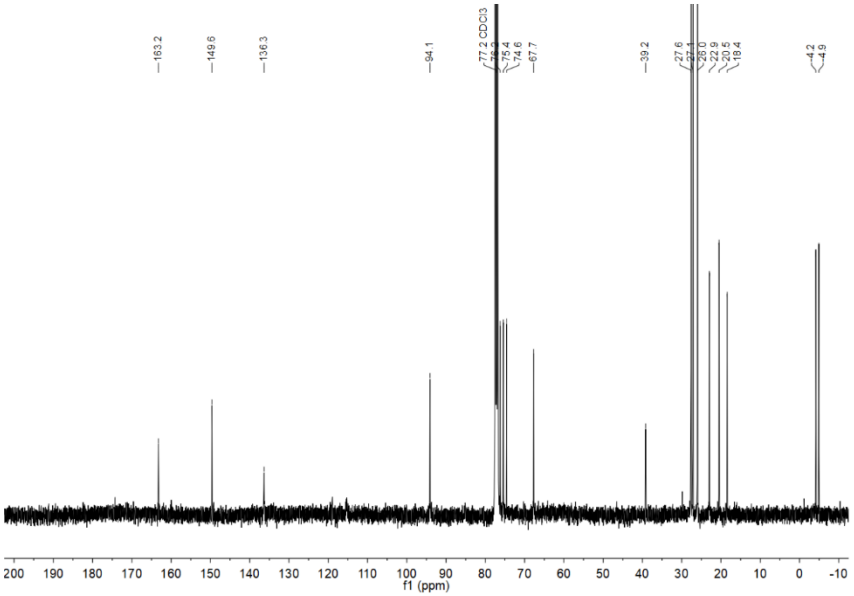
17. NMR spectra of synthesized compounds

^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 3a

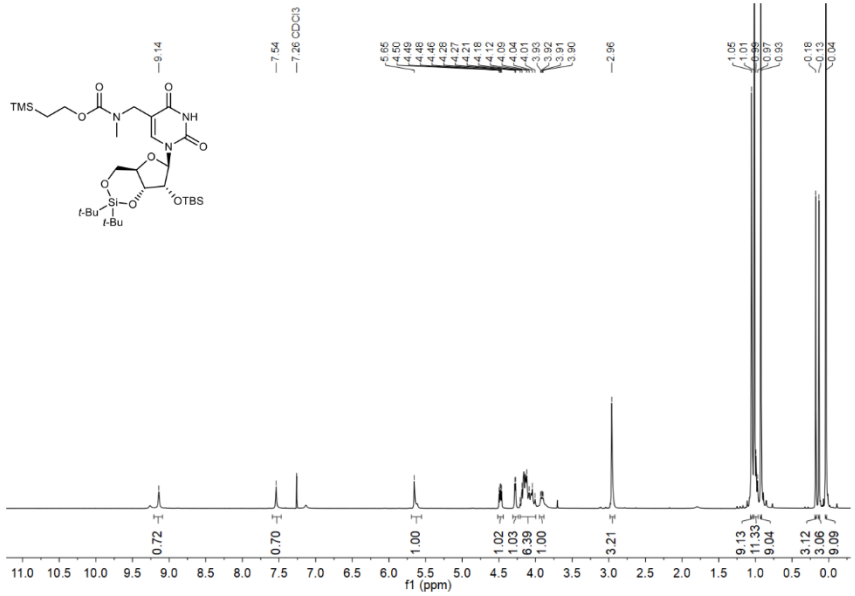


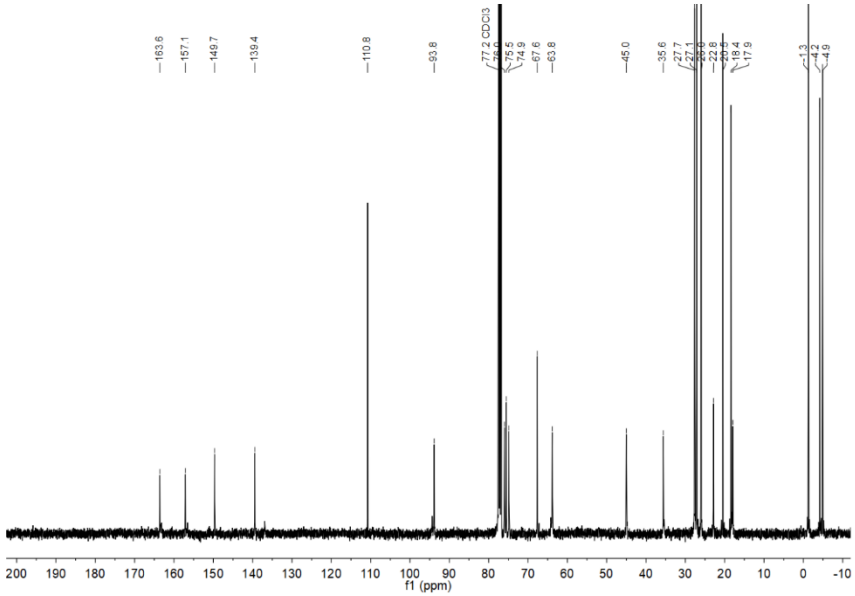
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 3b



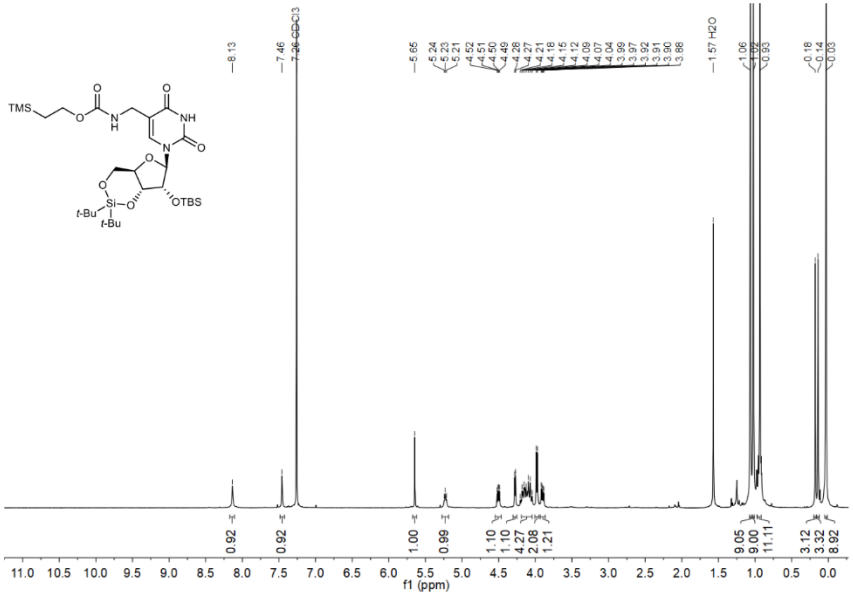


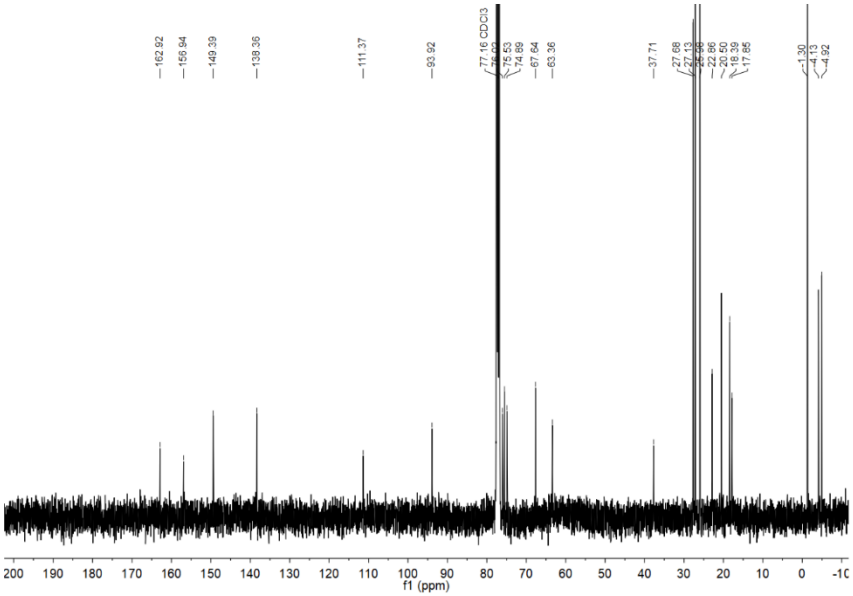
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4a



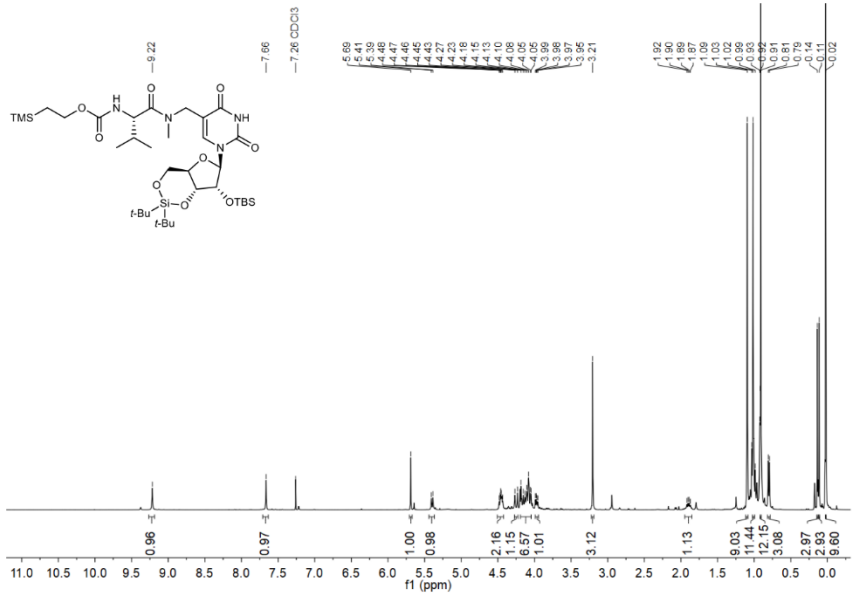


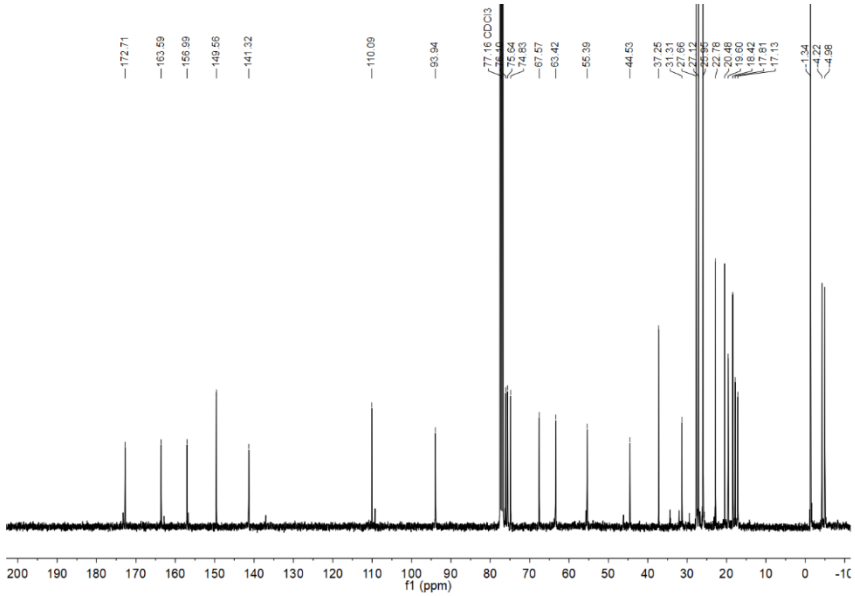
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4b



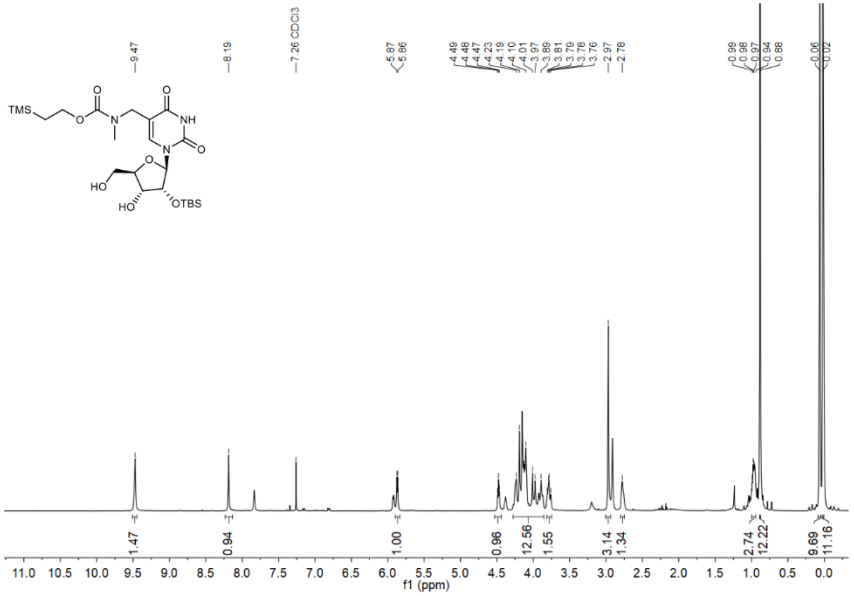


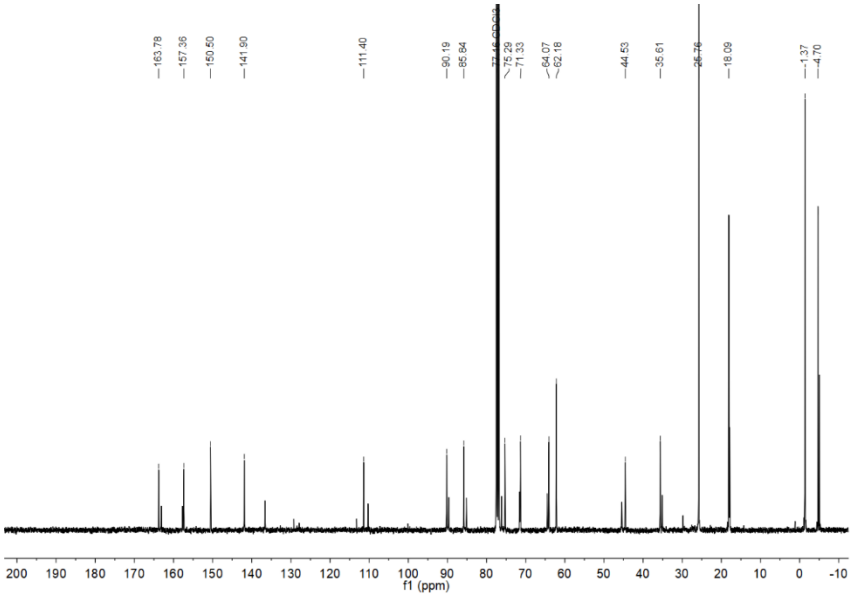
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4c



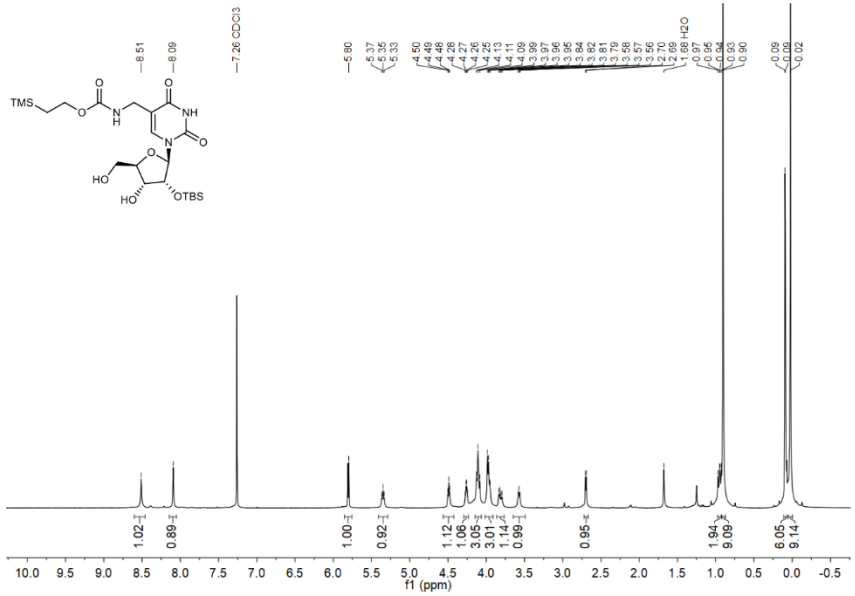


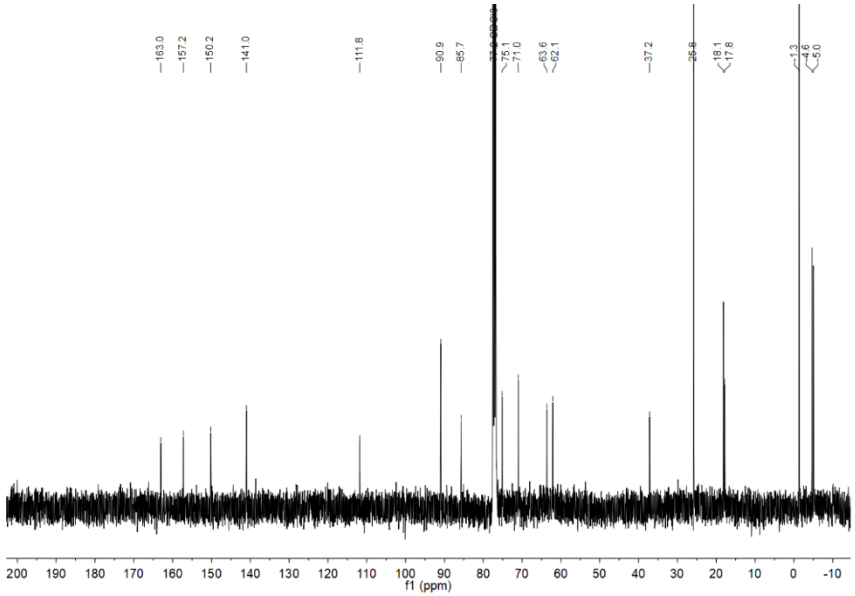
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5a



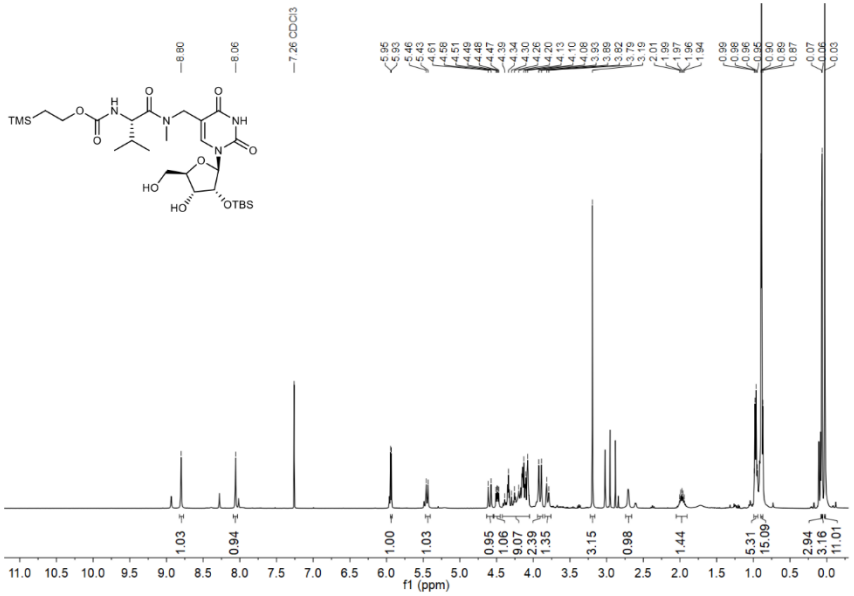


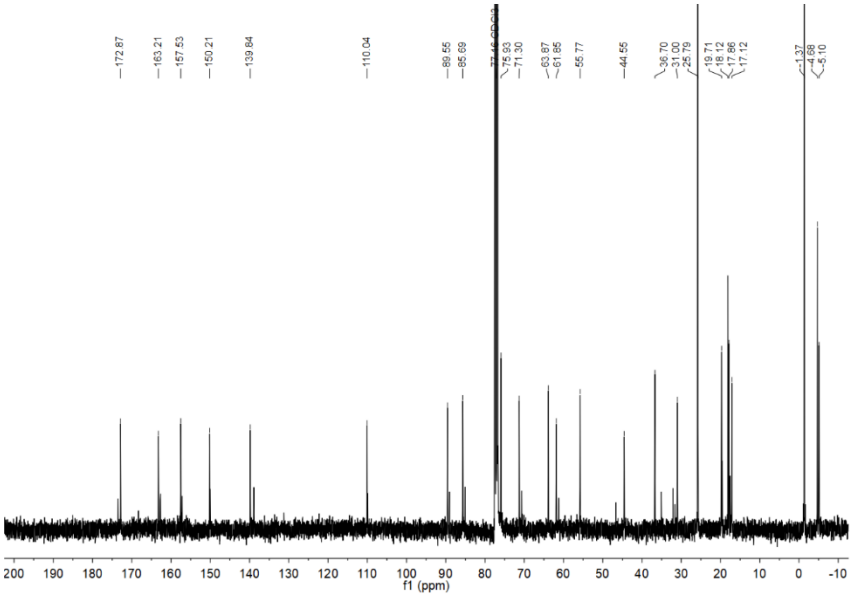
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5b



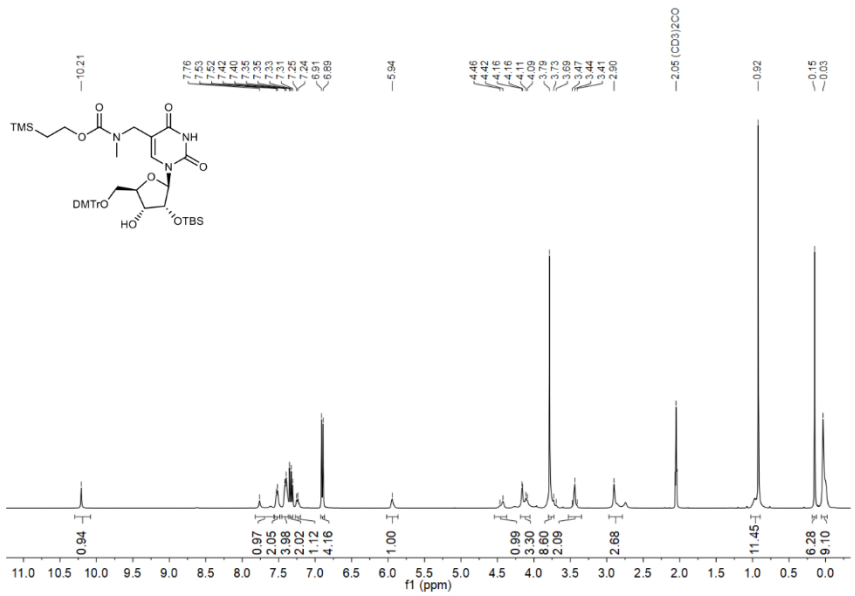


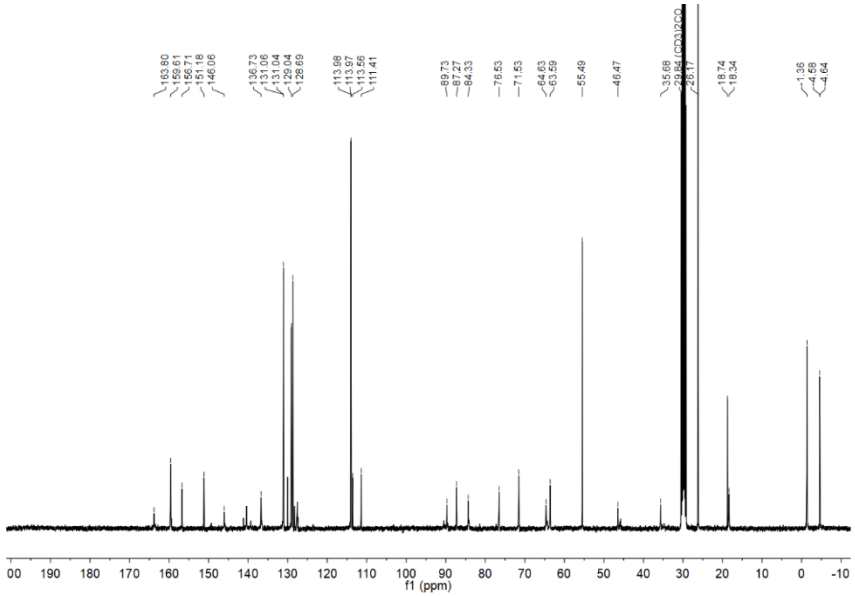
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5c



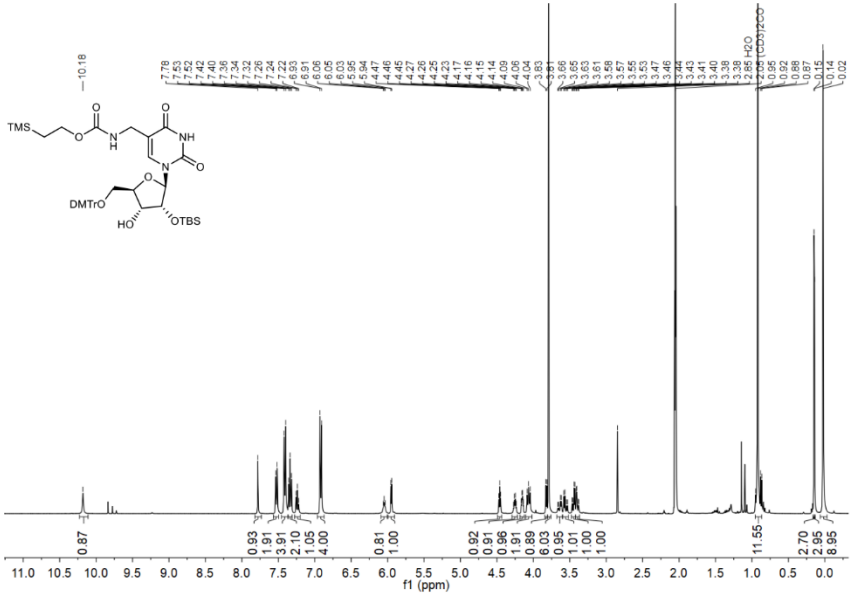


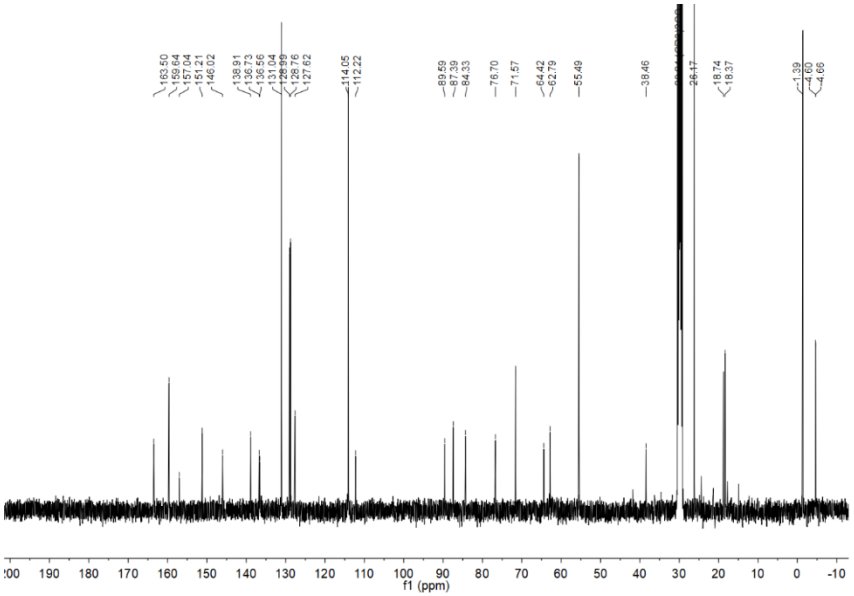
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6a



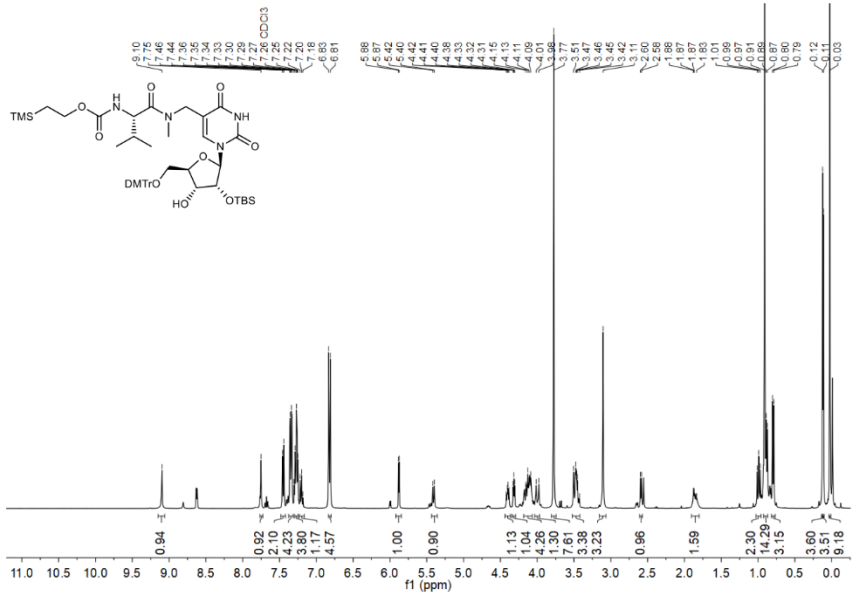


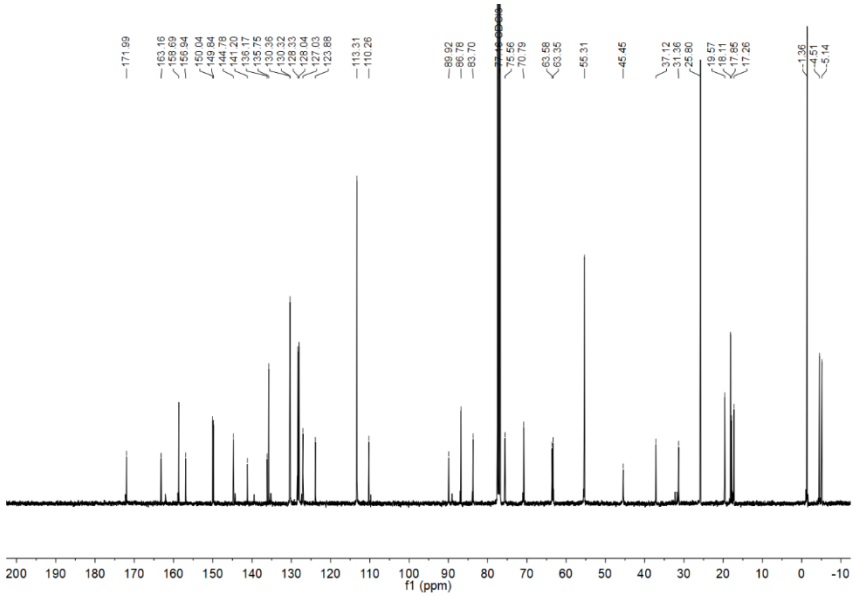
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6b



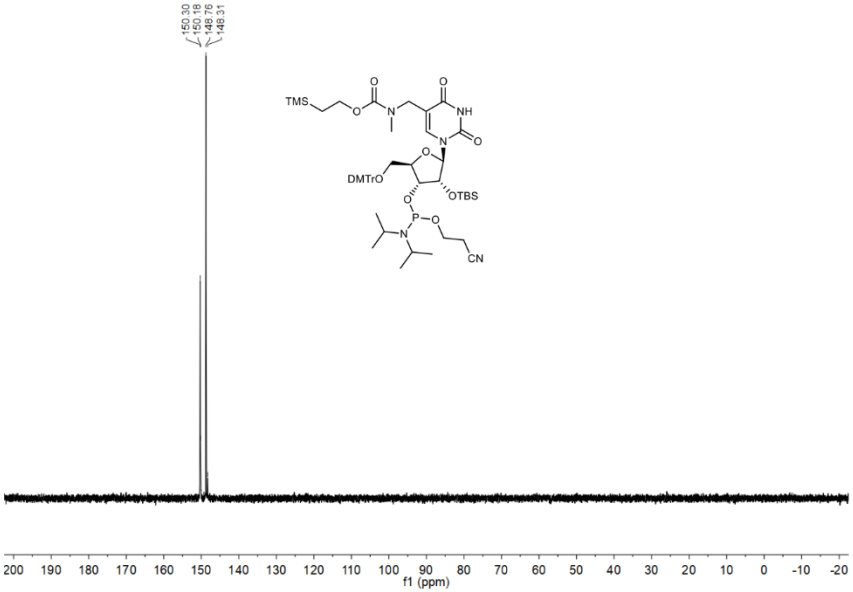


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6c

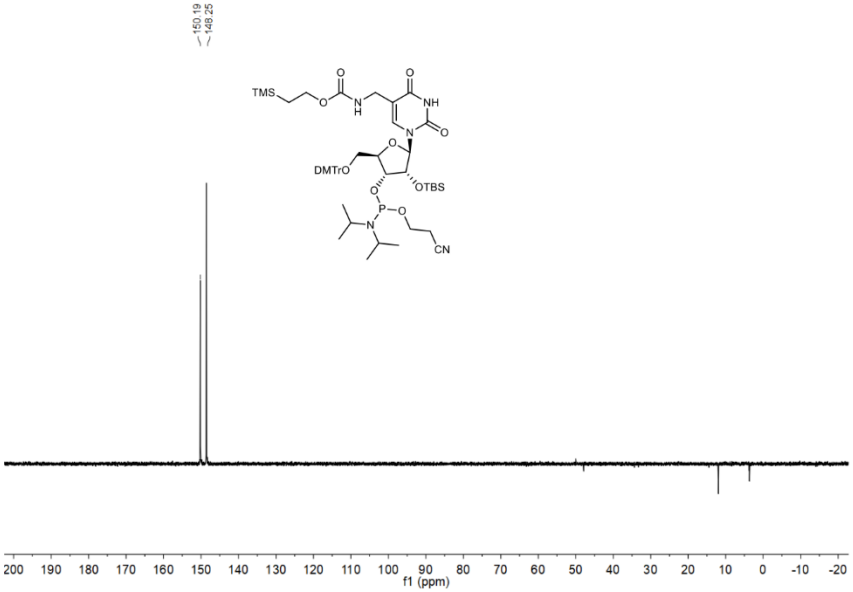




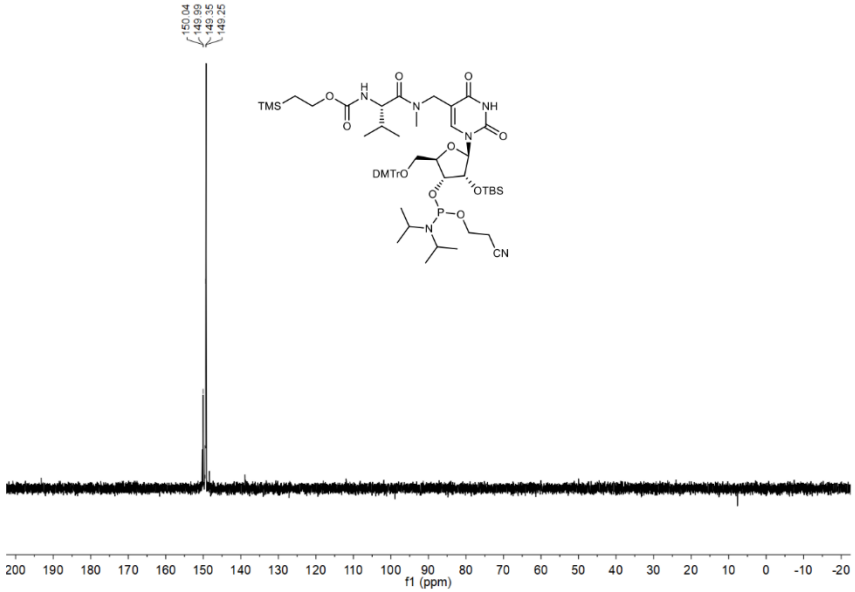
$^{31}P\{^{1}H\}$ NMR spectrum of compound 7a



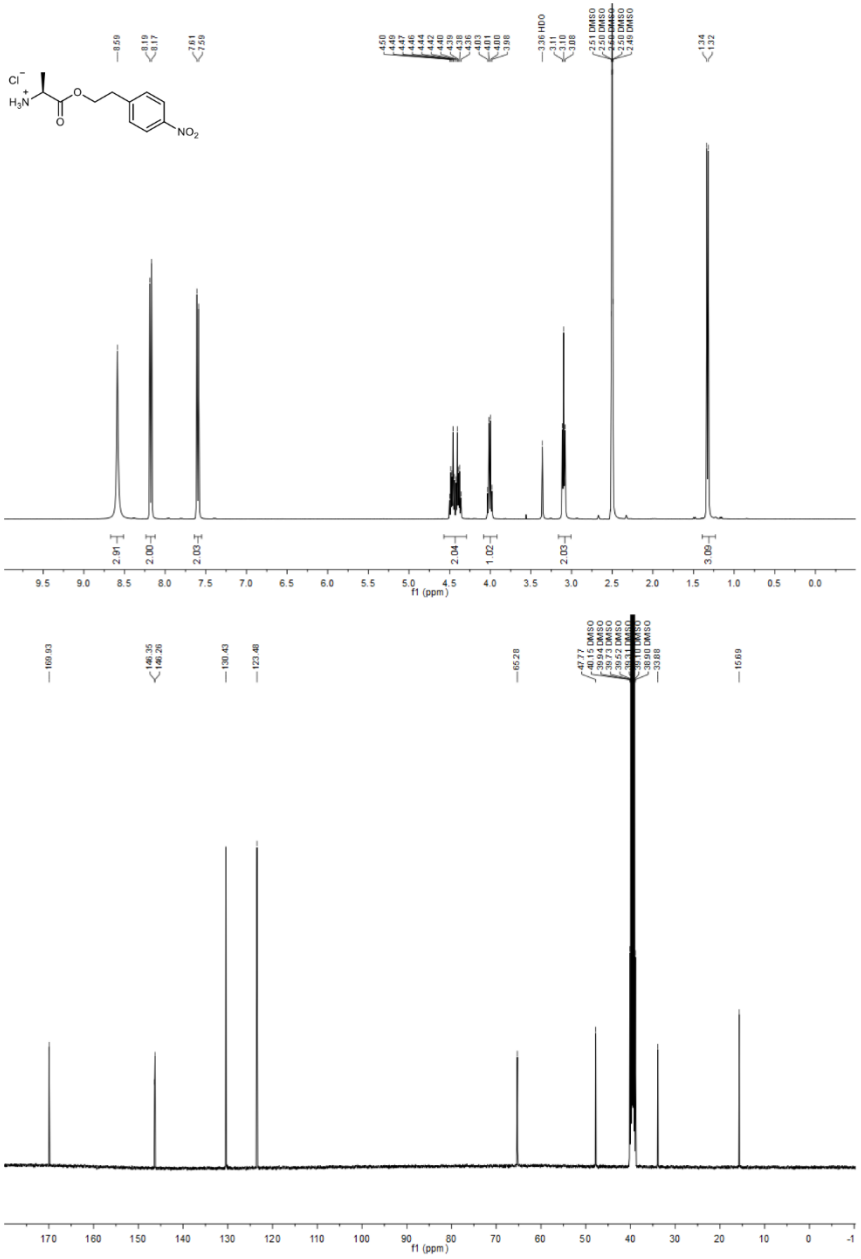
$^{31}P\{^{1}H\}$ NMR spectrum of compound 7b



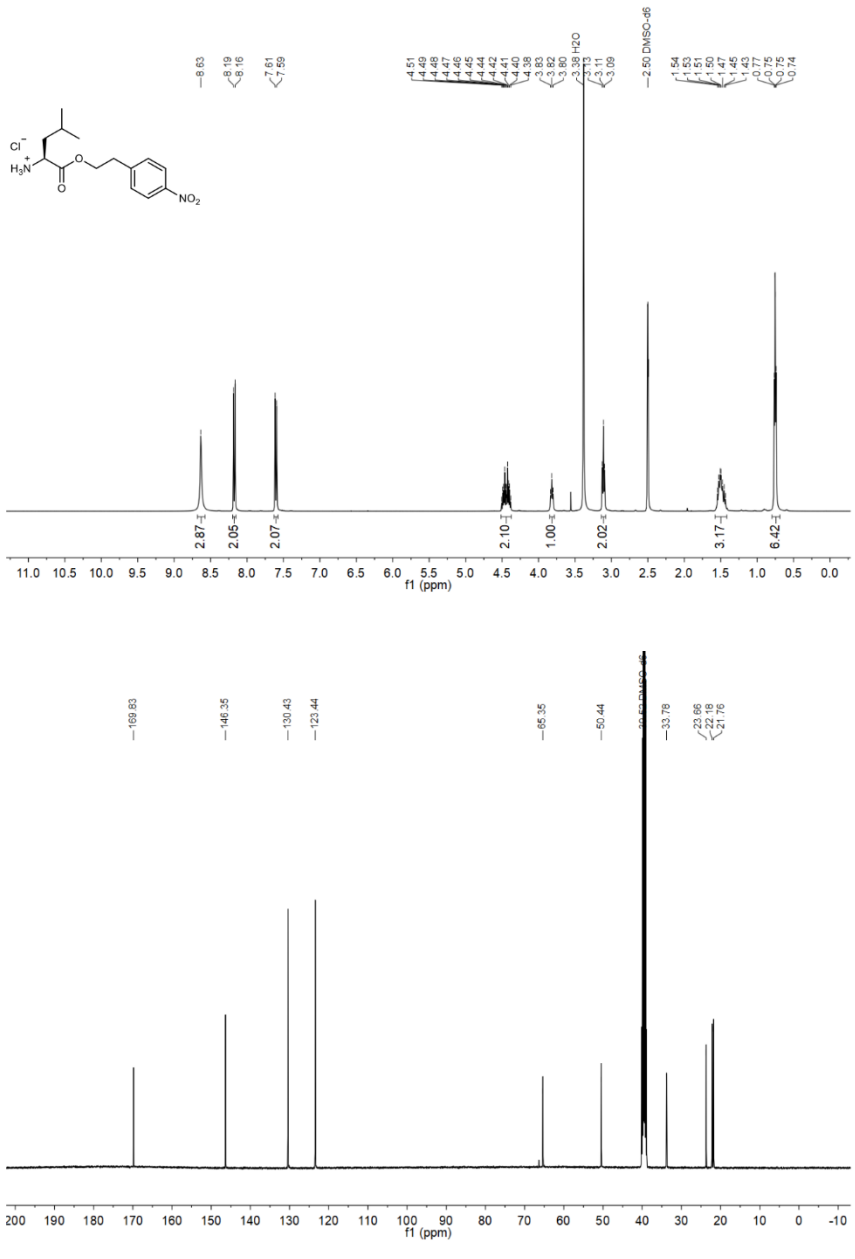
$^{31}P\{^{1}H\}$ NMR spectrum of compound 7c



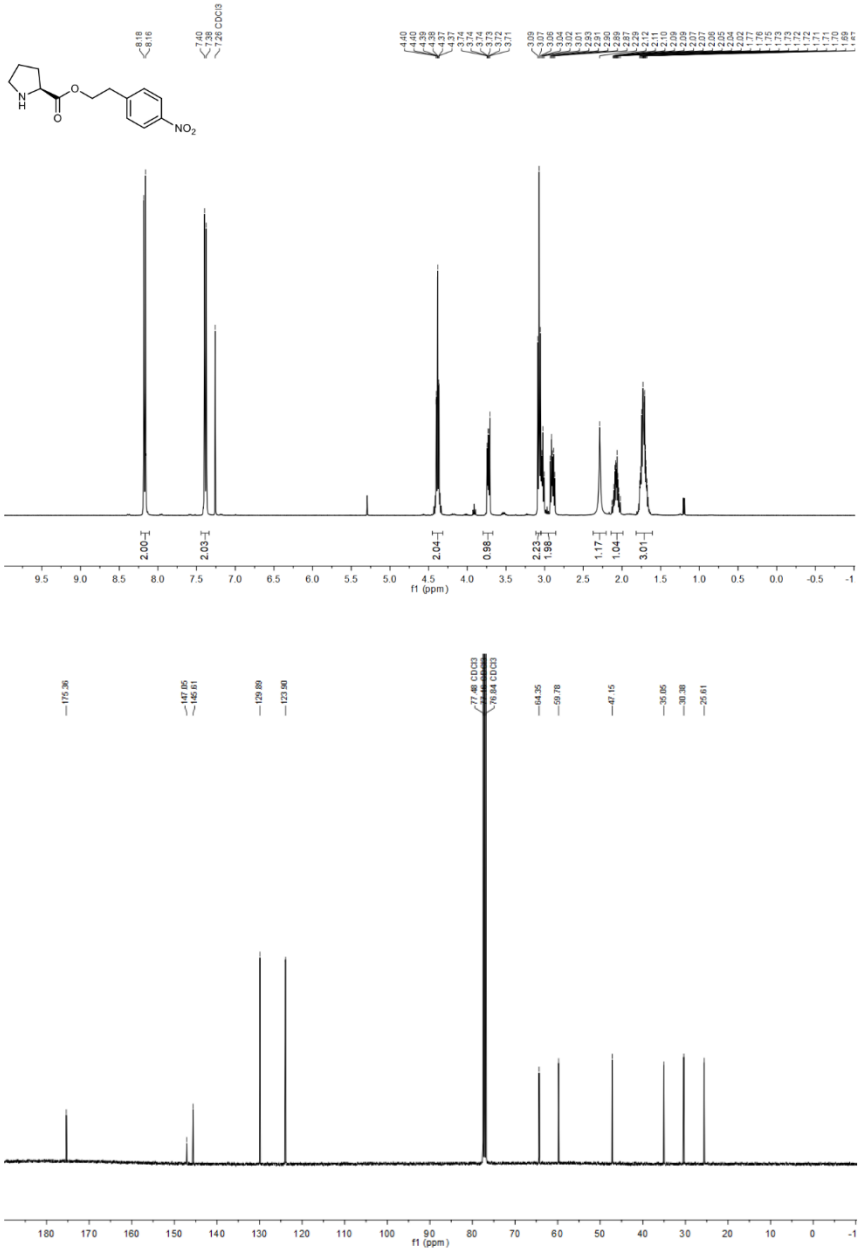
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Ala-Onpe•HCl



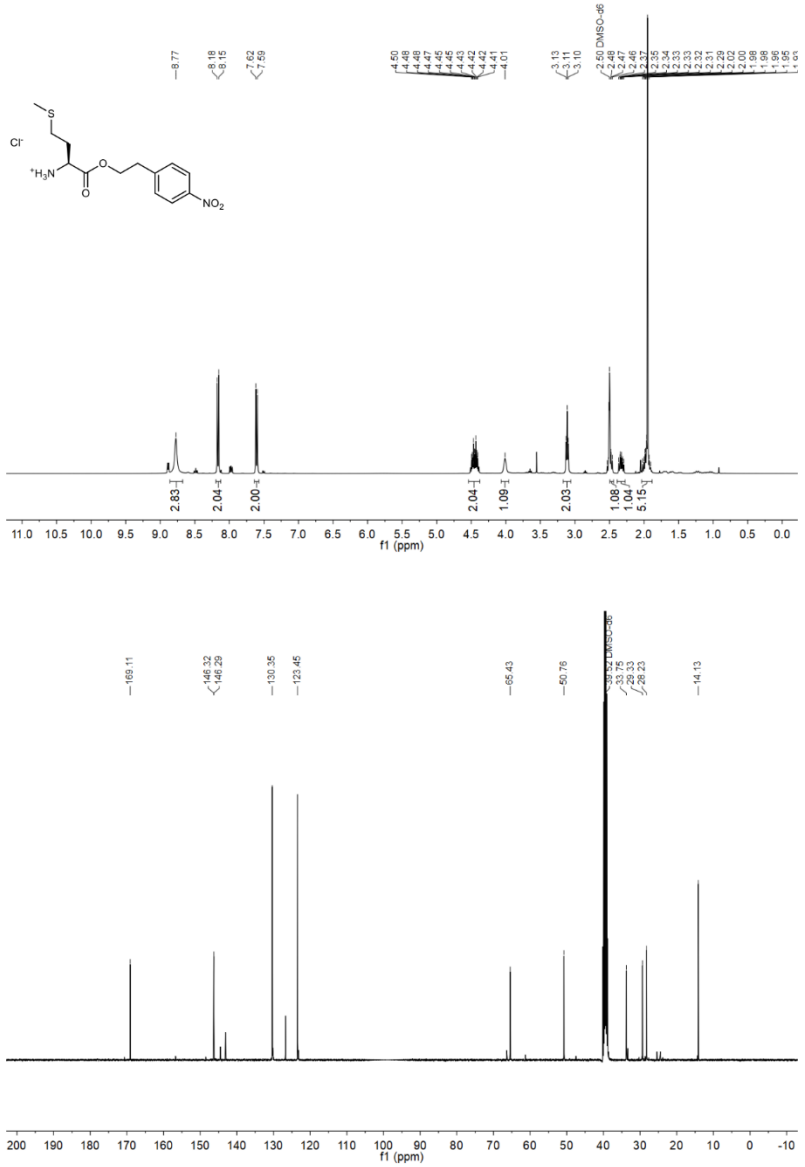
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Leu-Onpe•HCl



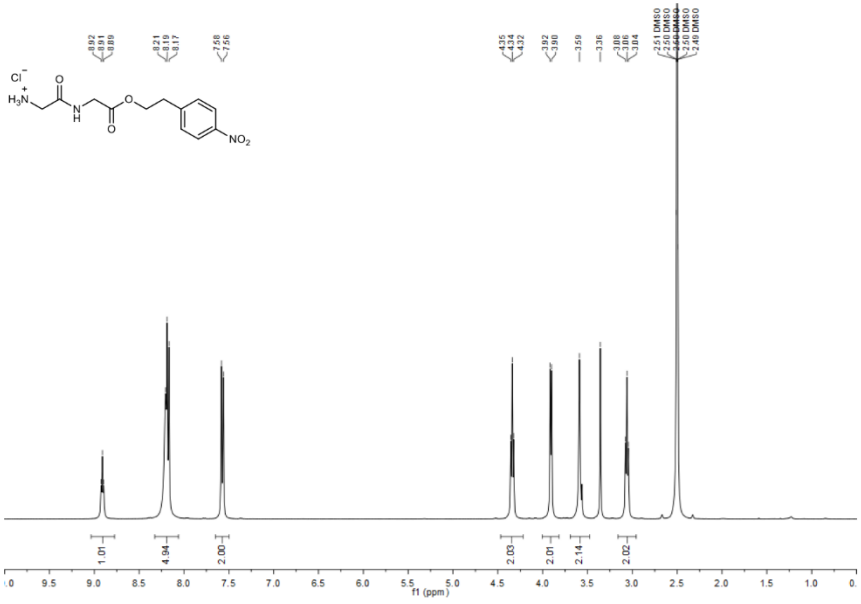
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Pro-Onpe

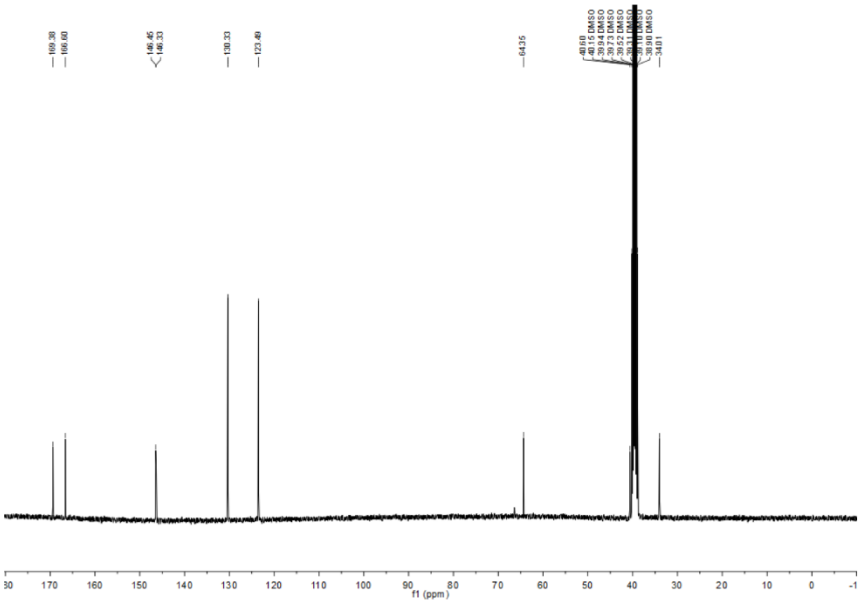


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Met-Onpe•HCl

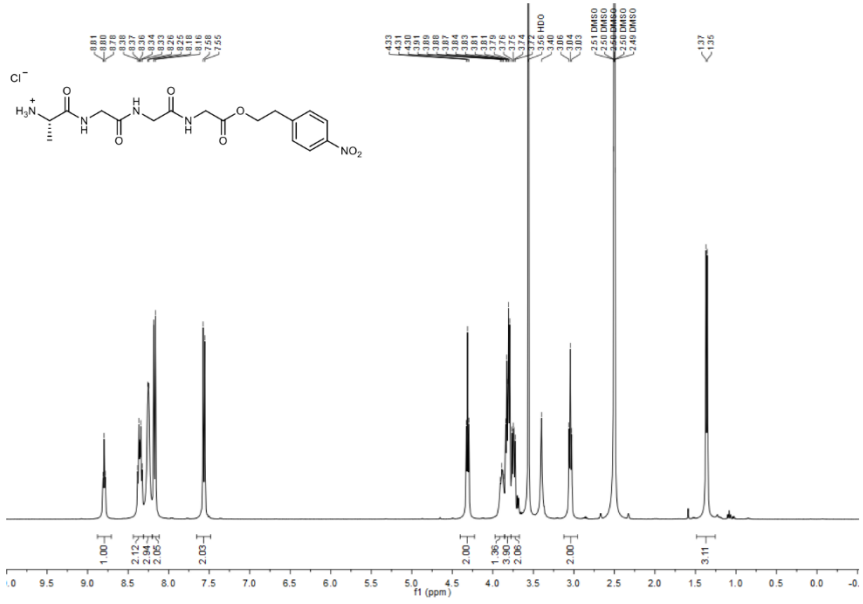


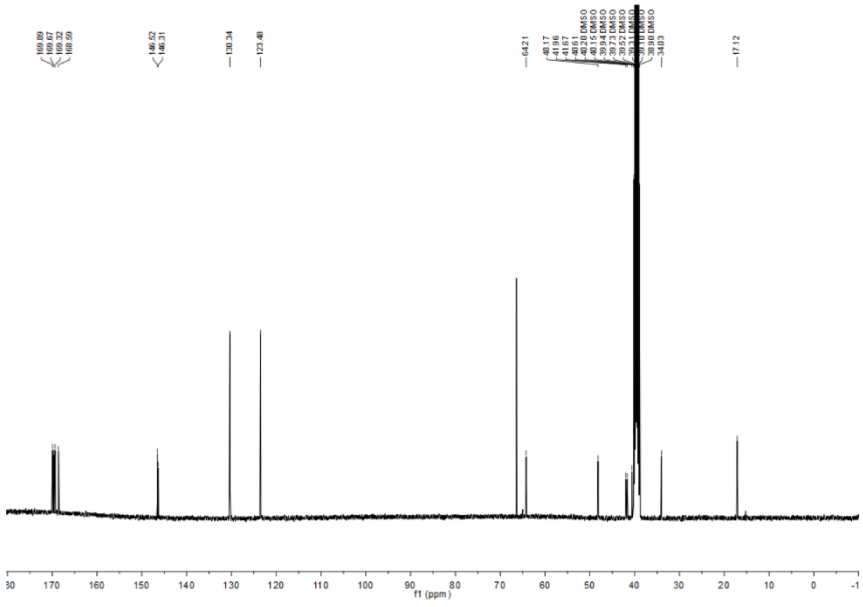
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Gly-Gly-Onpe•HCl



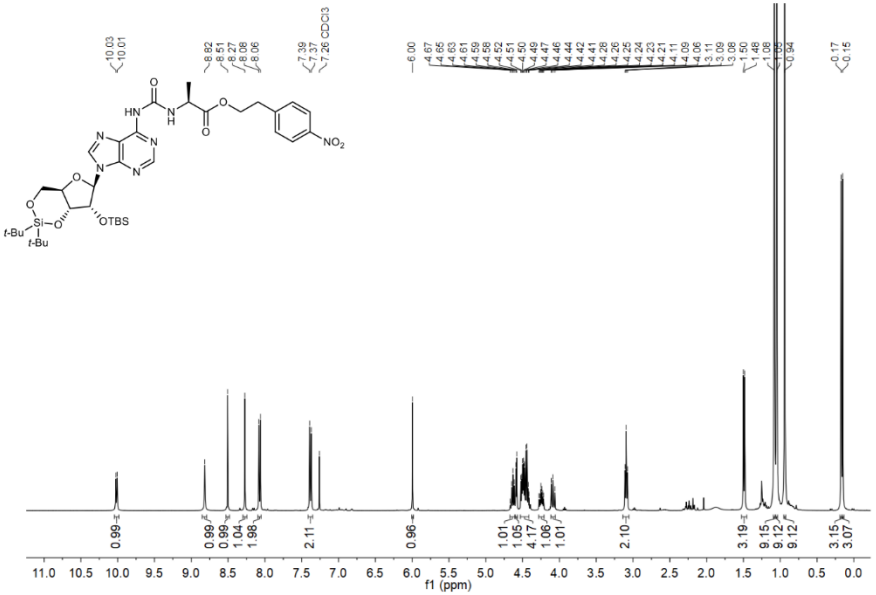


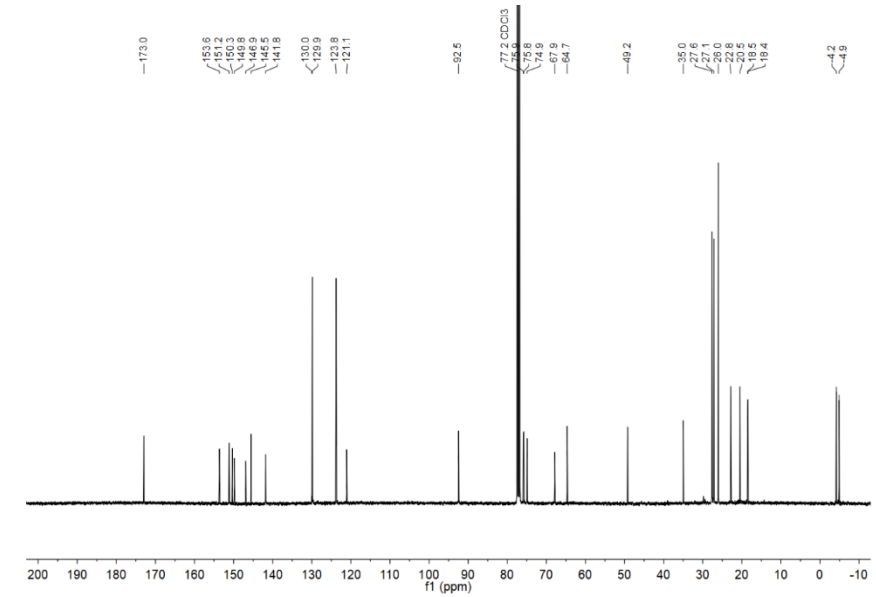
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound H-Ala-Gly-Gly-Gly-Onpe•HCl



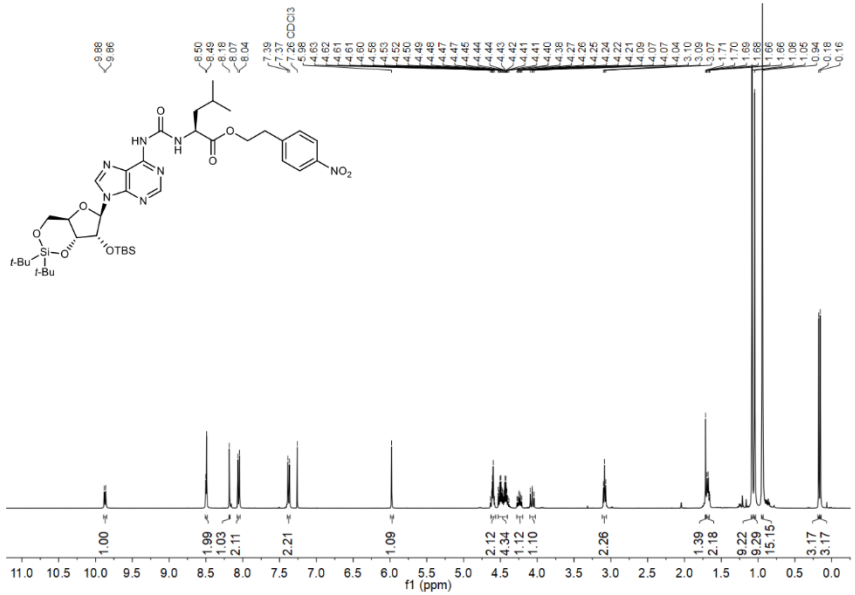


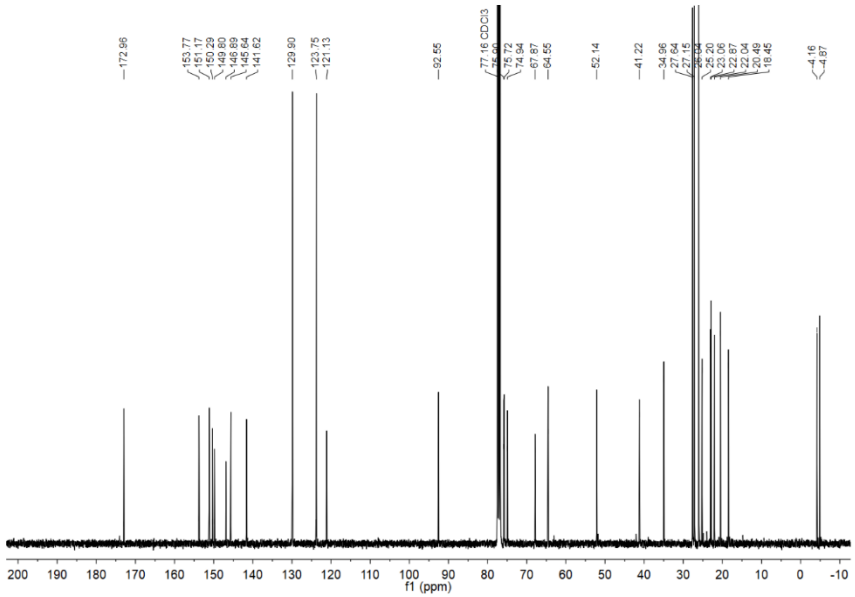
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10b



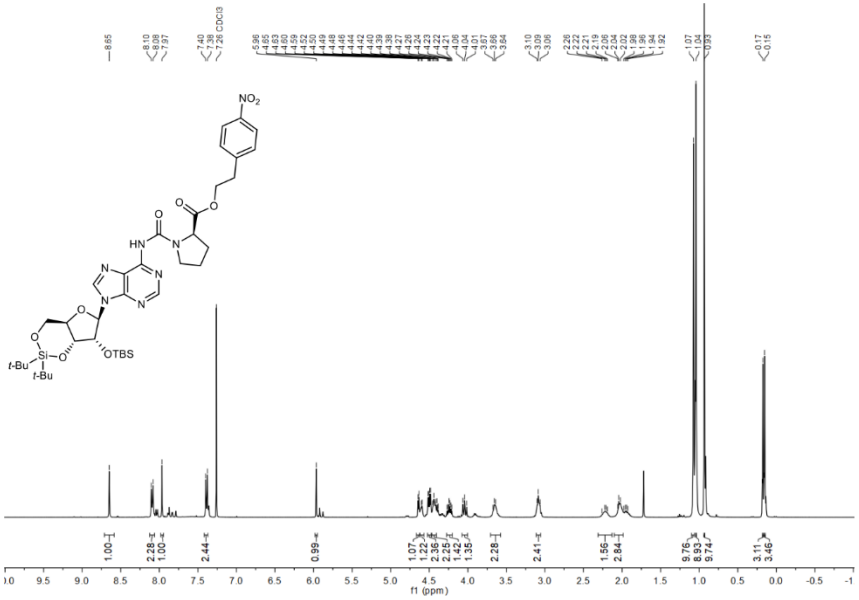


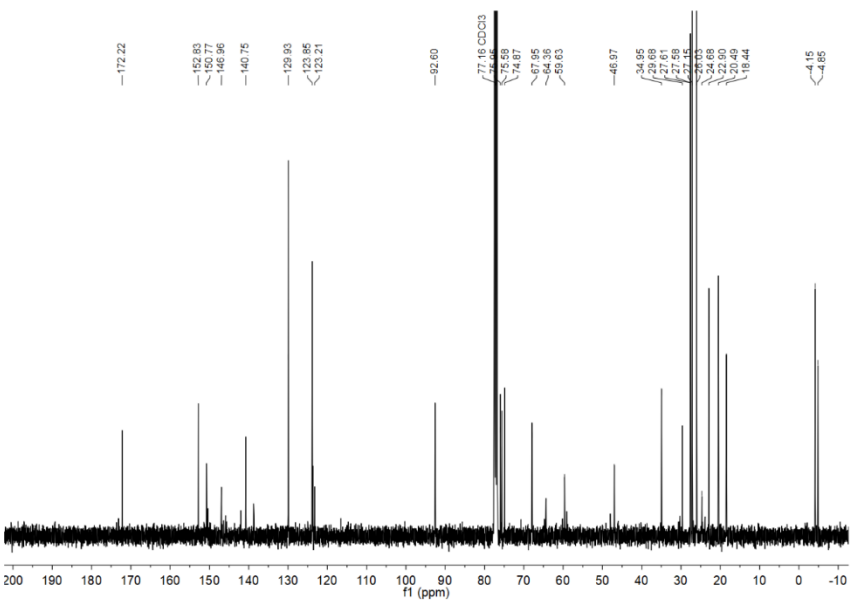
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10d



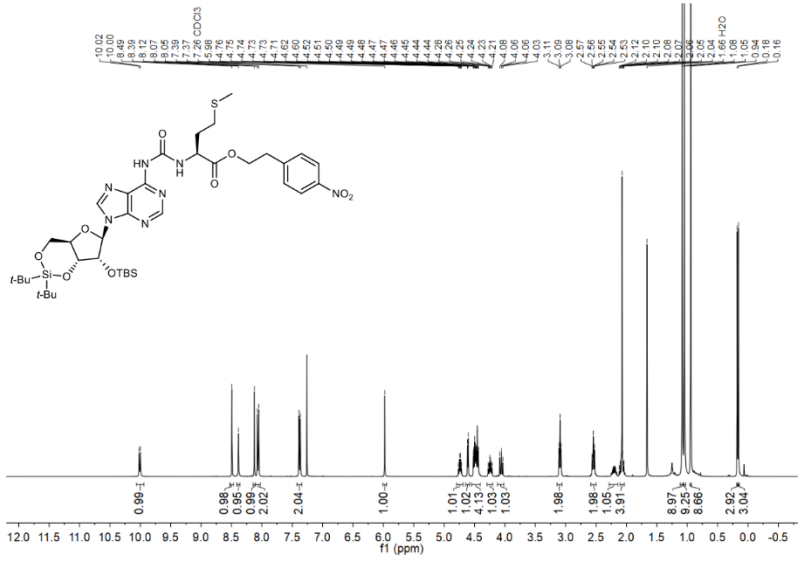


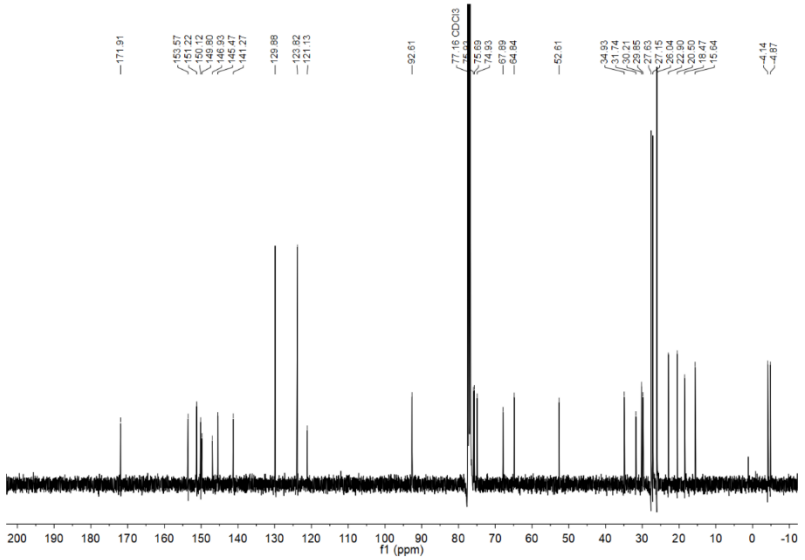
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10f



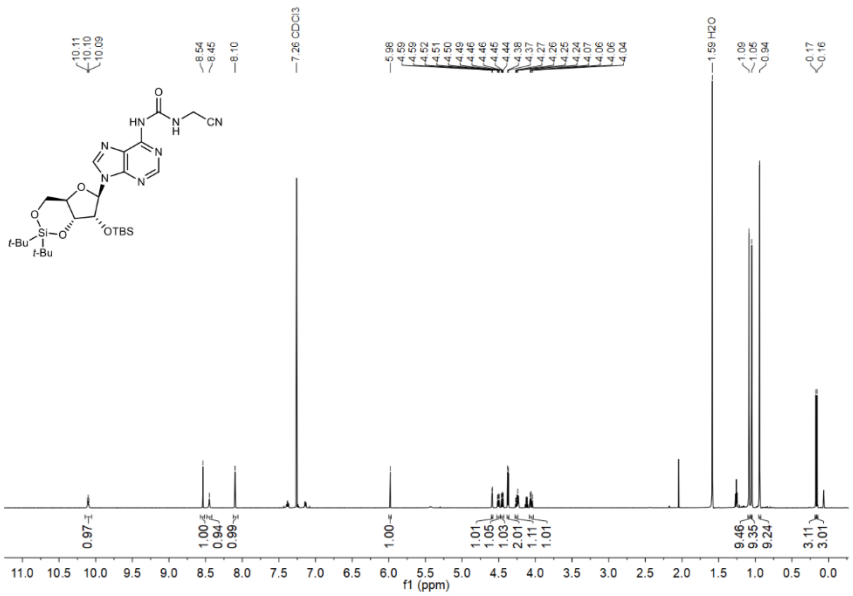


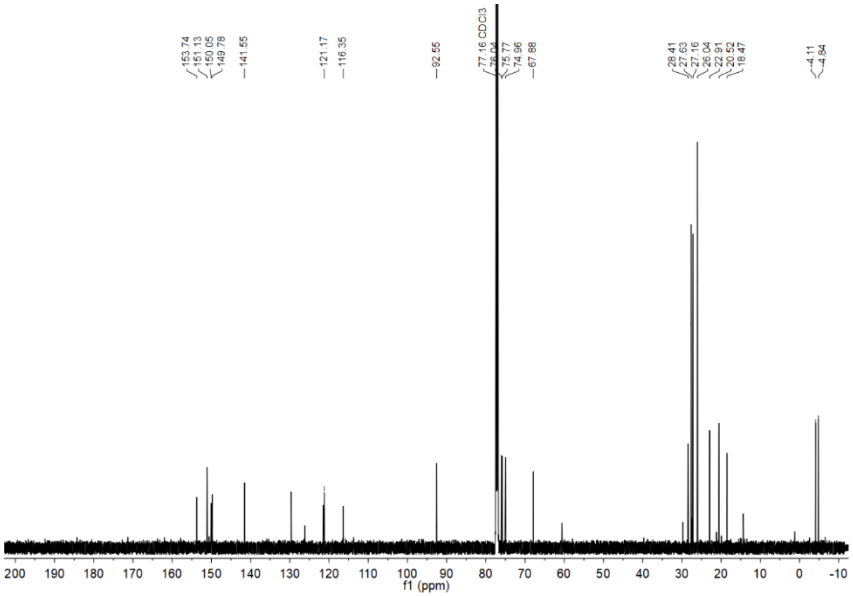
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10h



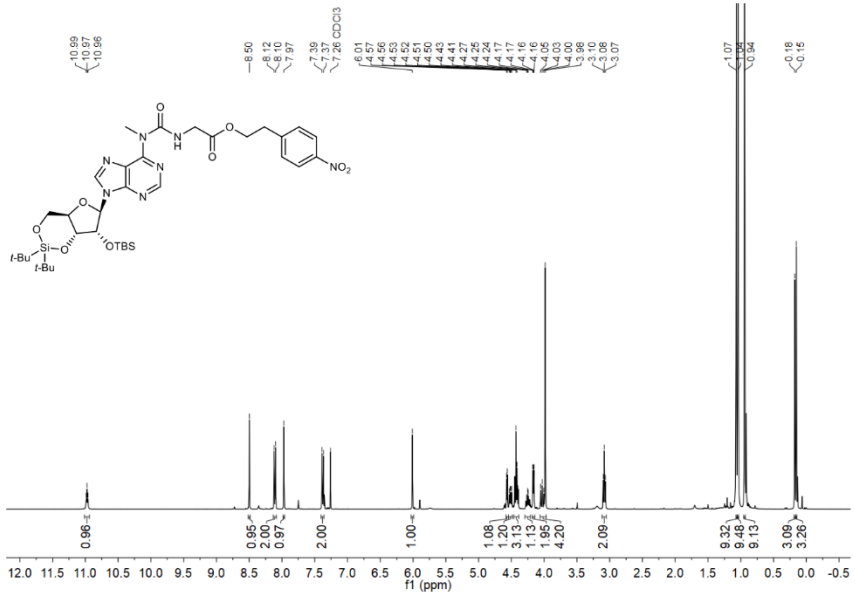


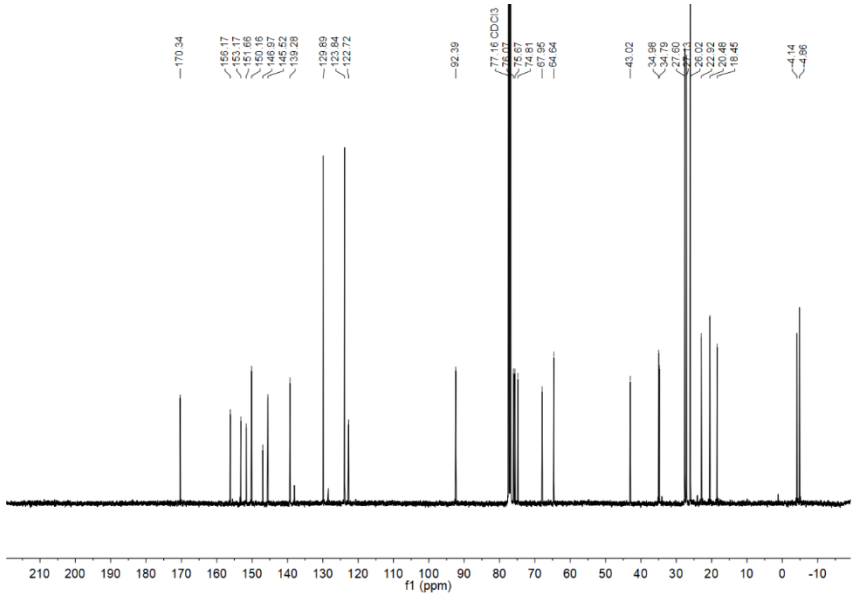
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 10j



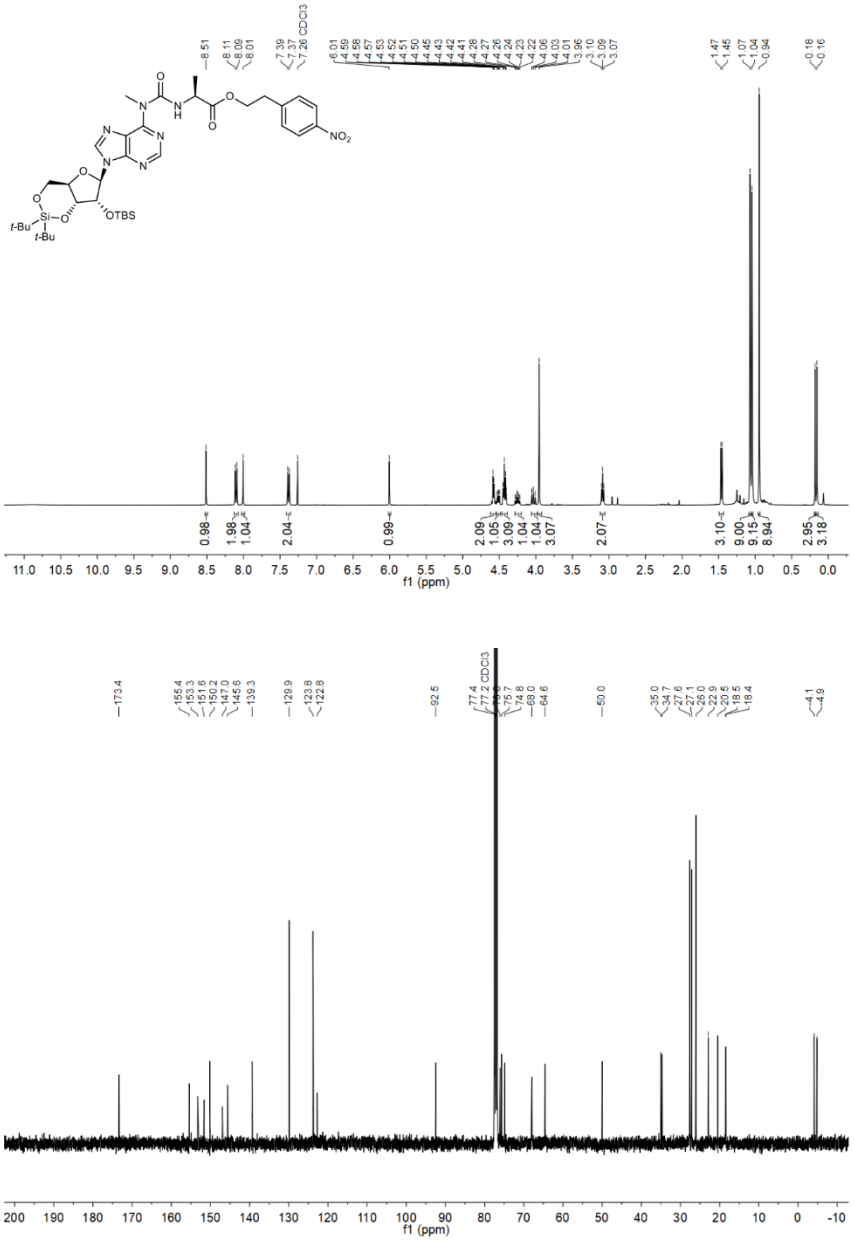


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11a

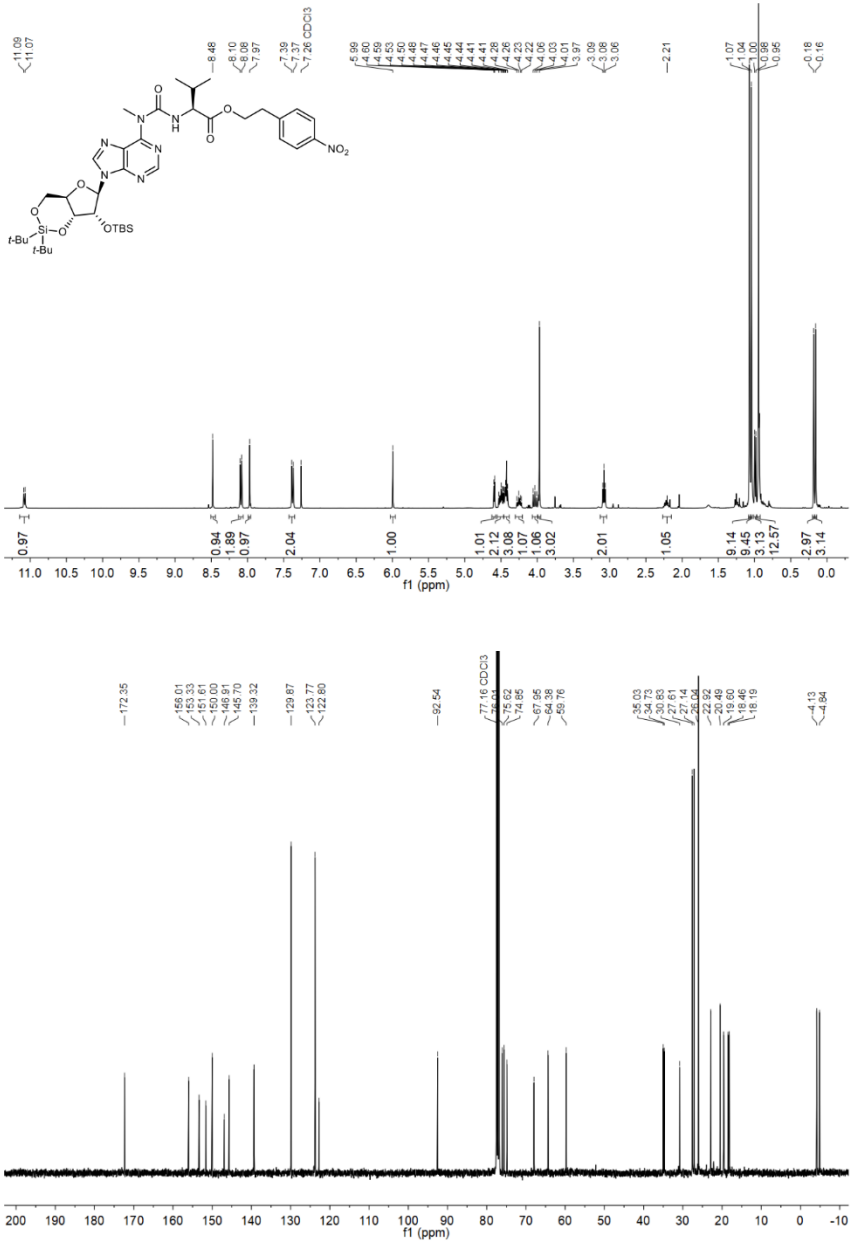




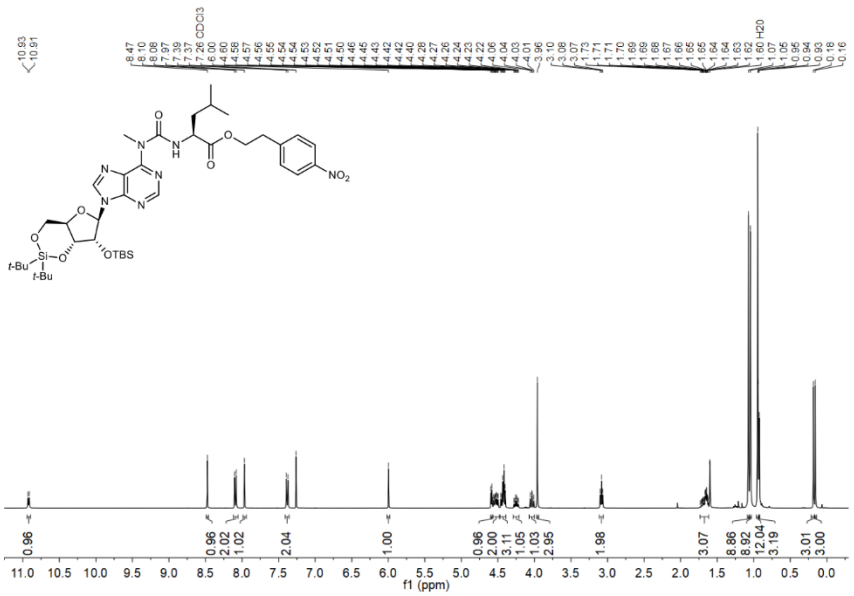
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11b

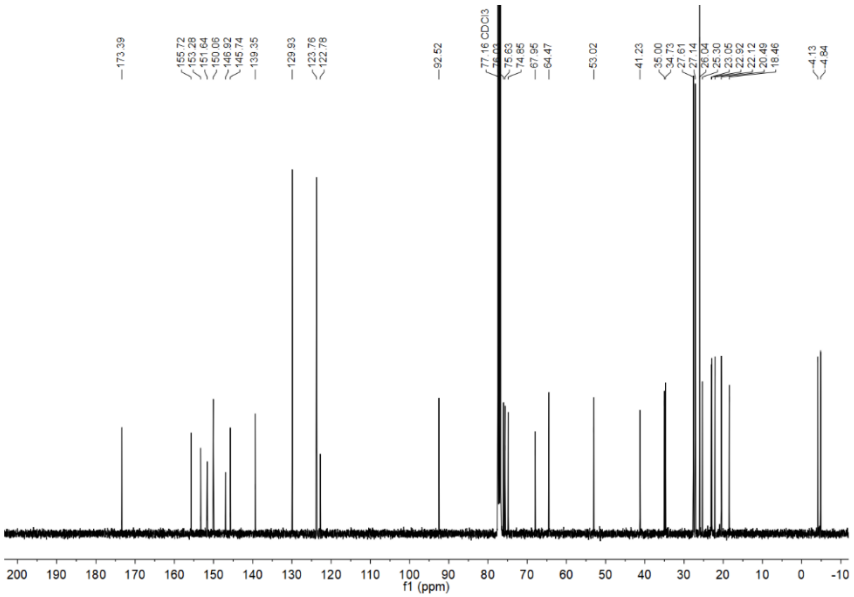


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11c

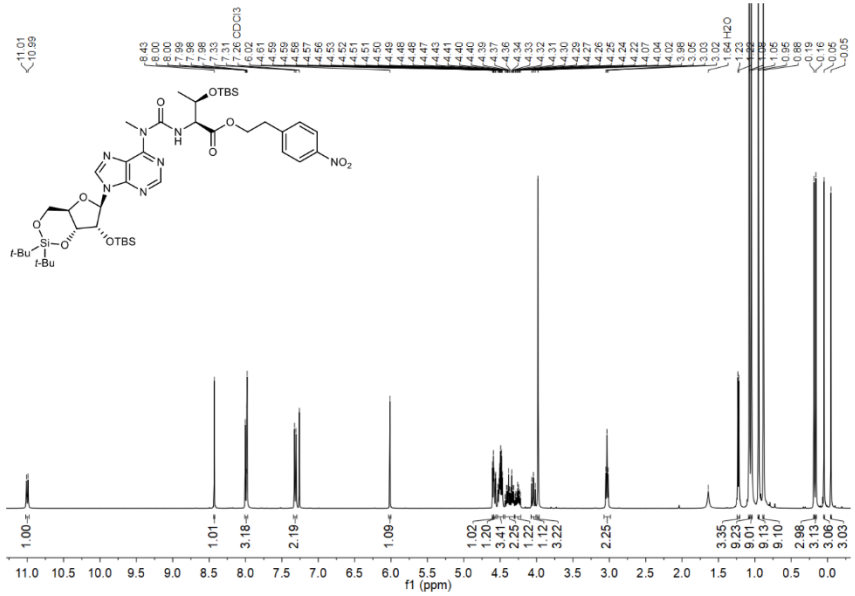


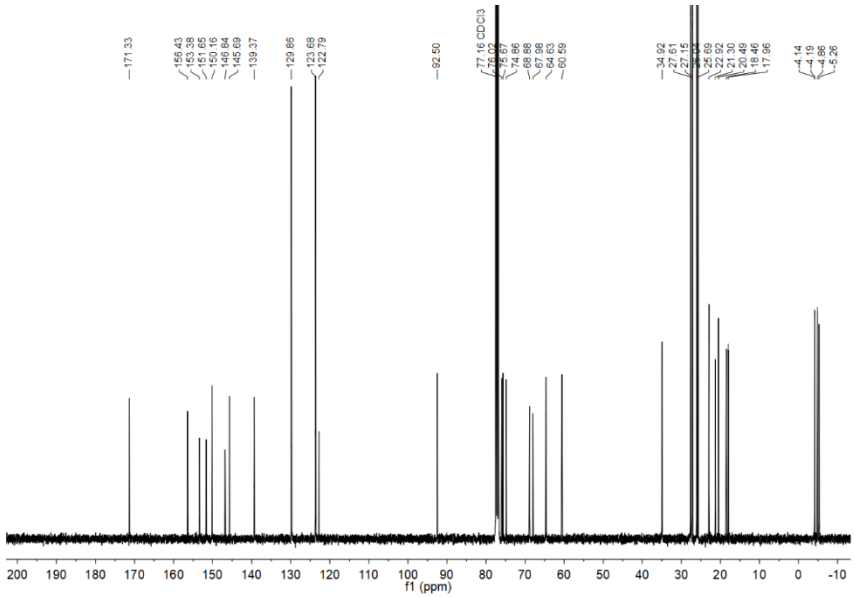
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11d



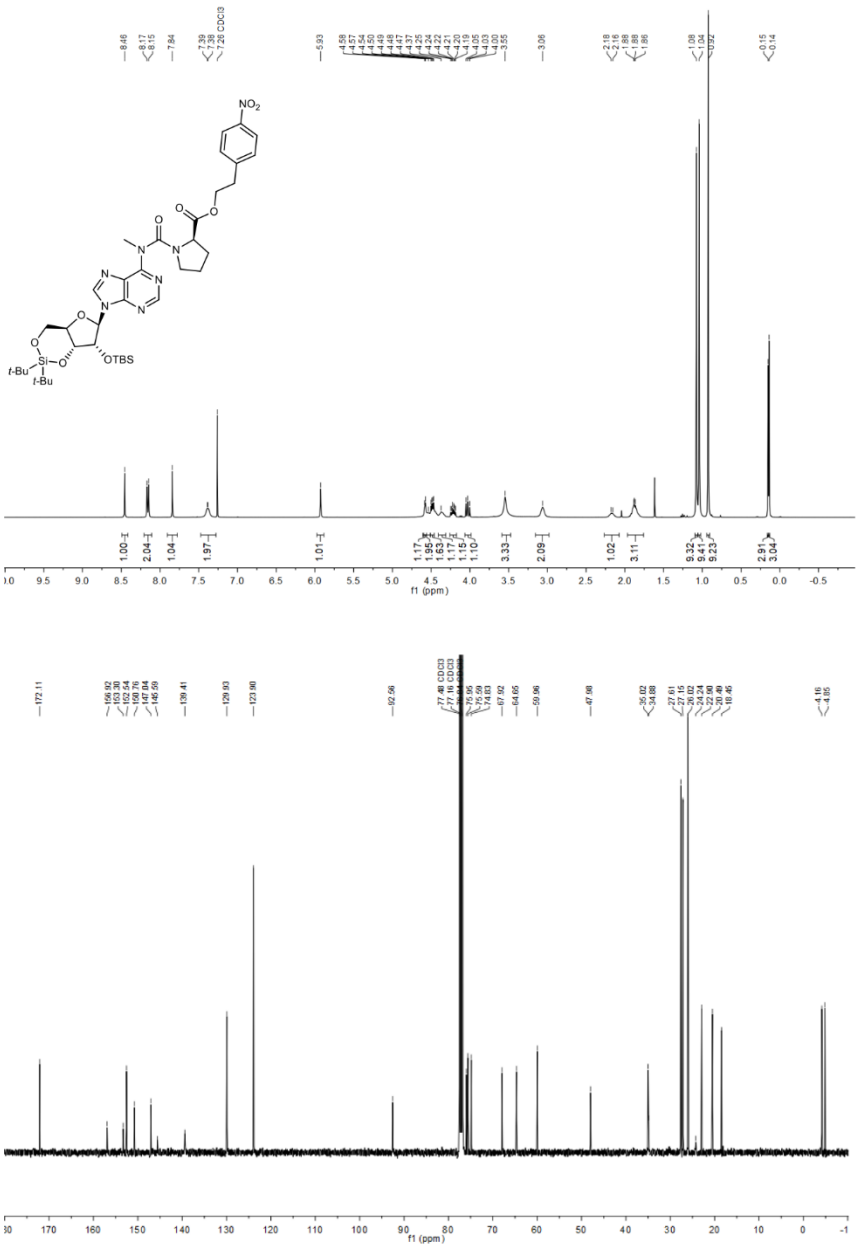


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11e

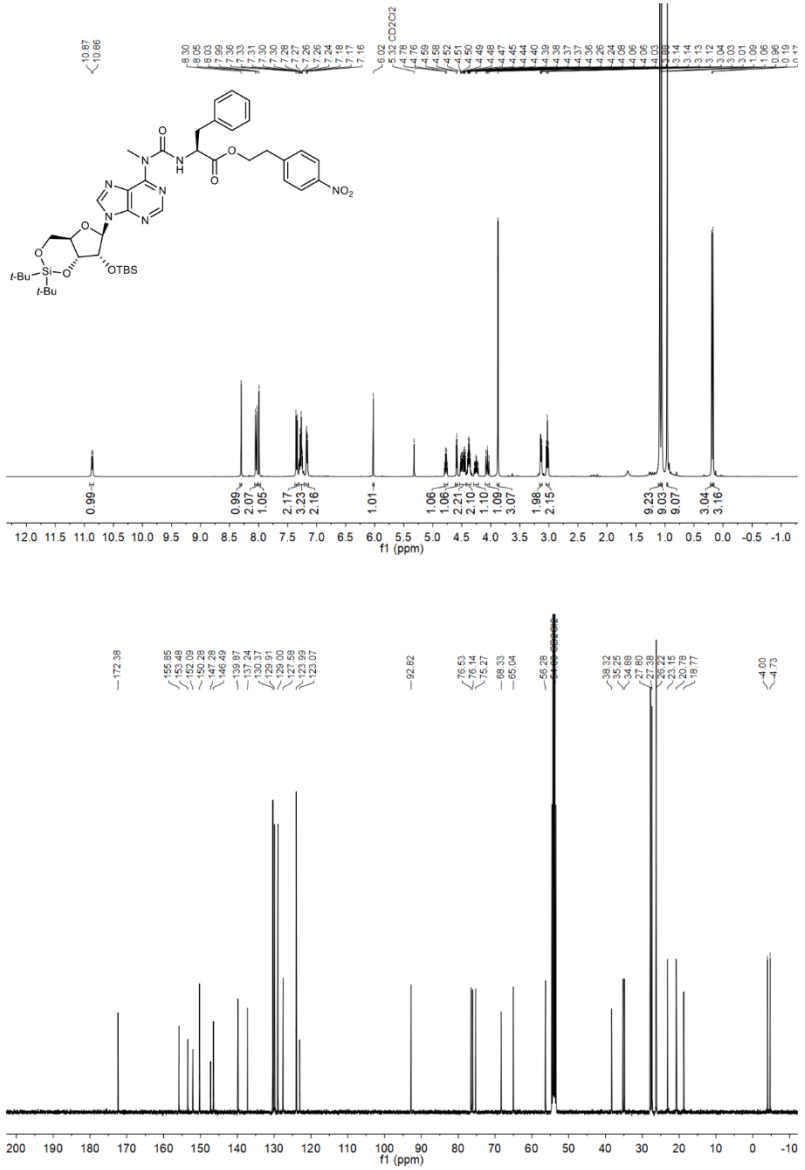




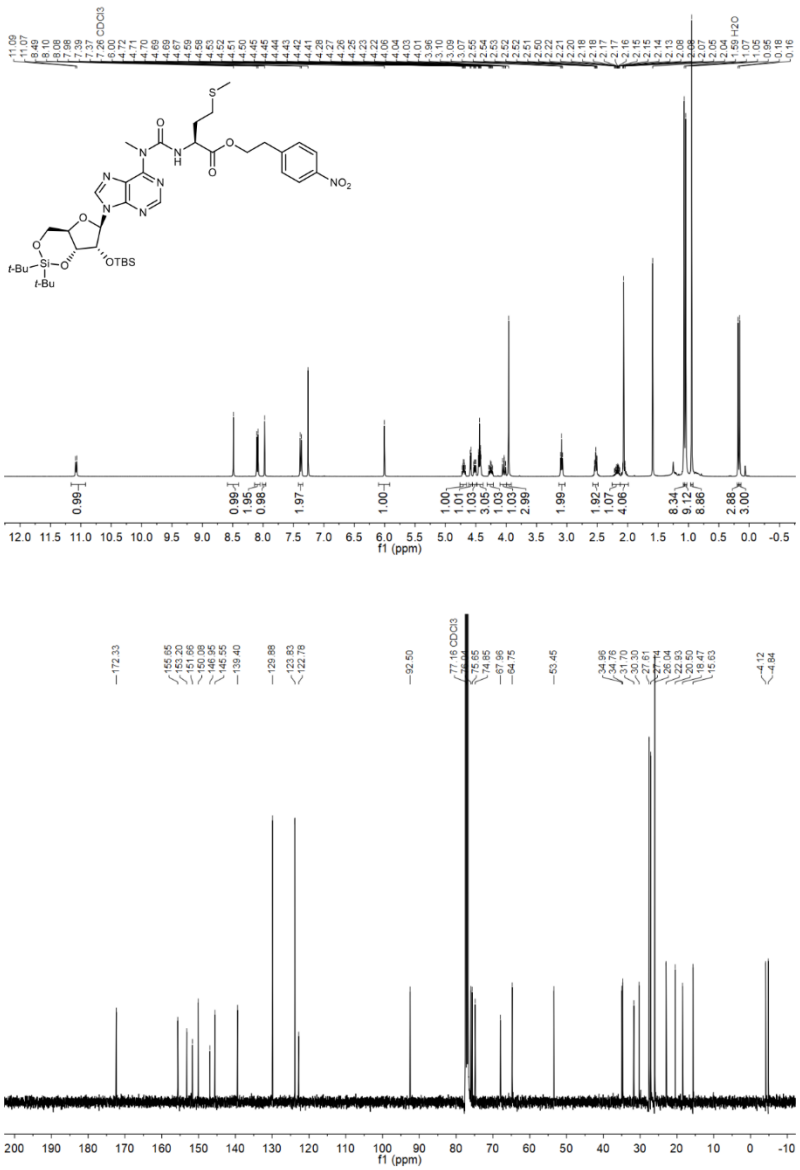
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11f



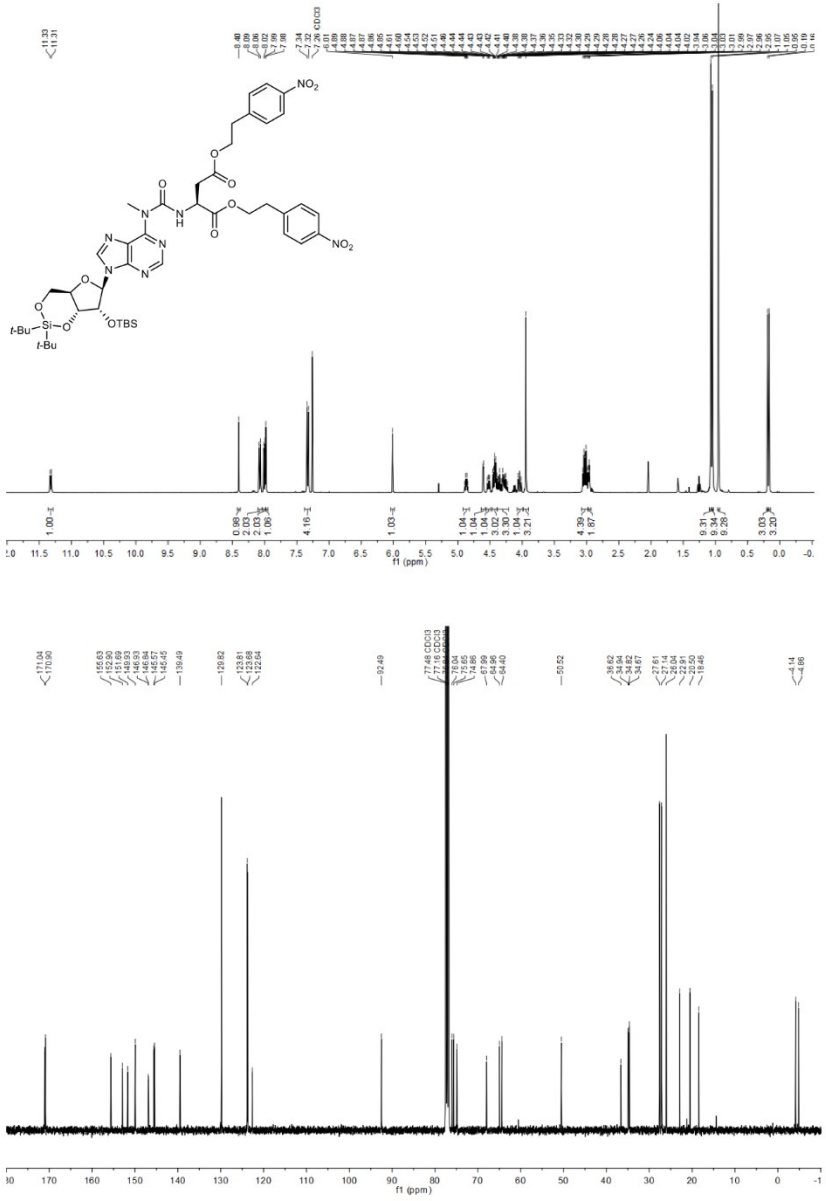
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11g



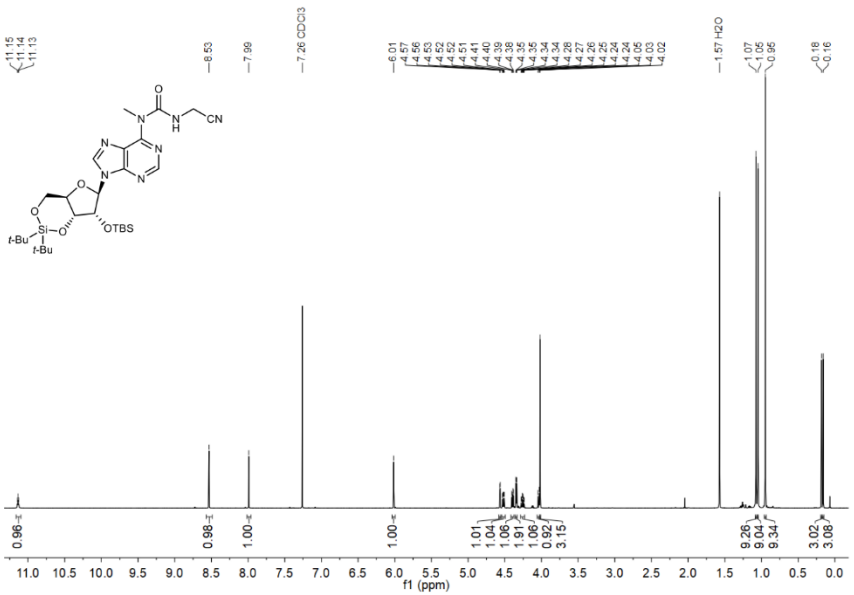
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11h

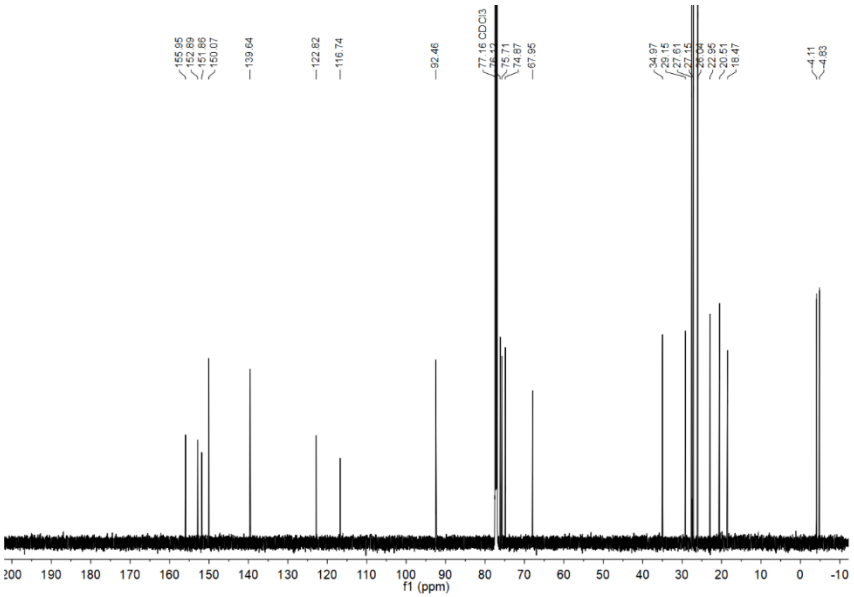


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11i

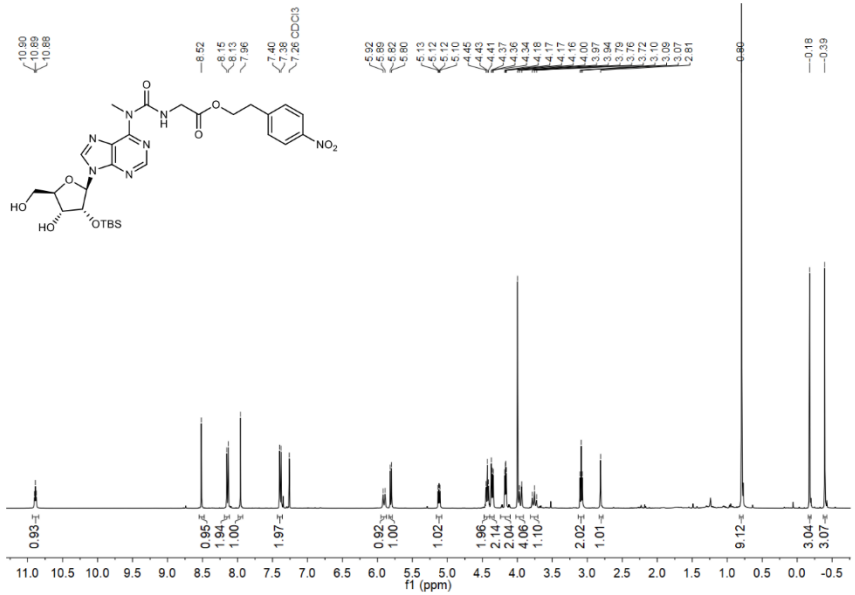


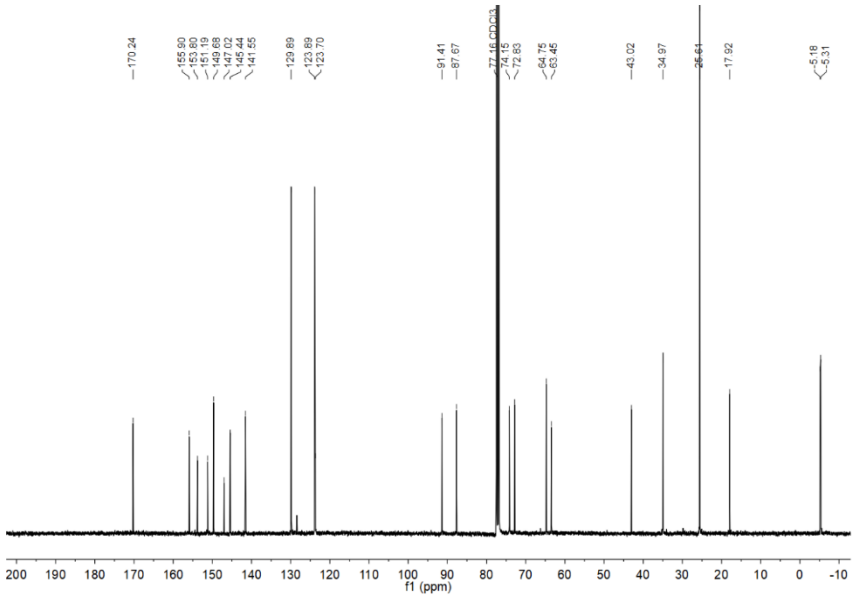
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 11j



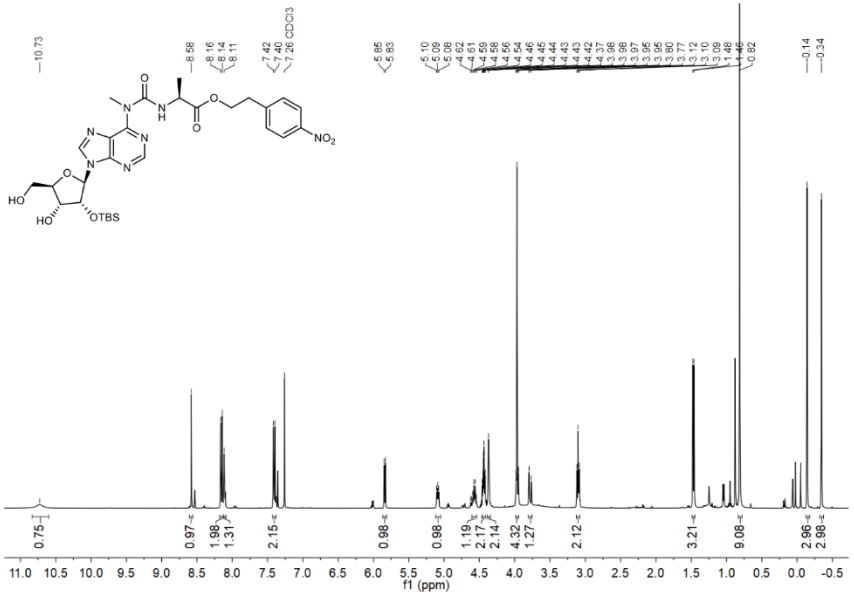


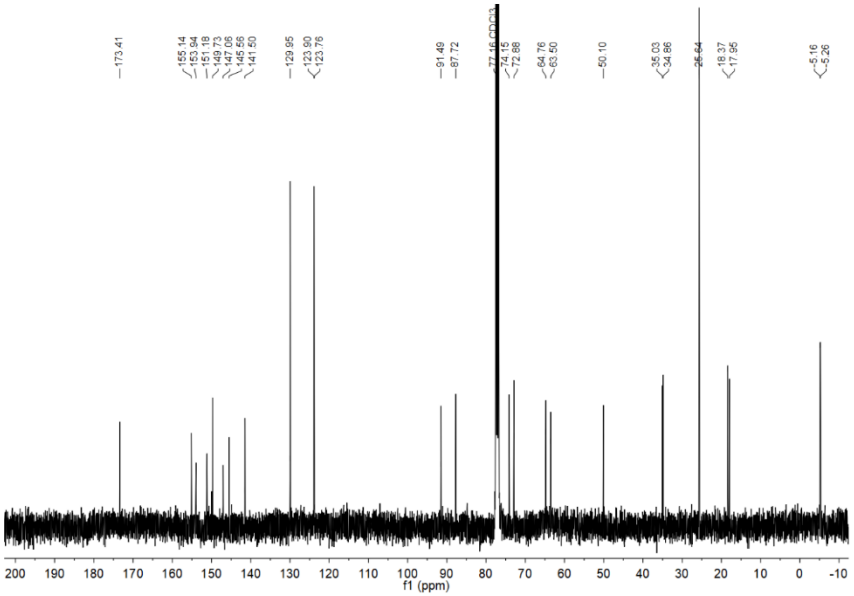
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12a



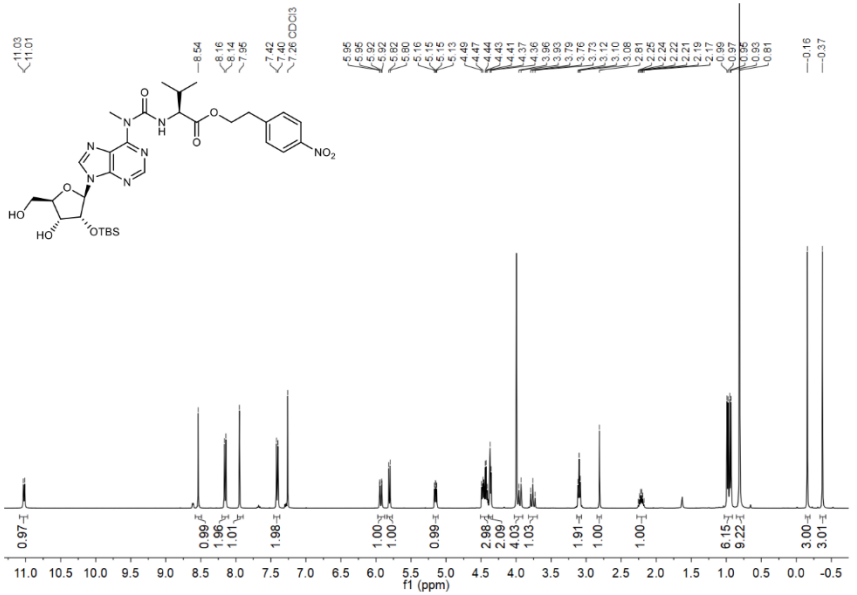


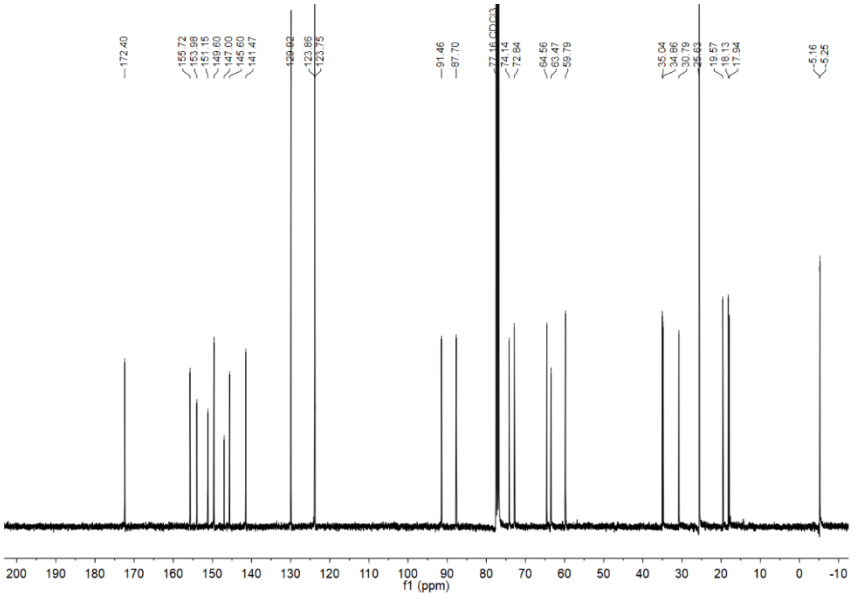
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12b



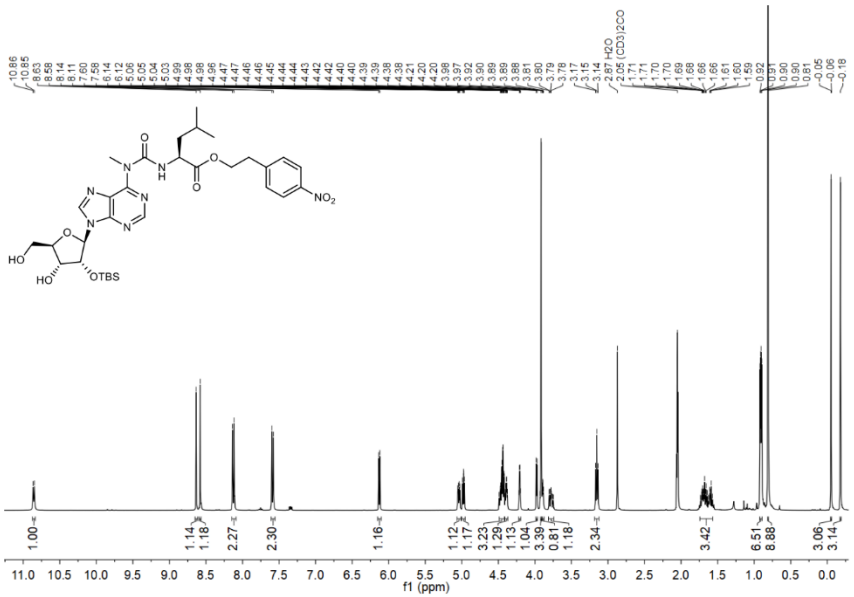


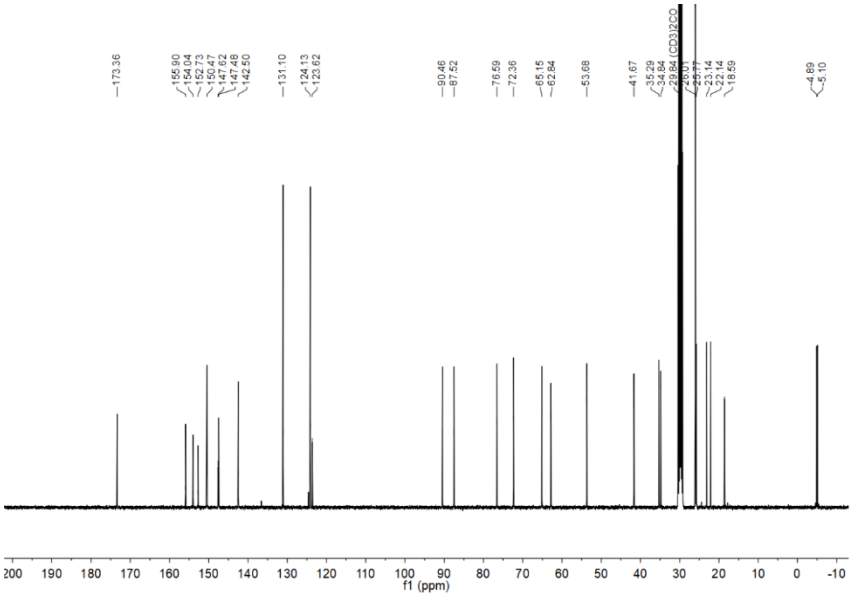
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12c



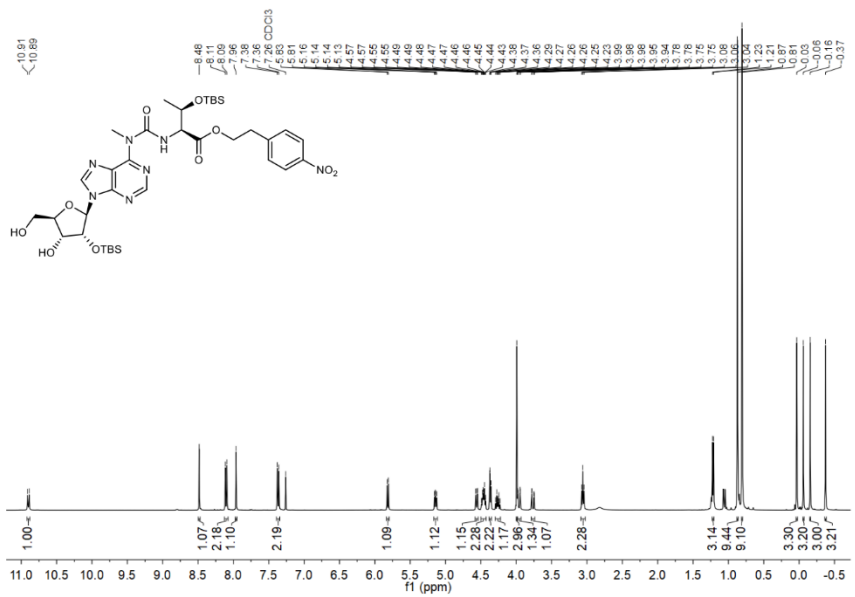


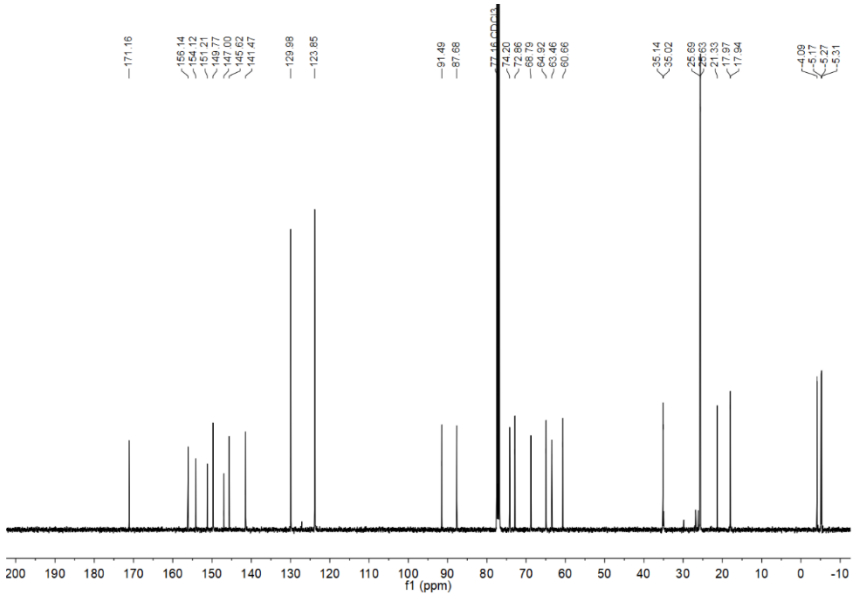
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12d



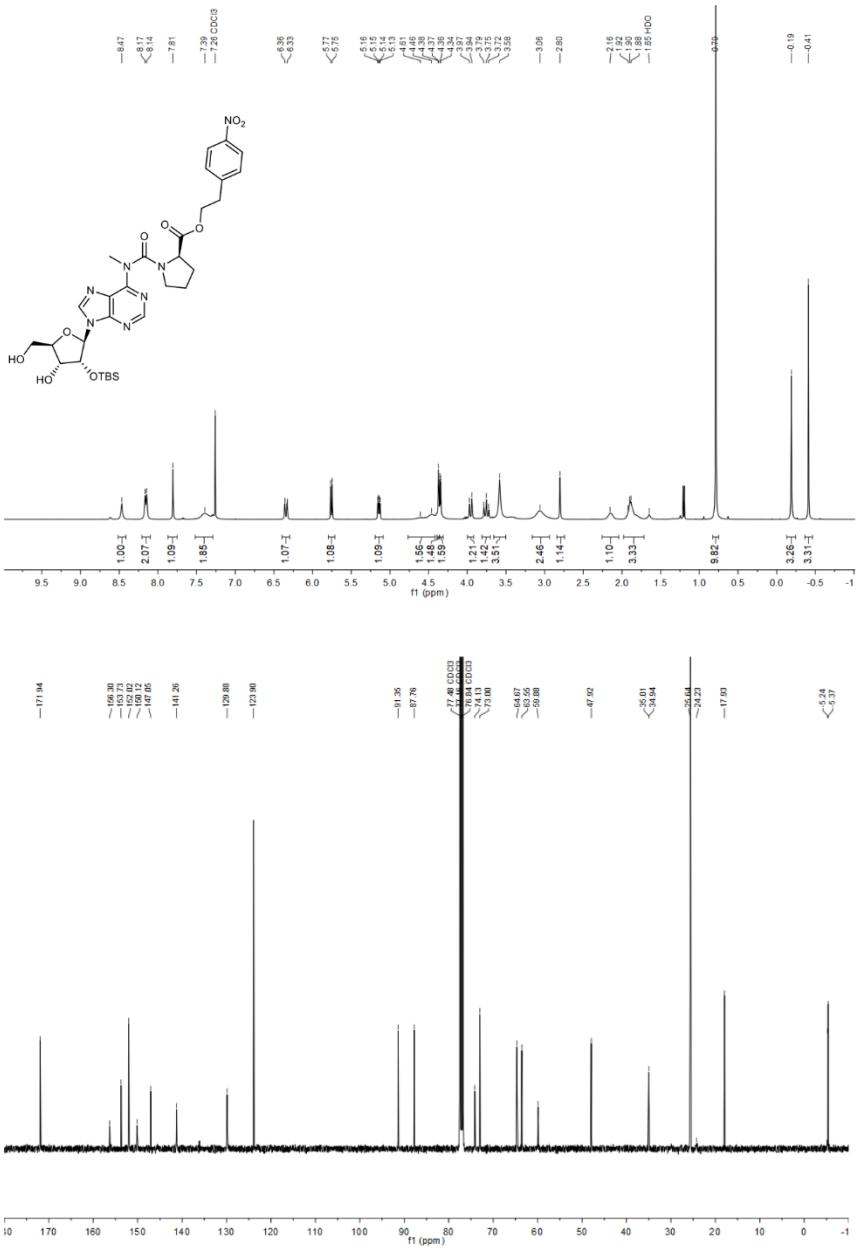


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12e

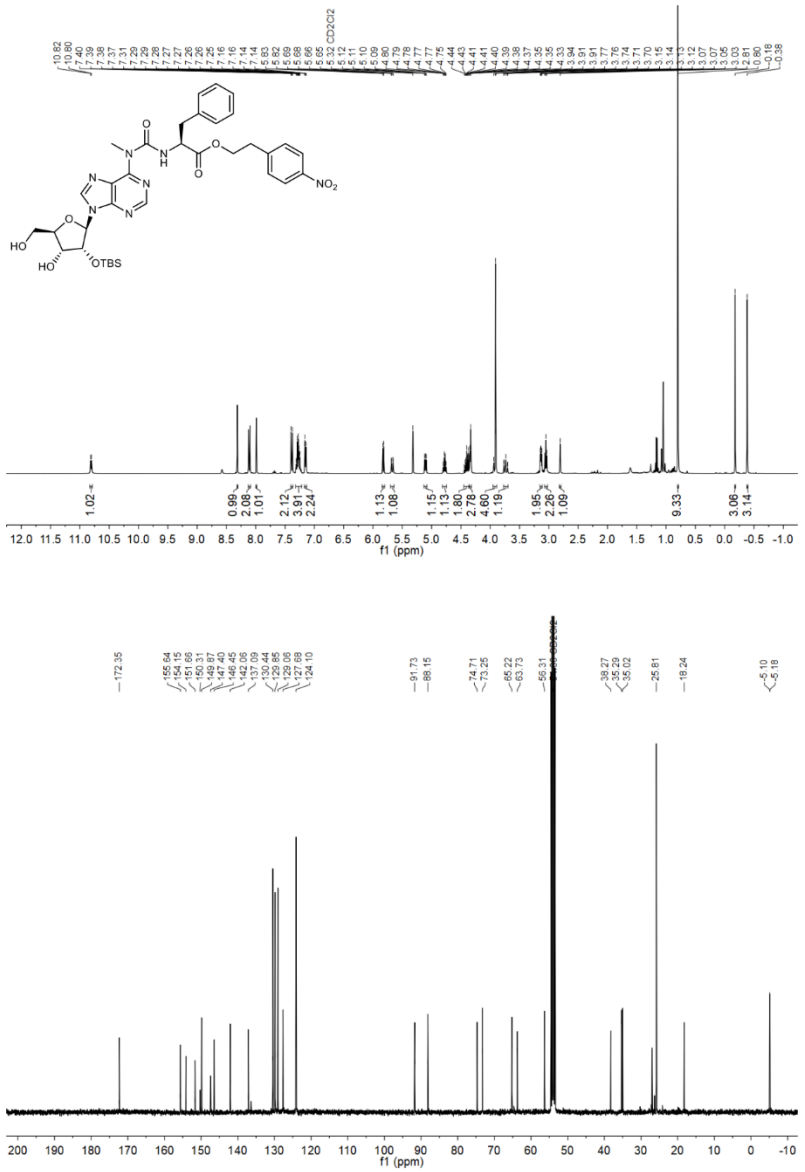




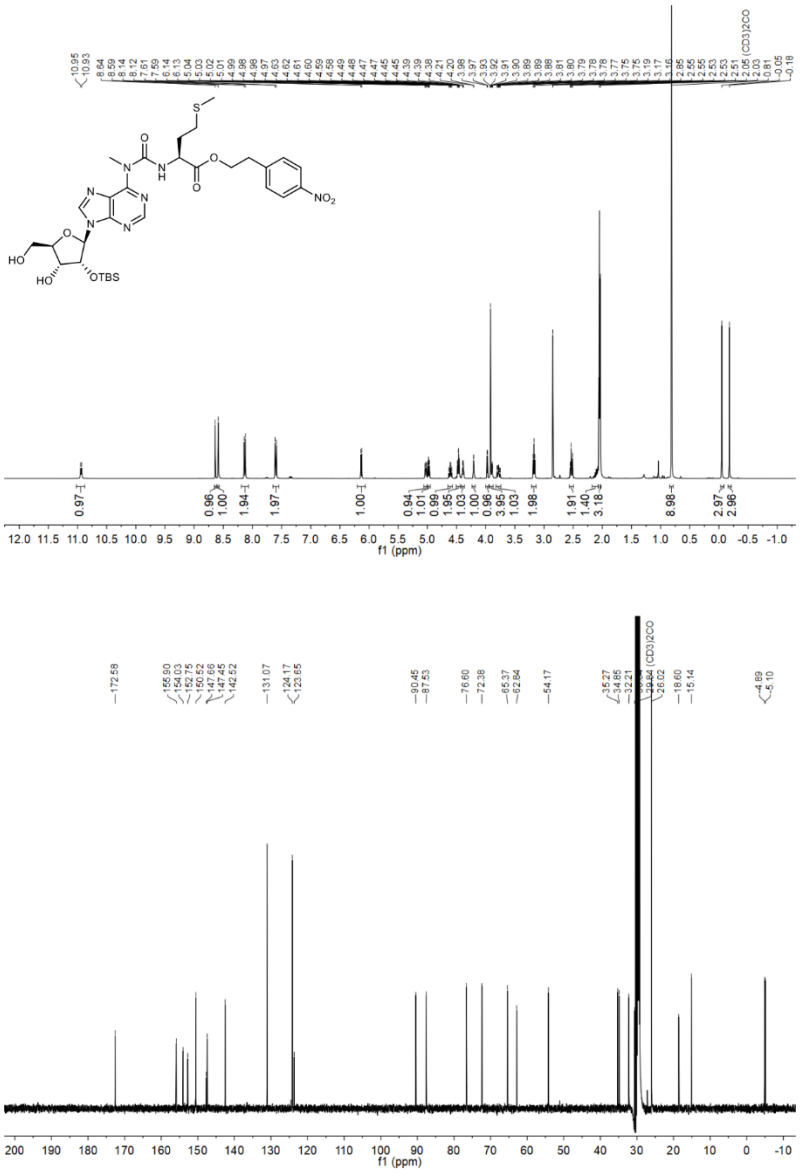
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12f



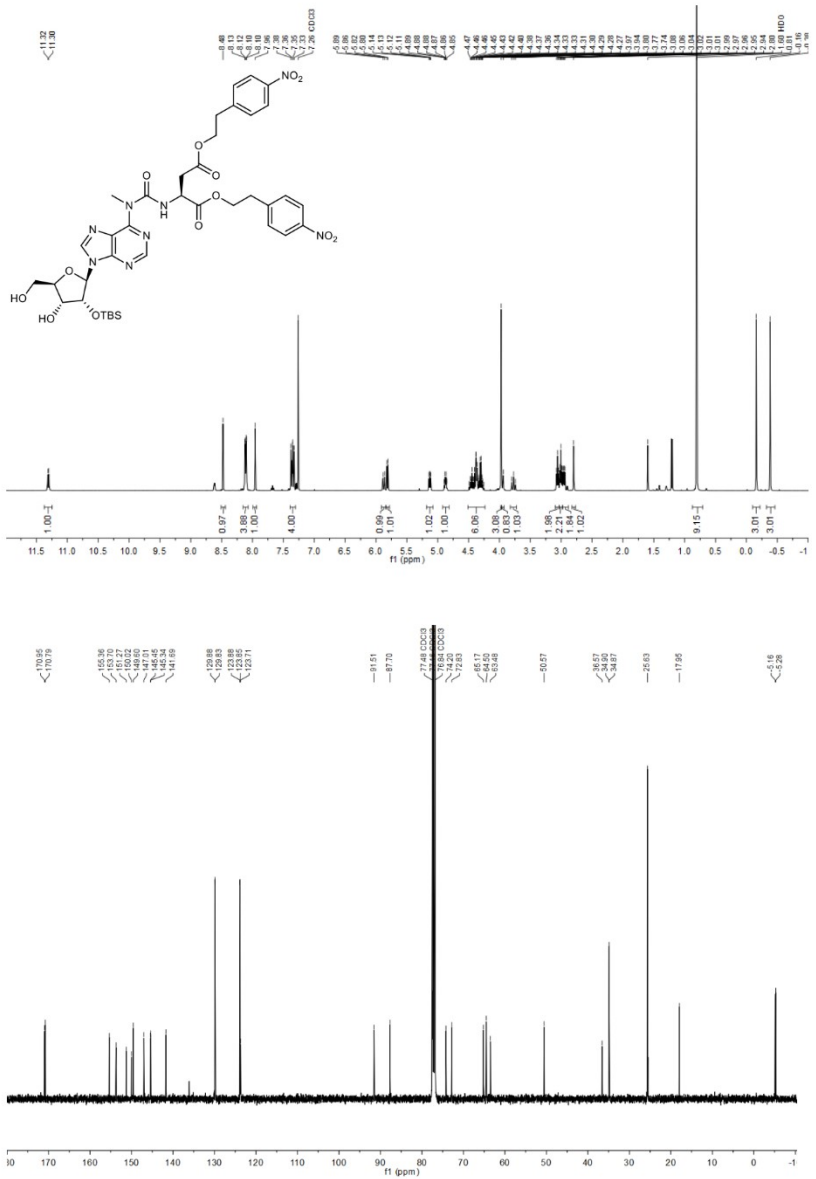
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12g



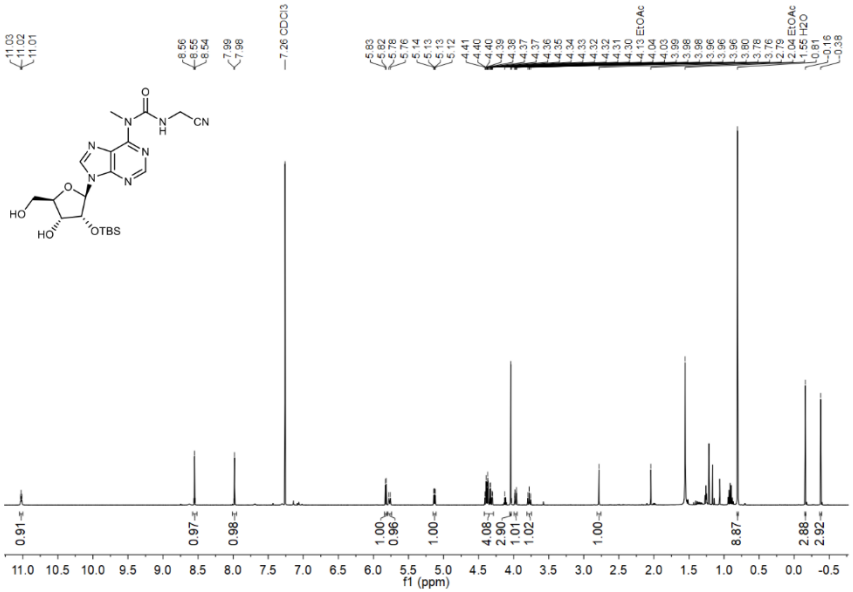
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12h

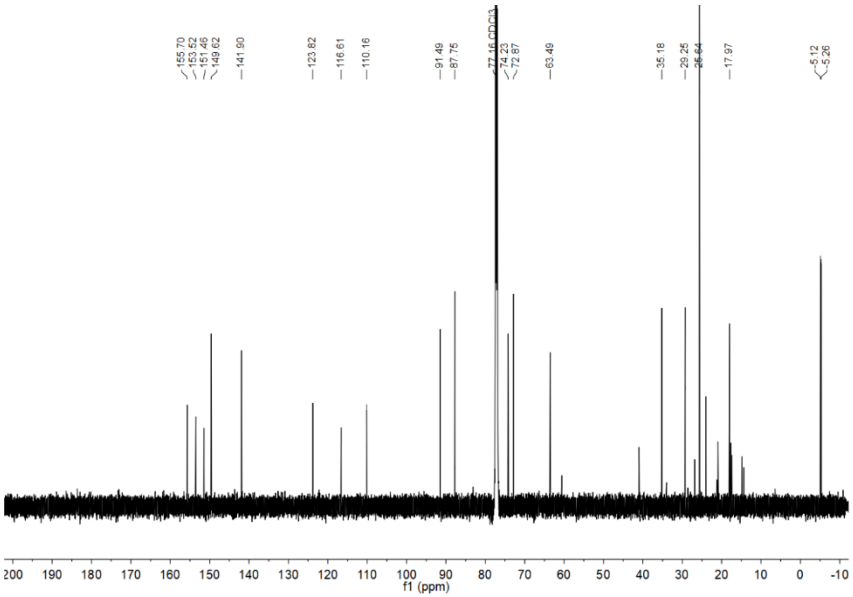


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12i

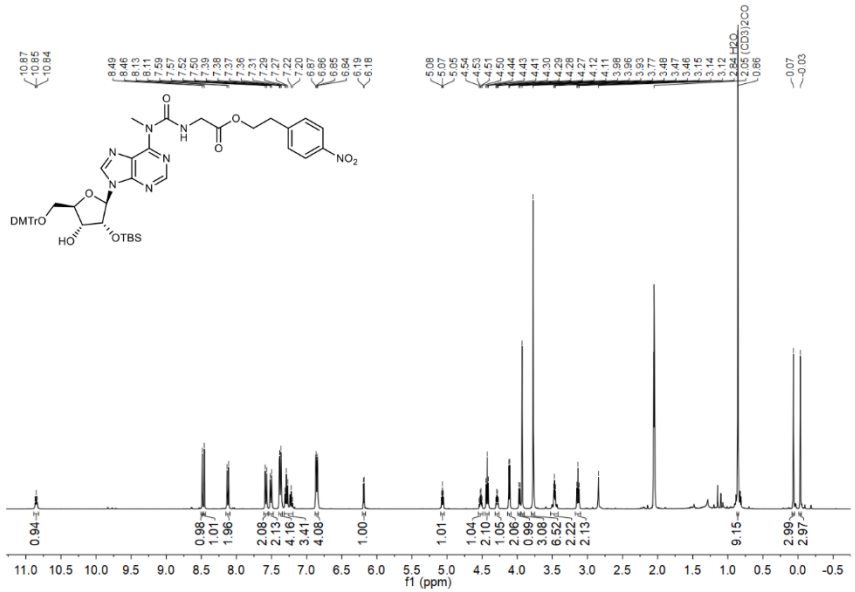


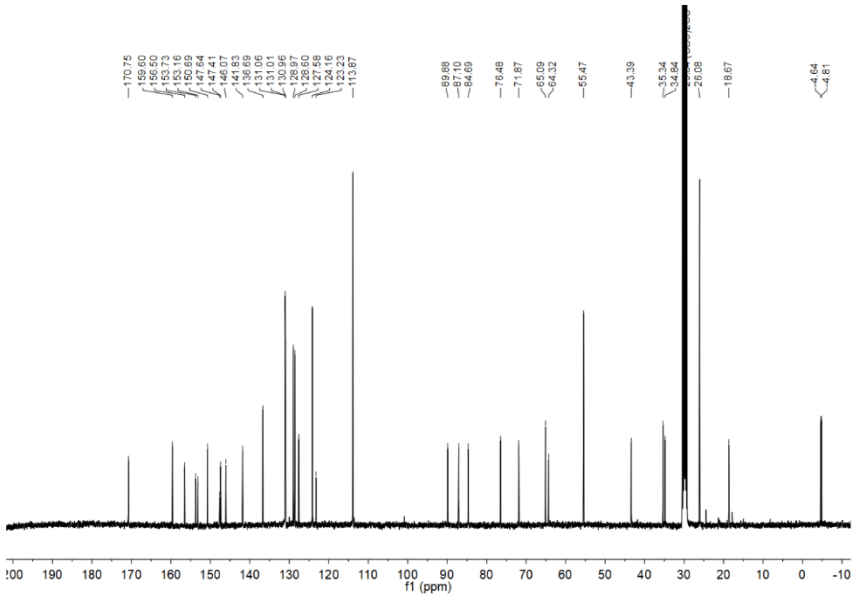
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 12j



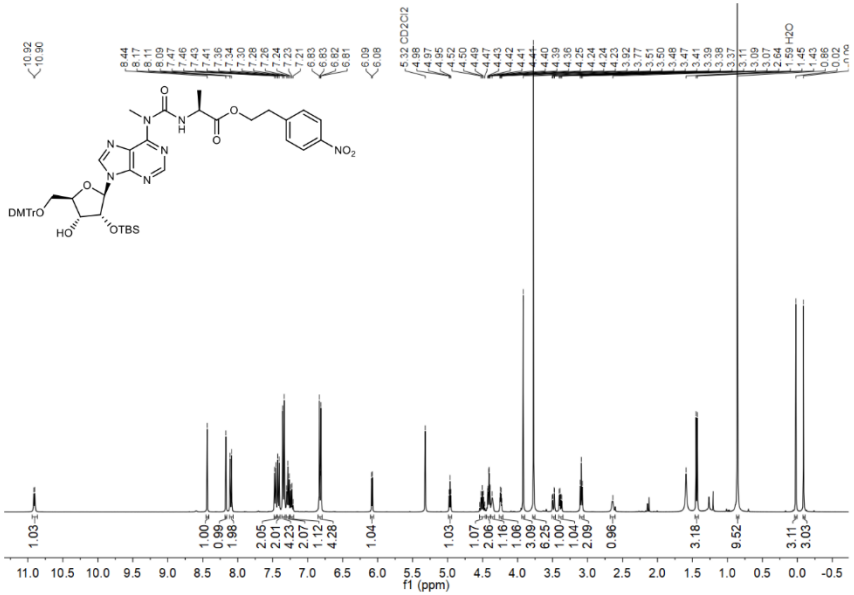


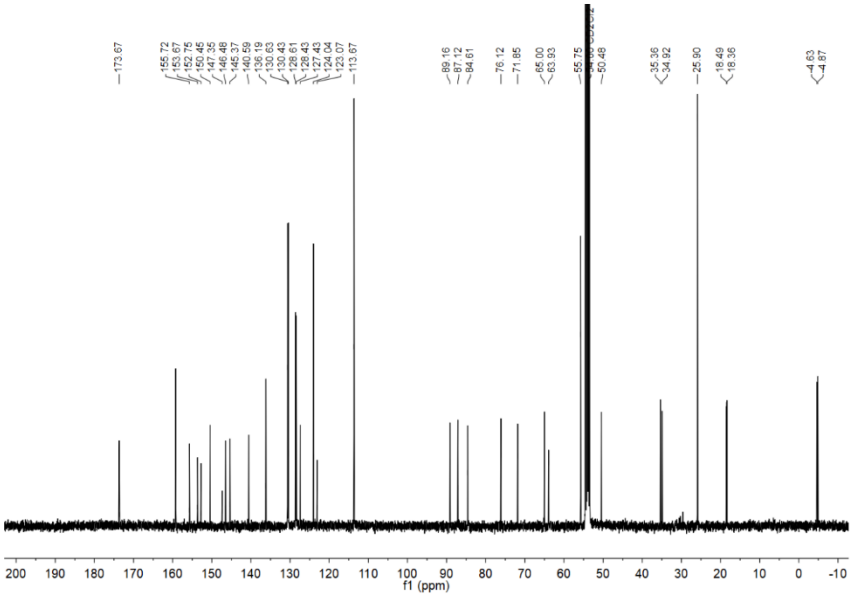
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13a



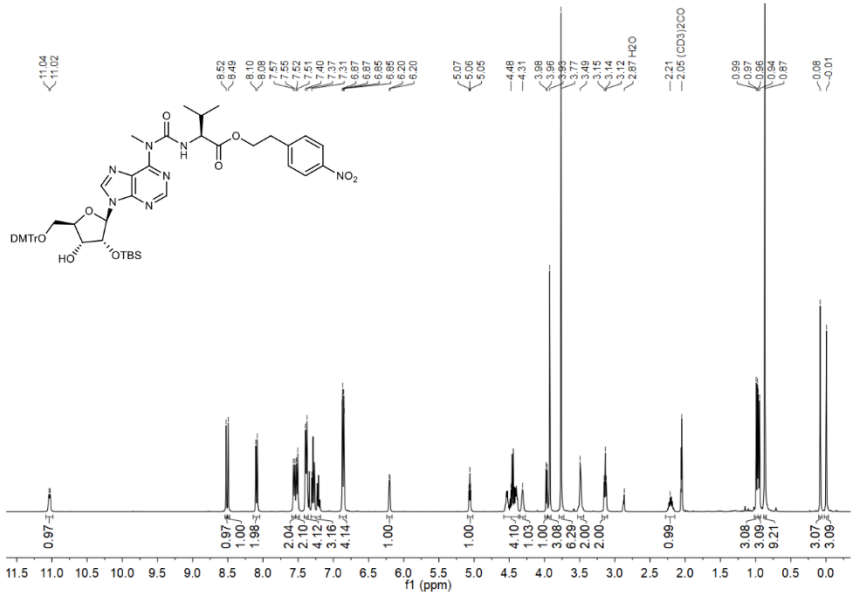


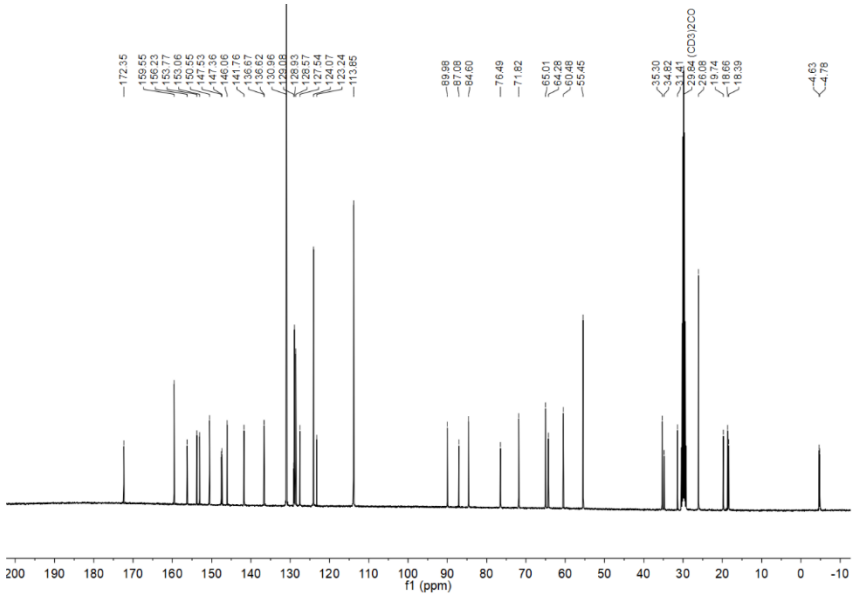
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13b



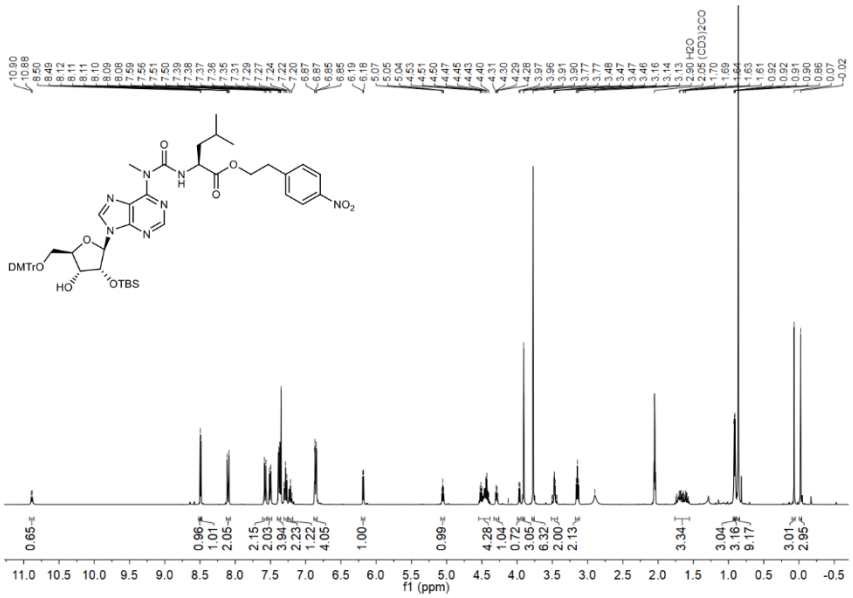


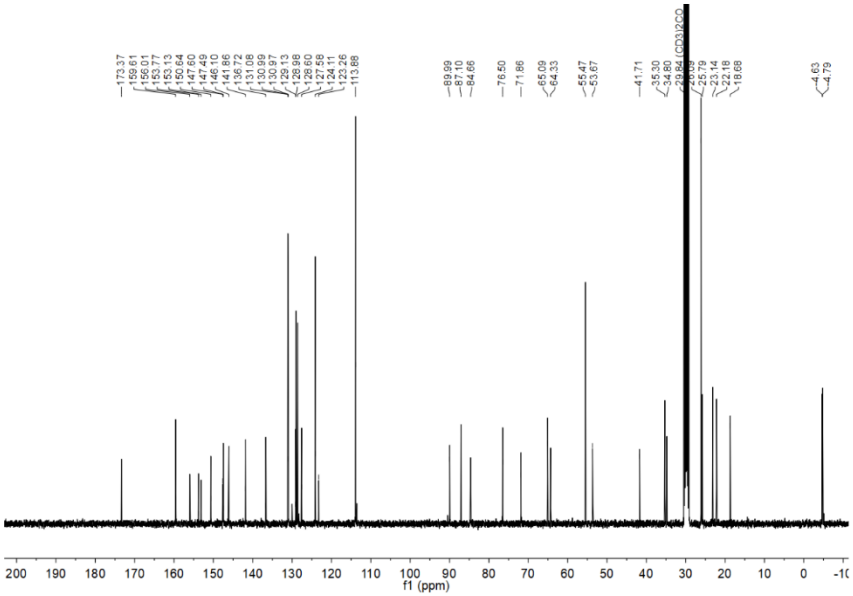
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13c



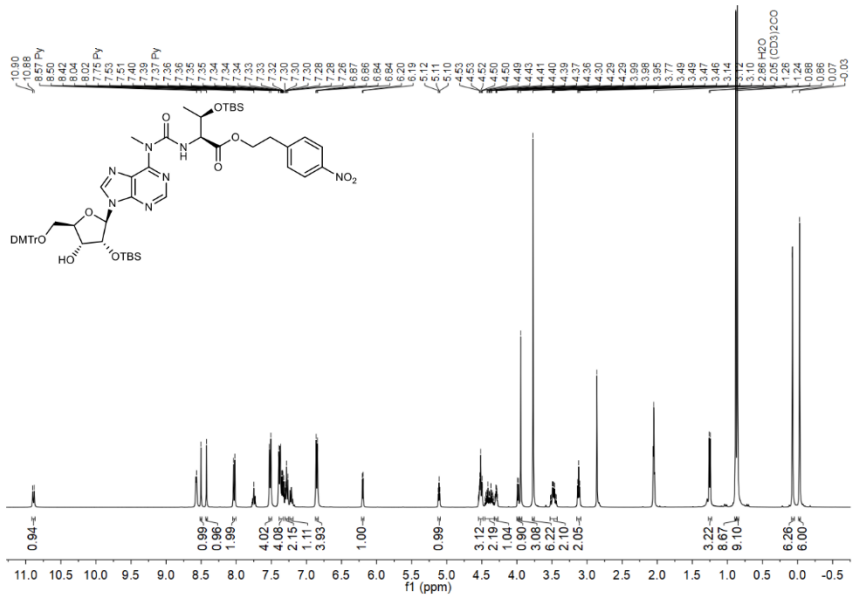


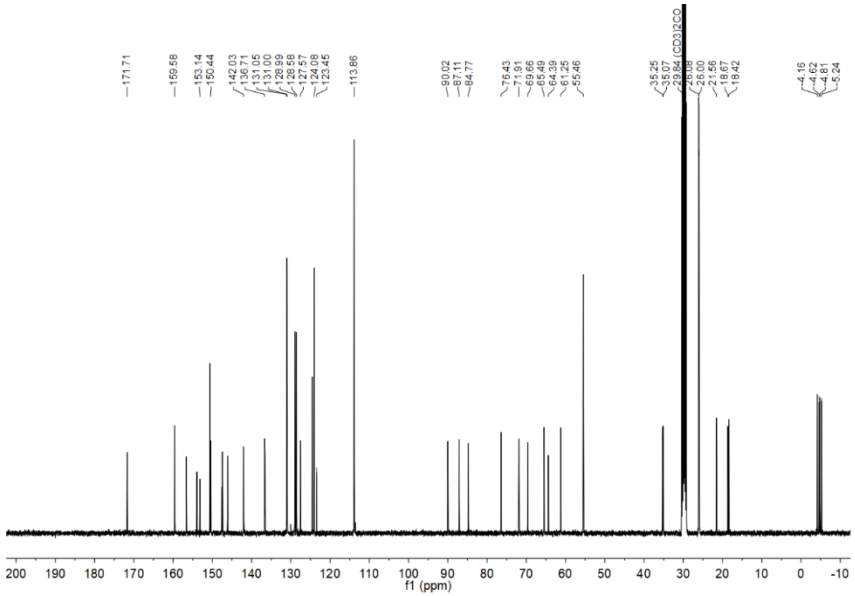
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13d



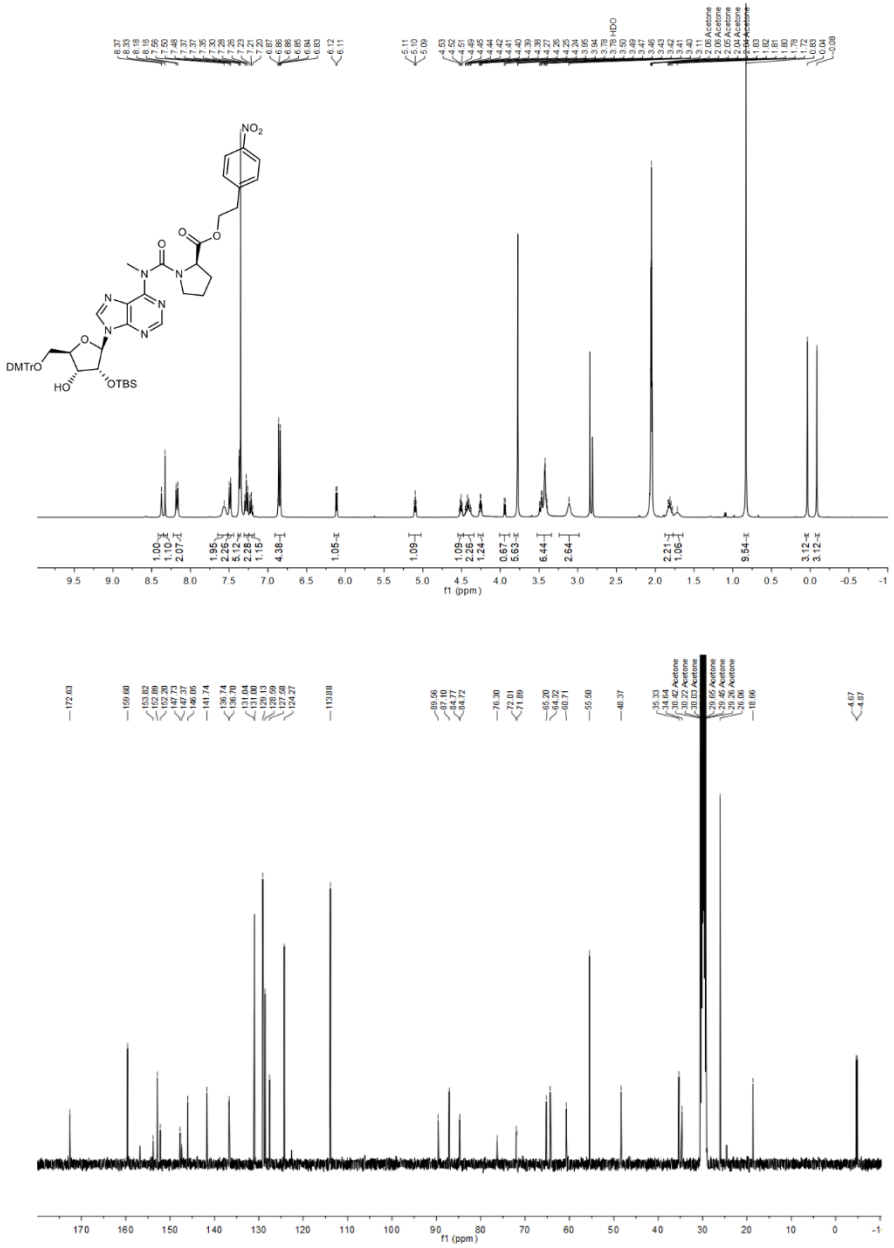


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13e

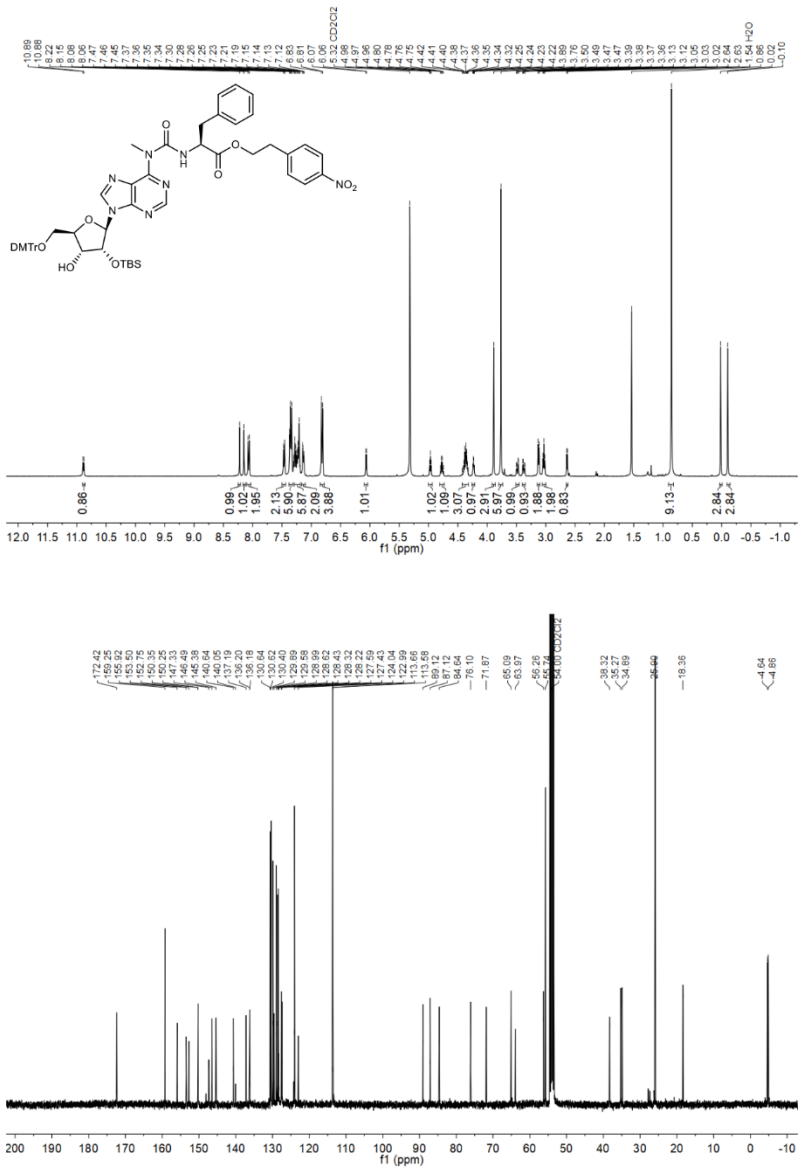




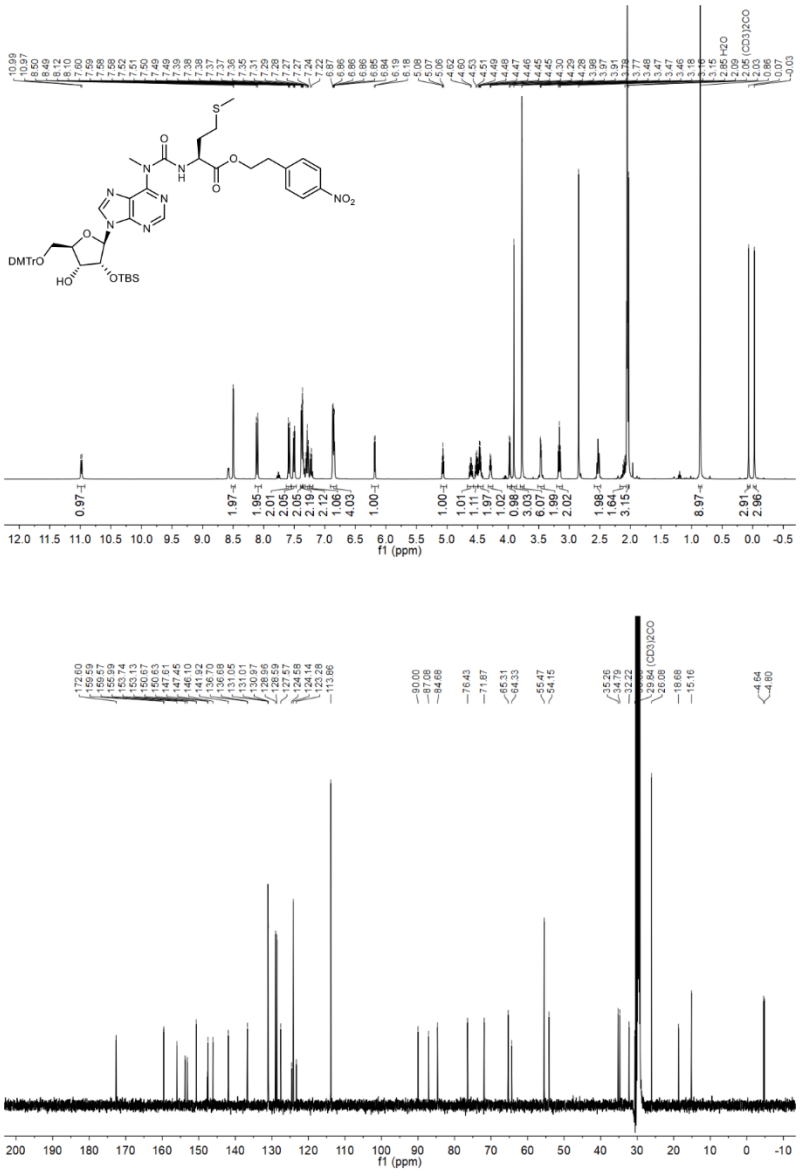
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13f



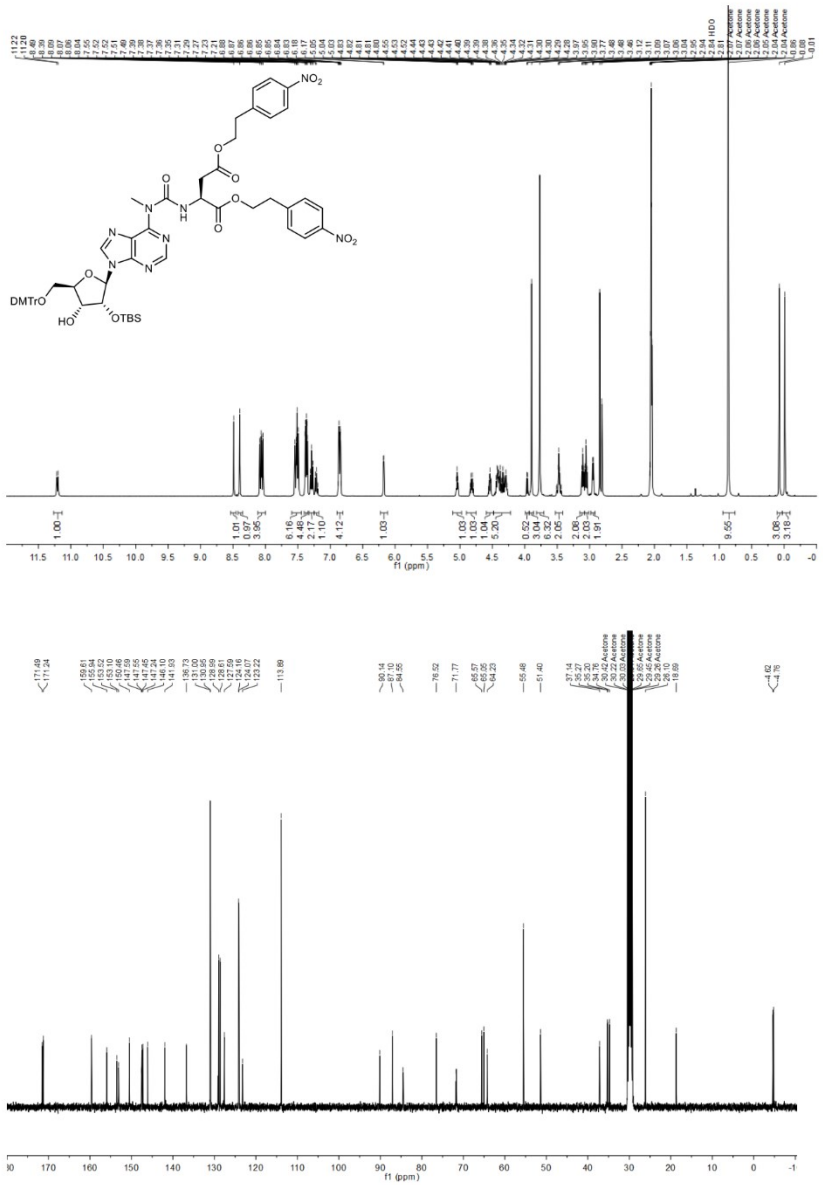
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13g



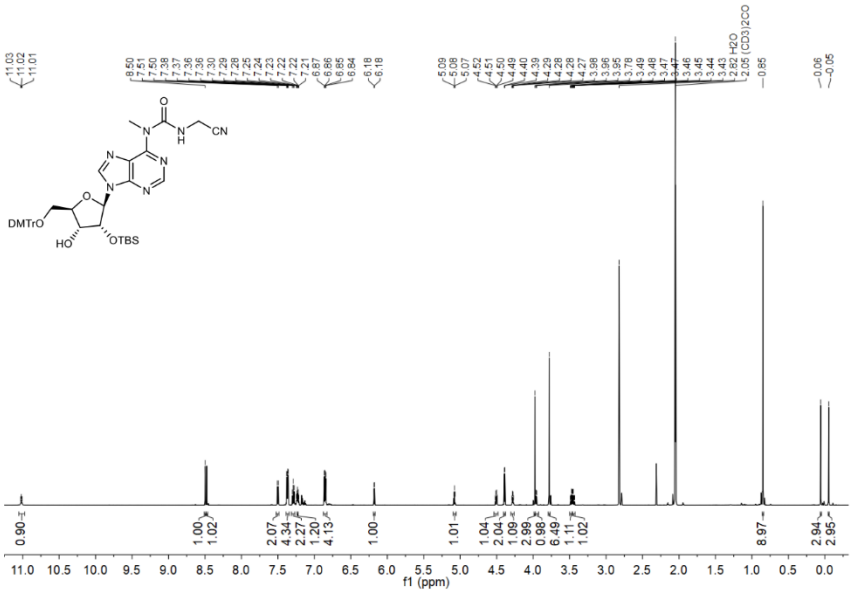
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13h

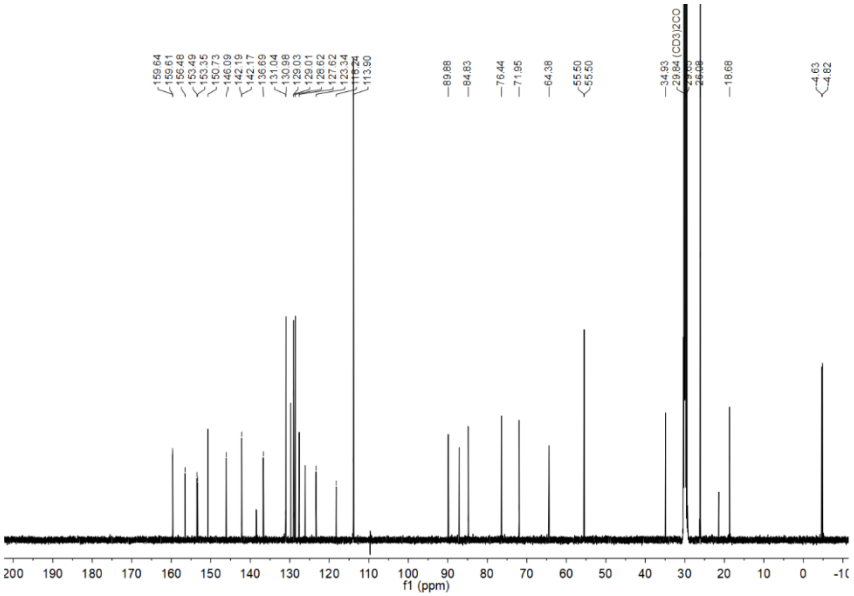


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13i

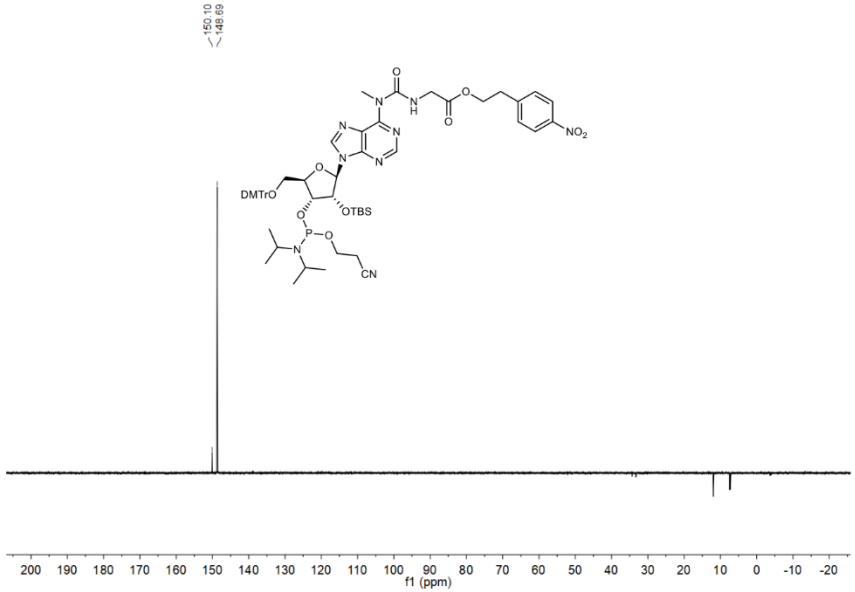


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13j

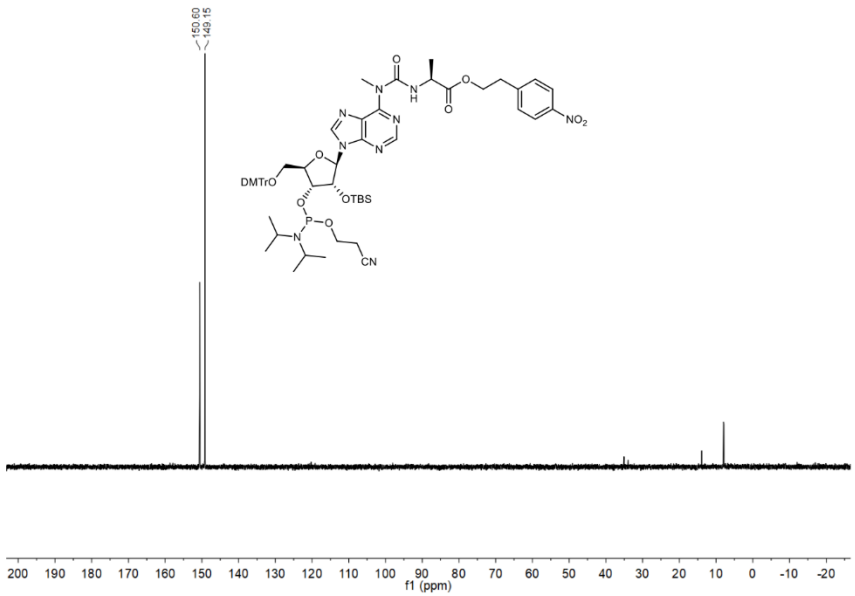




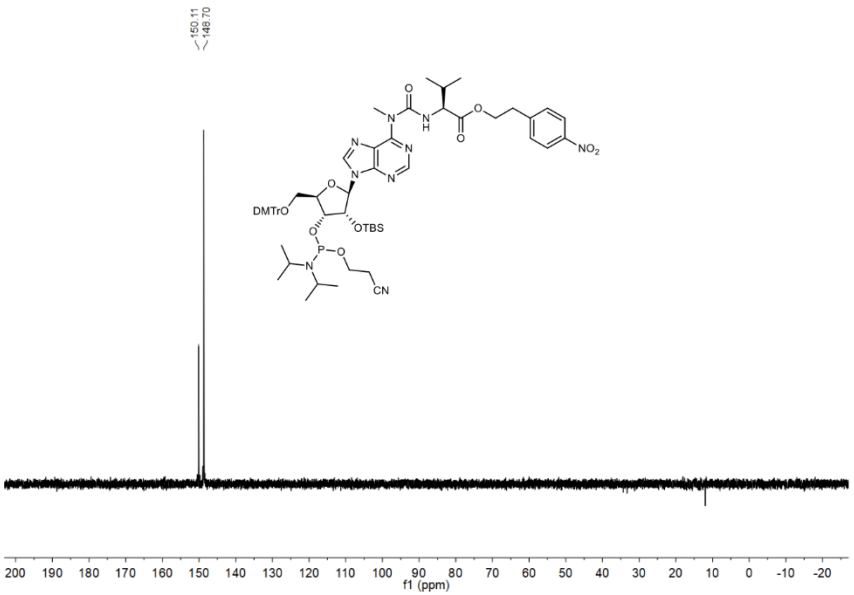
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14a



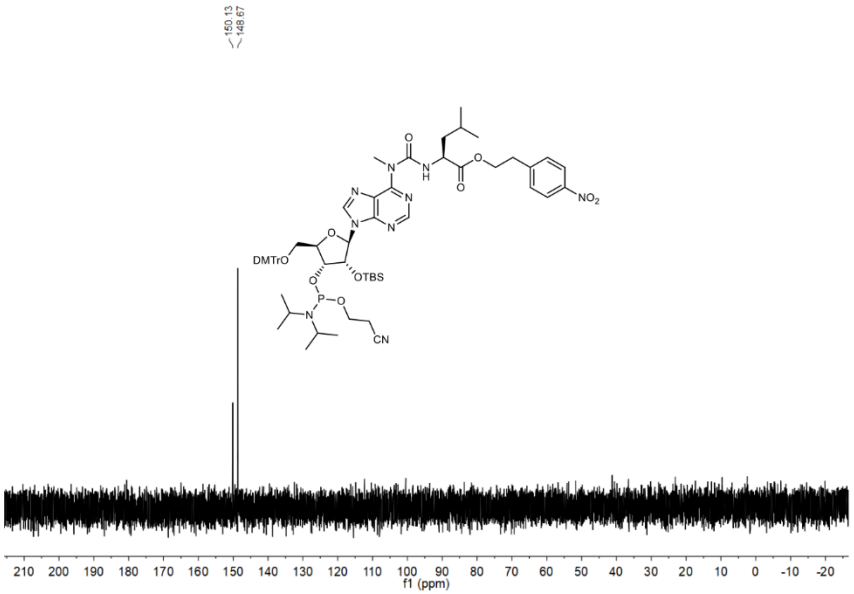
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14b



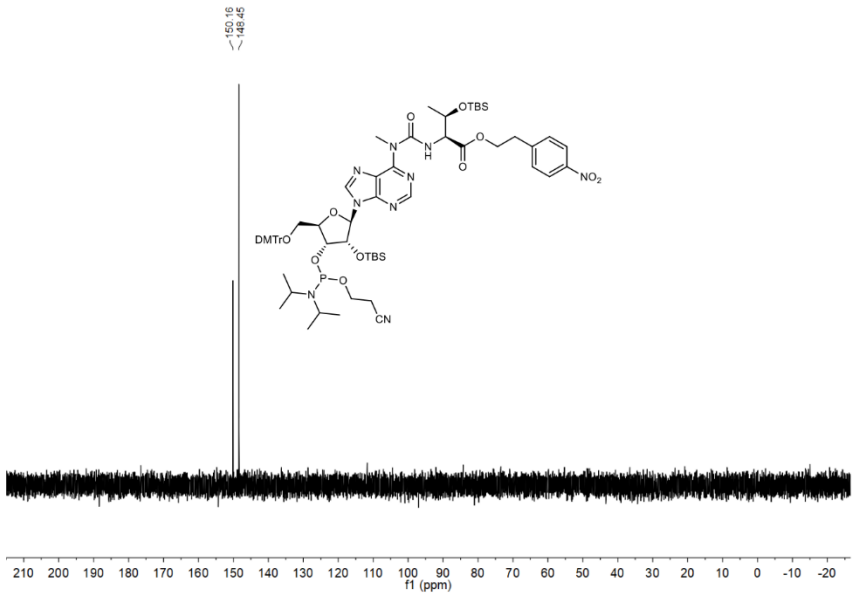
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14c



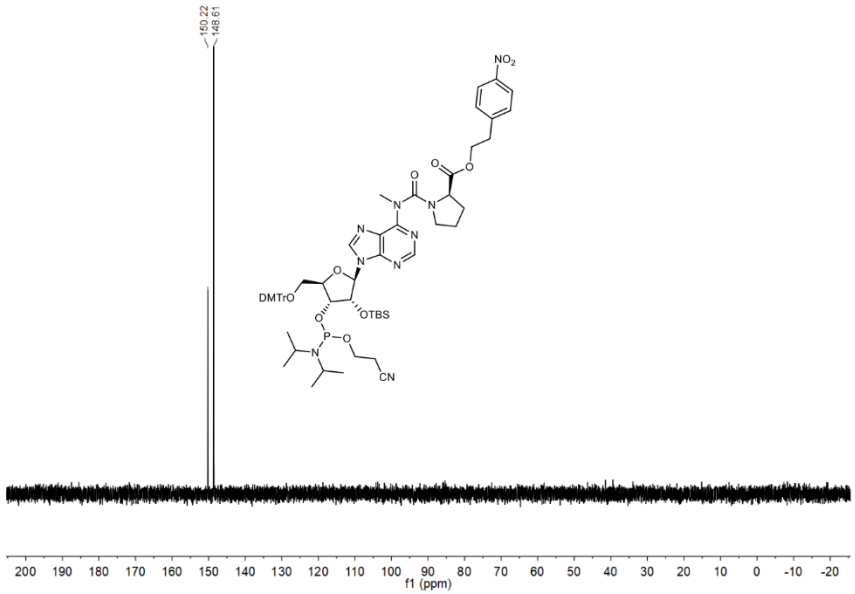
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14d



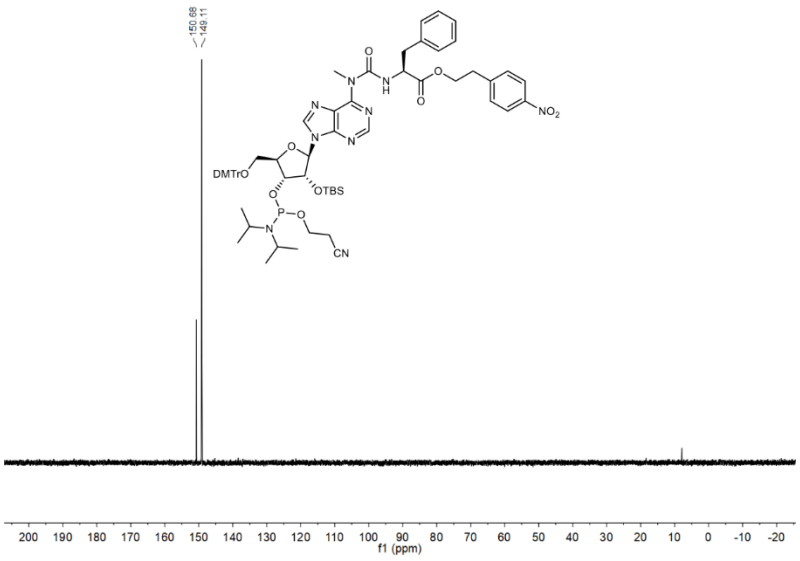
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14e



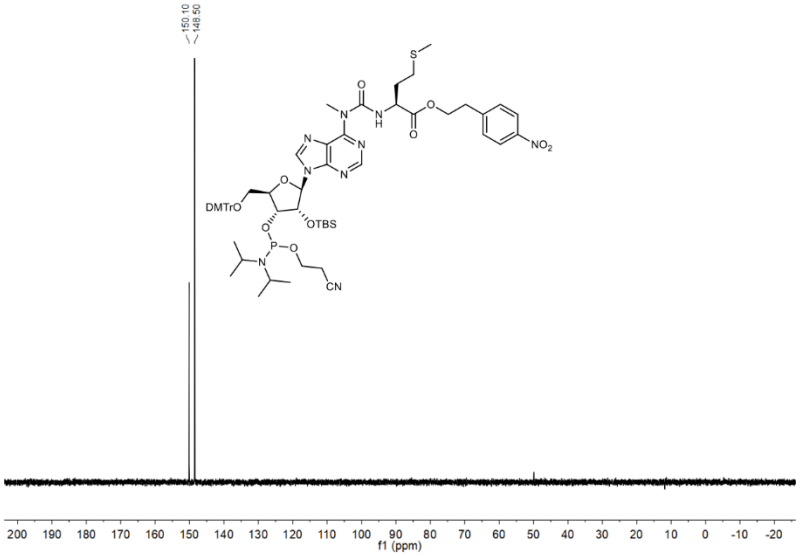
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14f



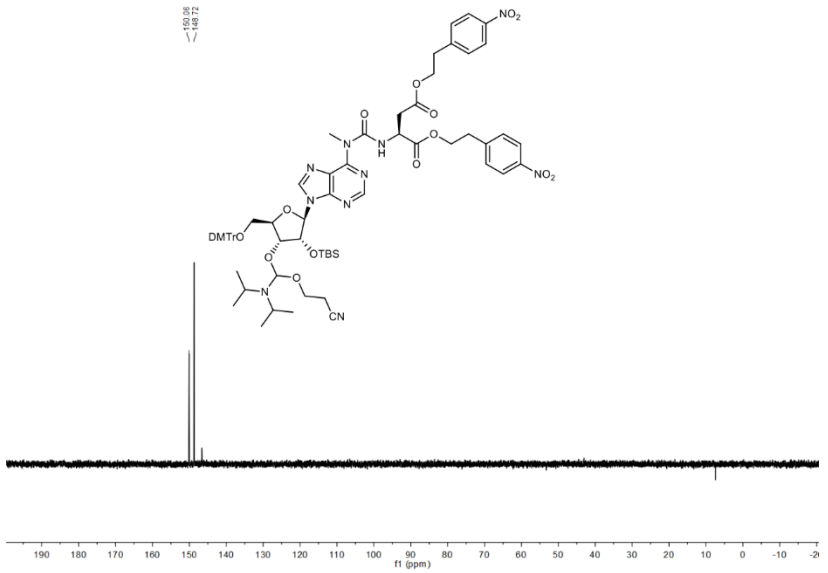
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14g



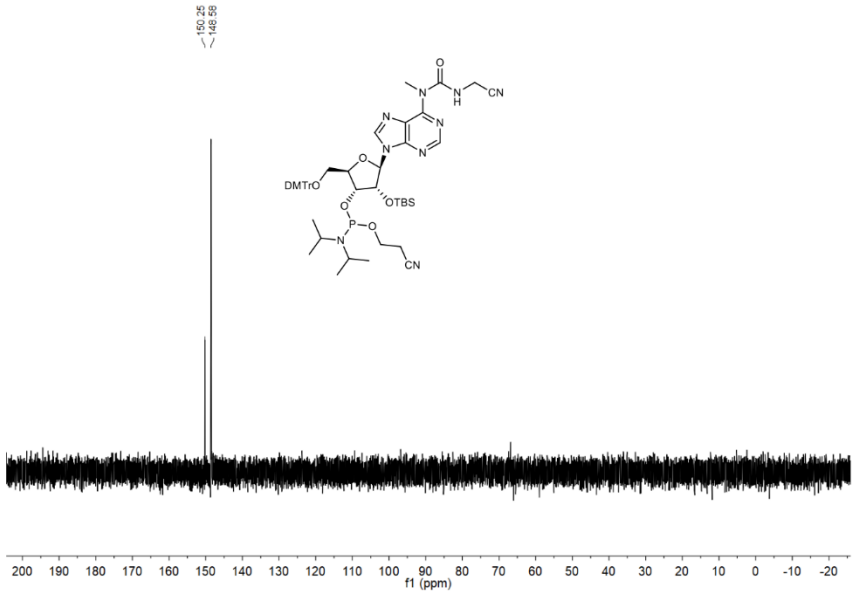
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14h



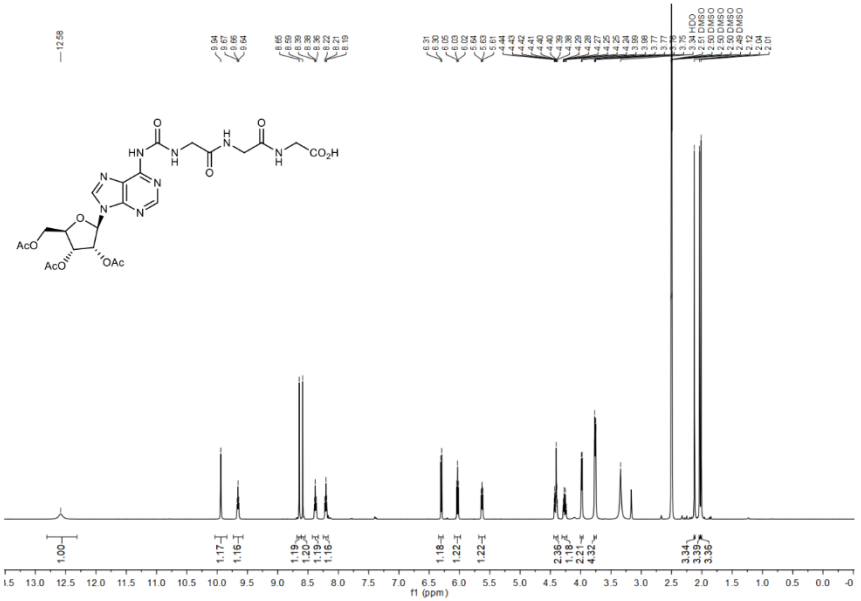
$^{31}P\{^{1}H\}$ NMR spectrum of compound 14i

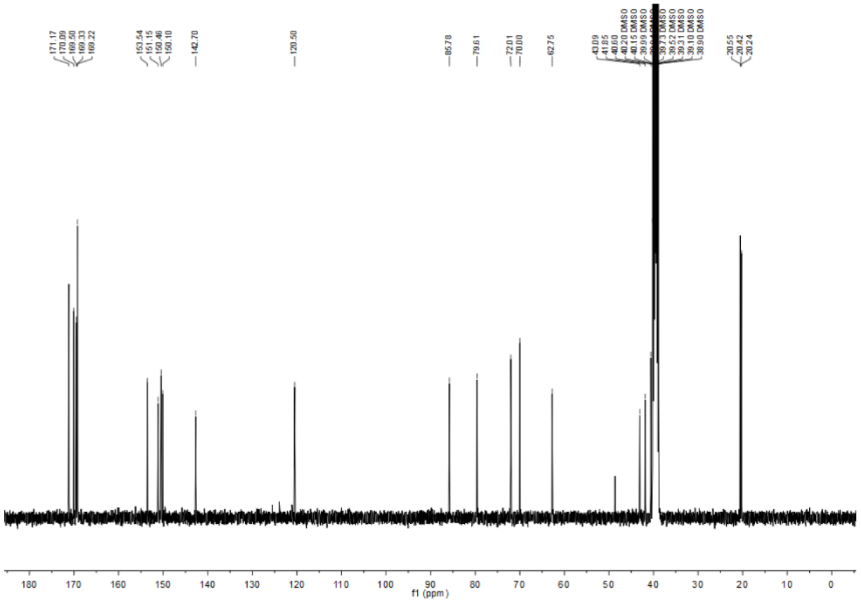


$^{31}P\{^{1}H\}$ NMR spectrum of compound 14j

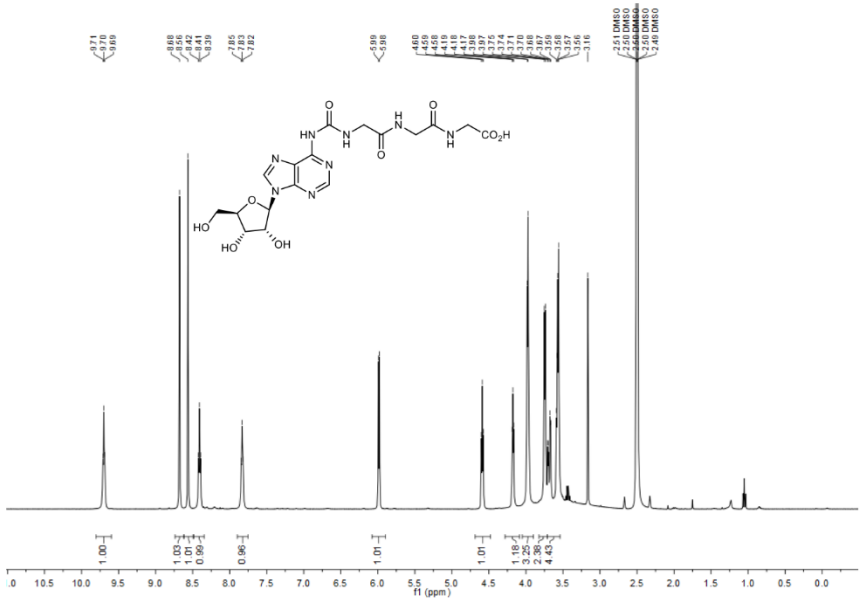


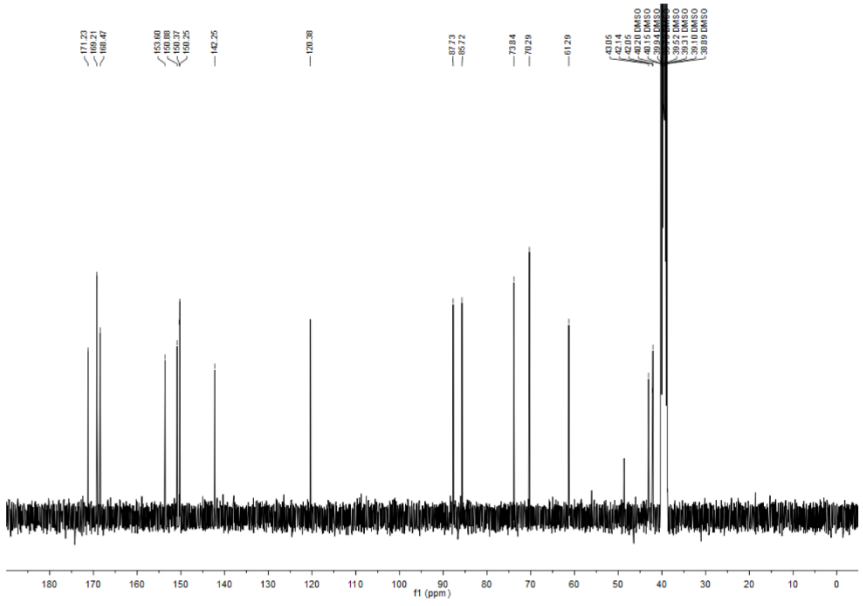
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 16



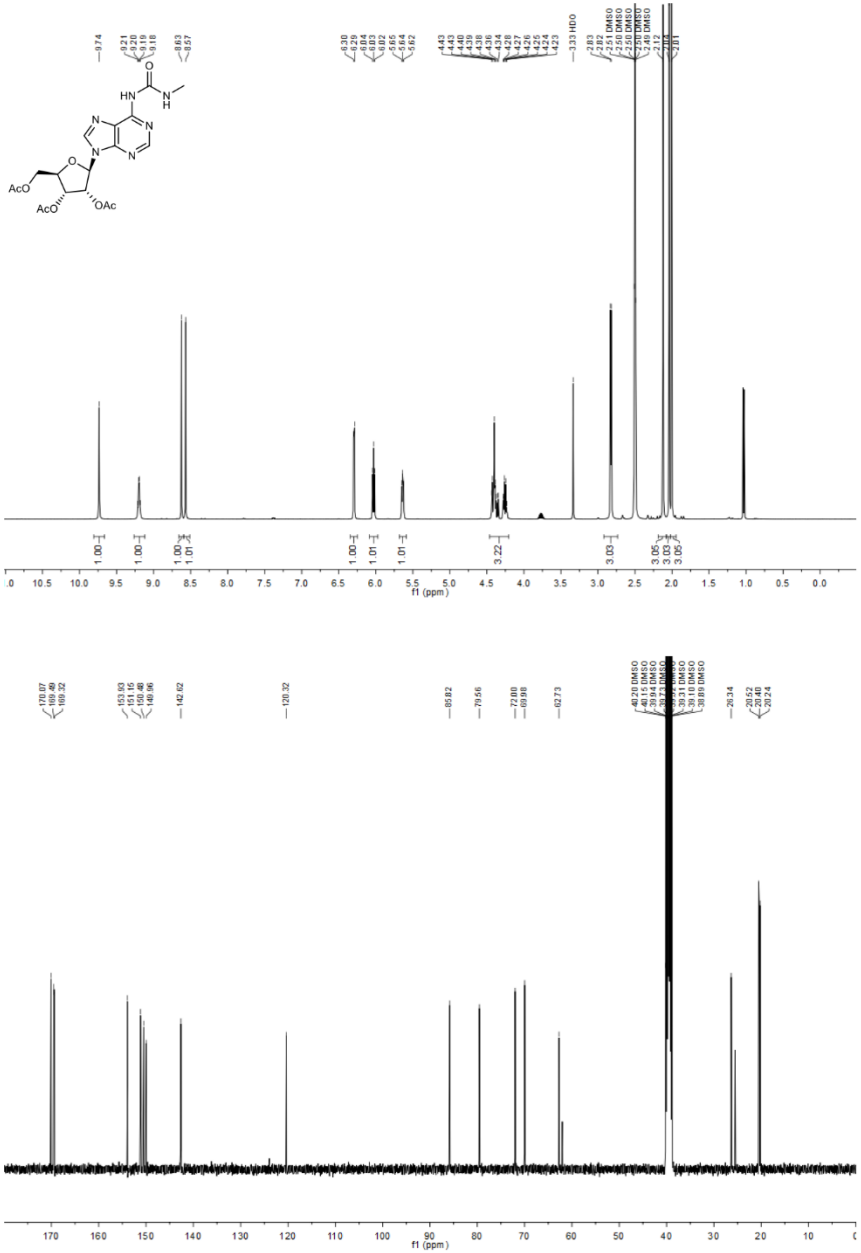


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 17

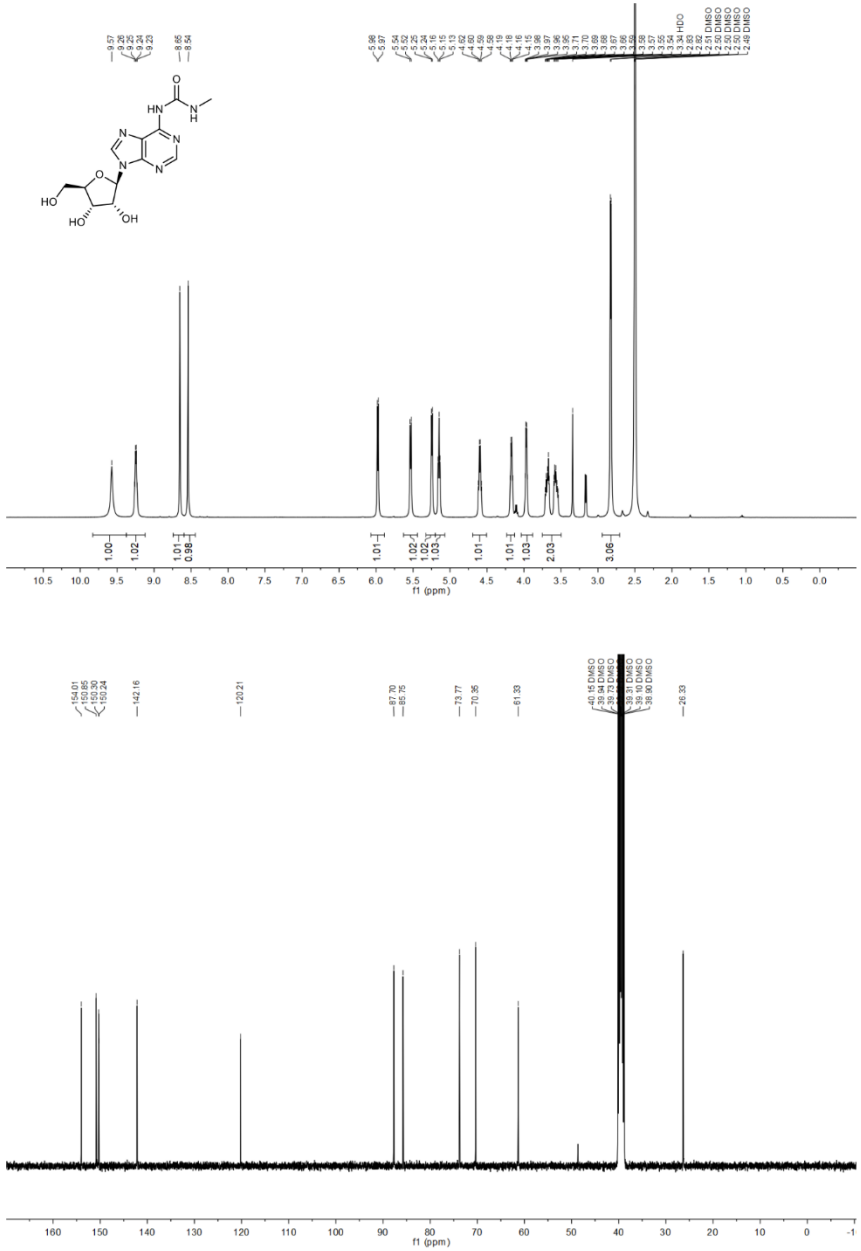




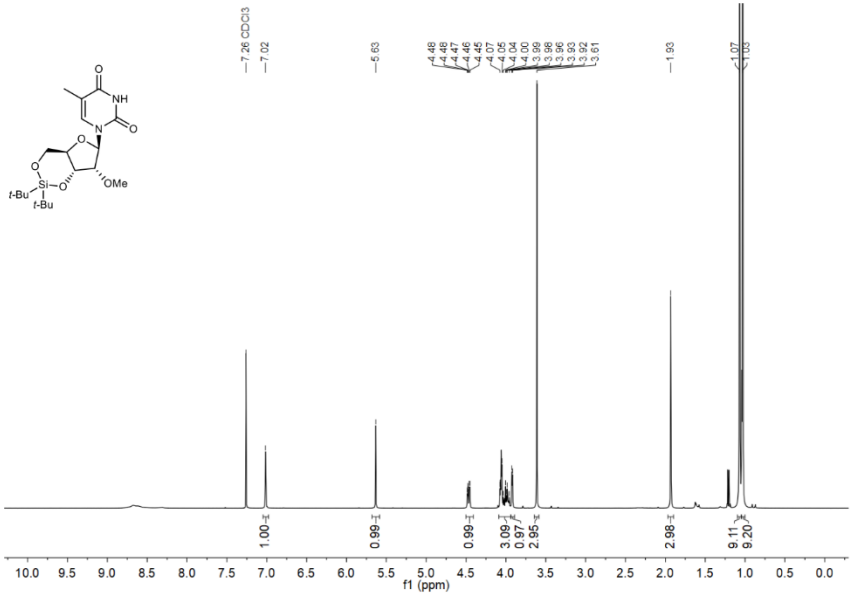
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 18

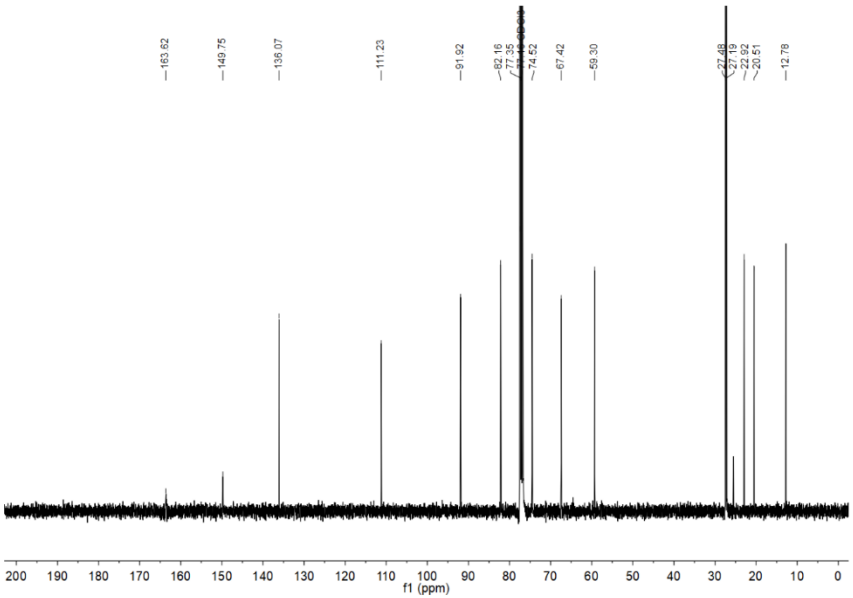


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 19

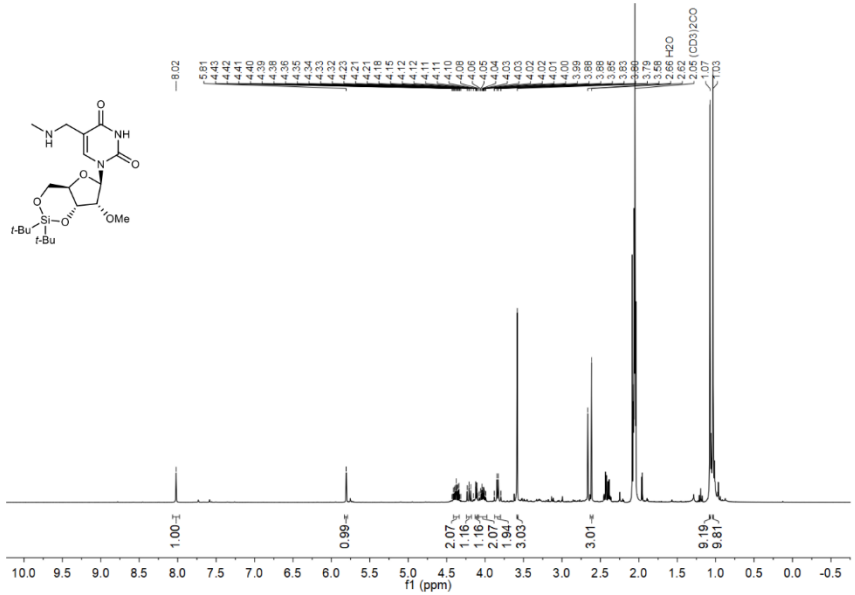


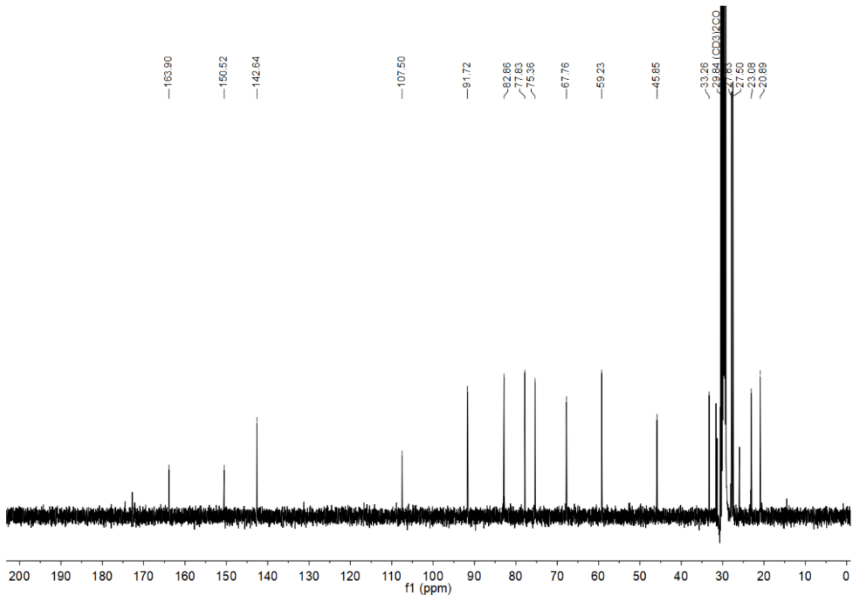
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 21



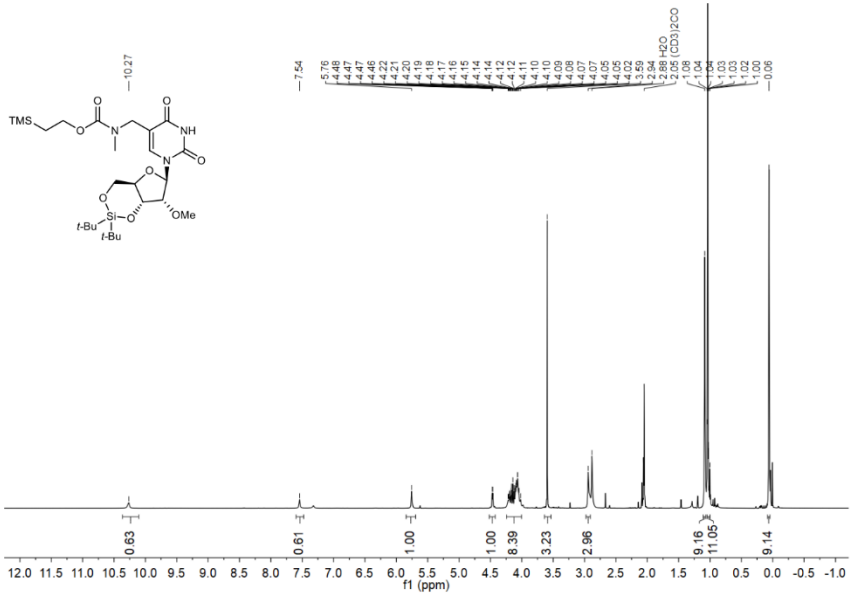


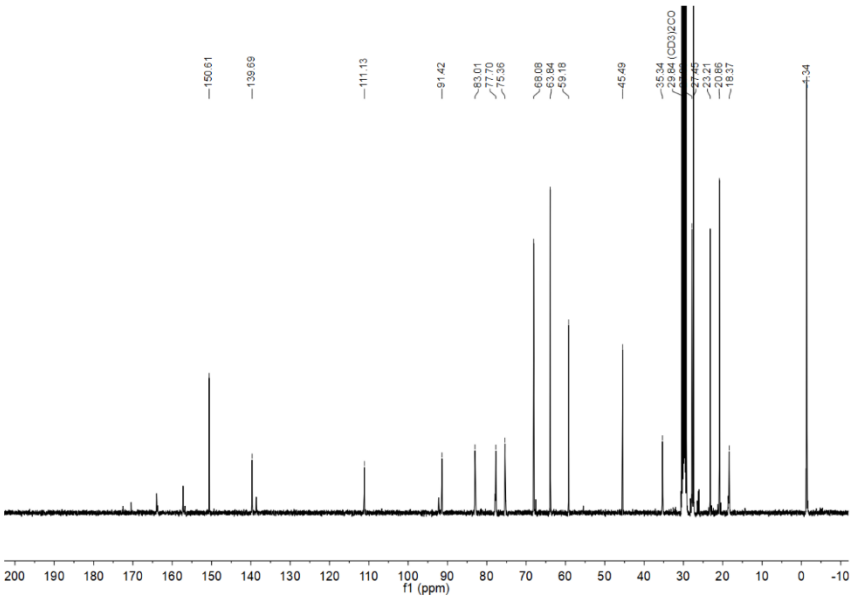
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 22



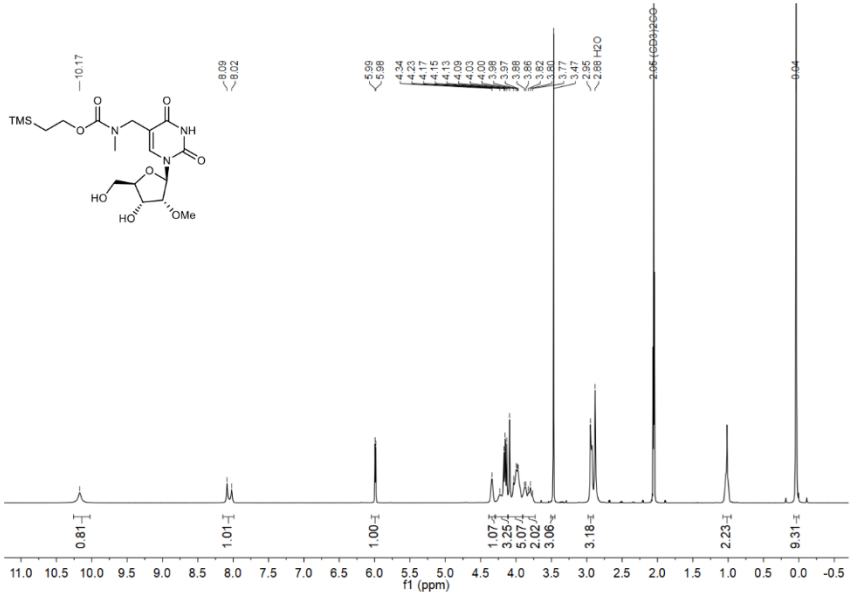


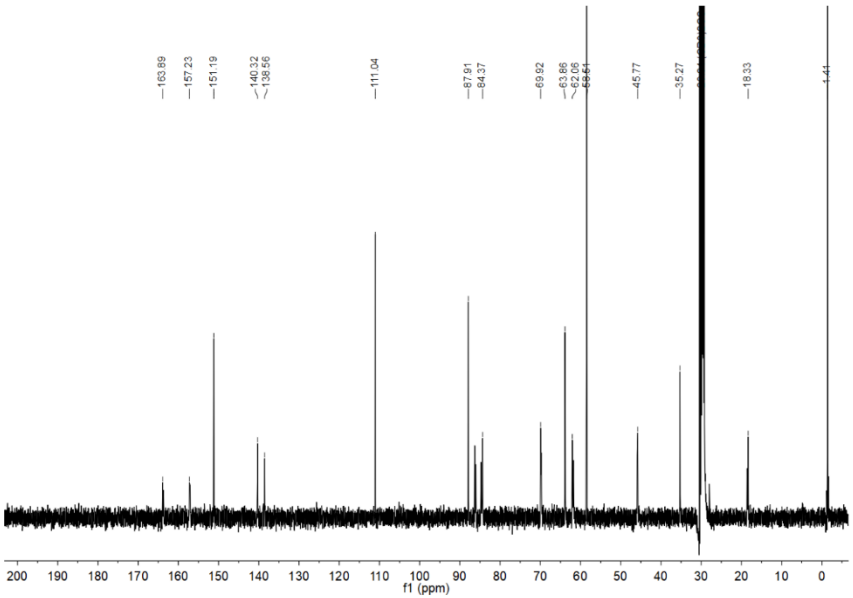
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 23



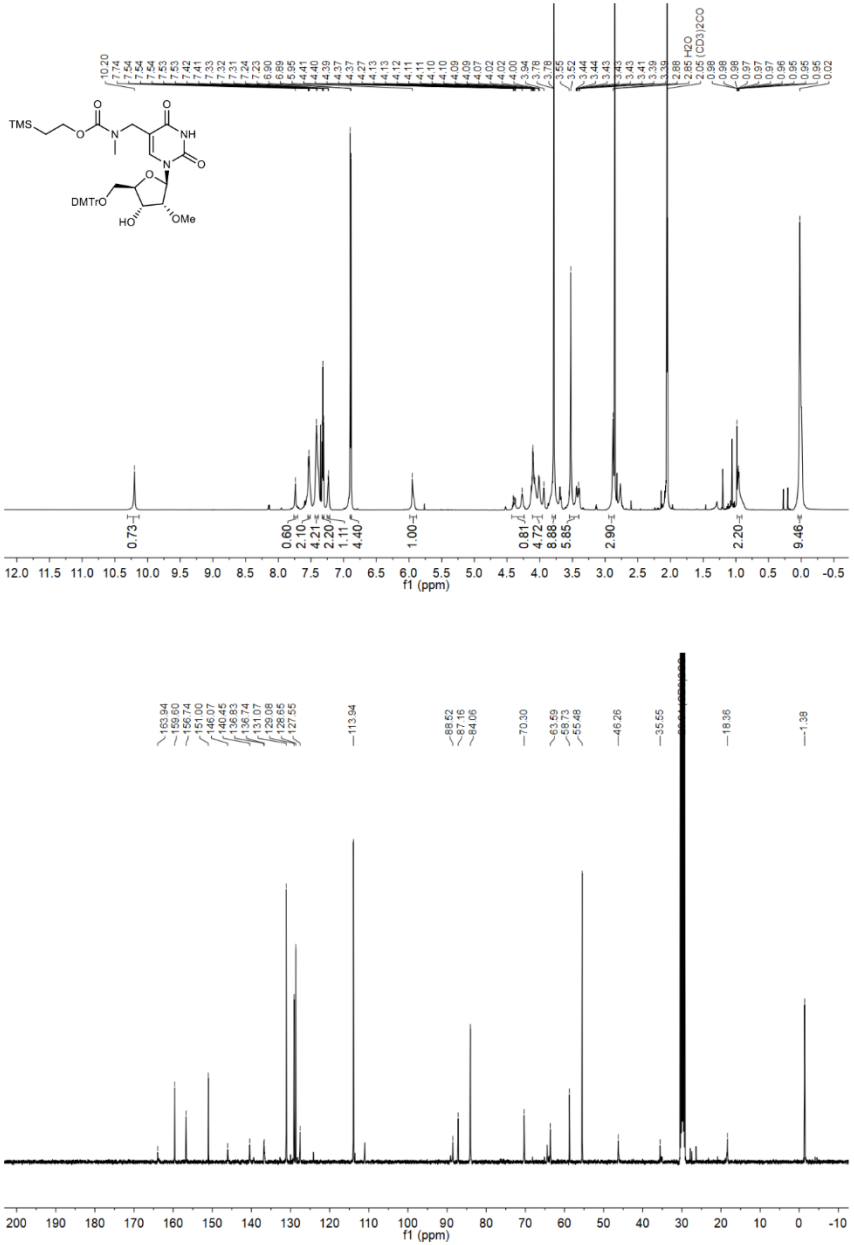


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 24

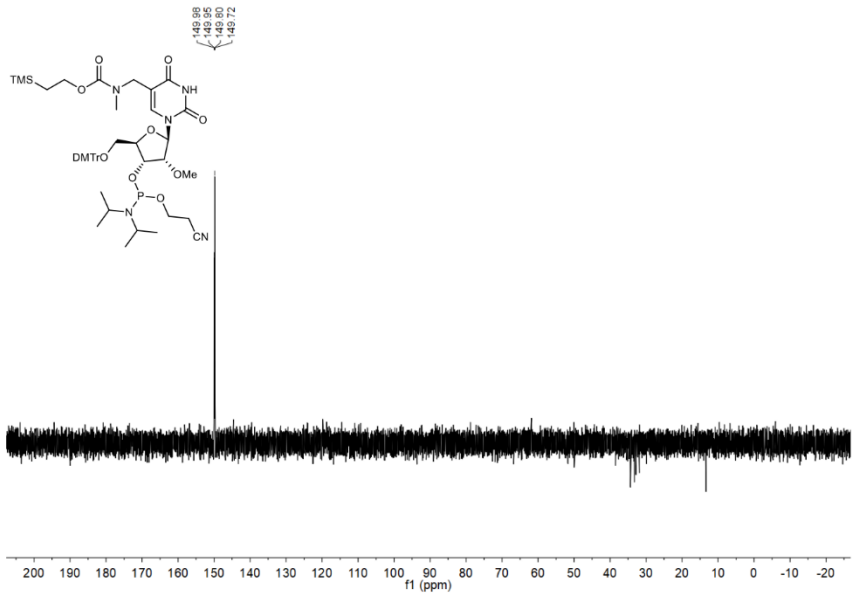




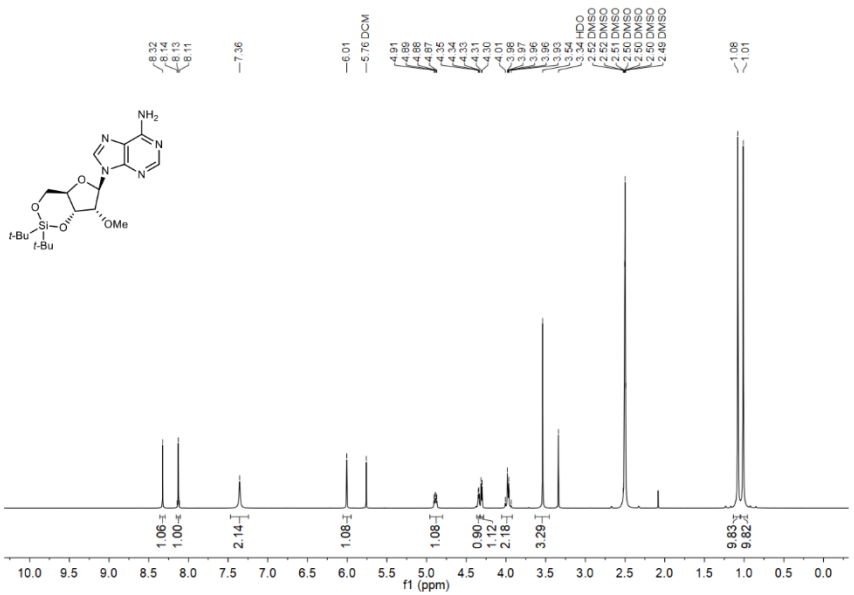
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 25

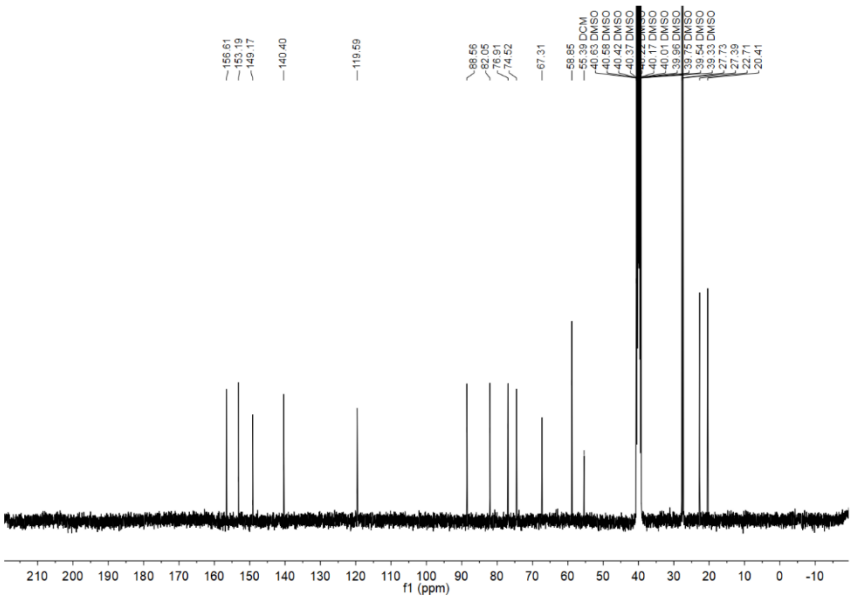


$^{31}P\{^{1}H\}$ NMR spectrum of compound 26

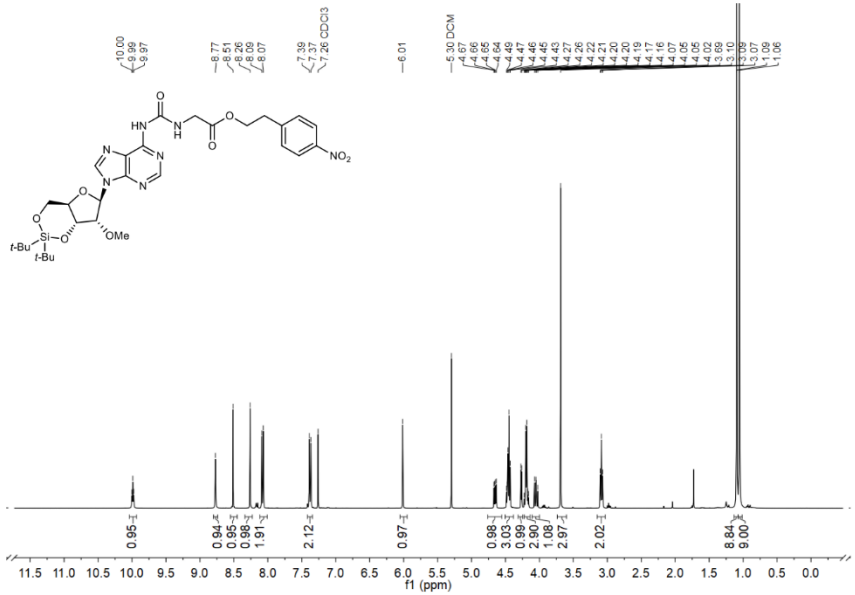


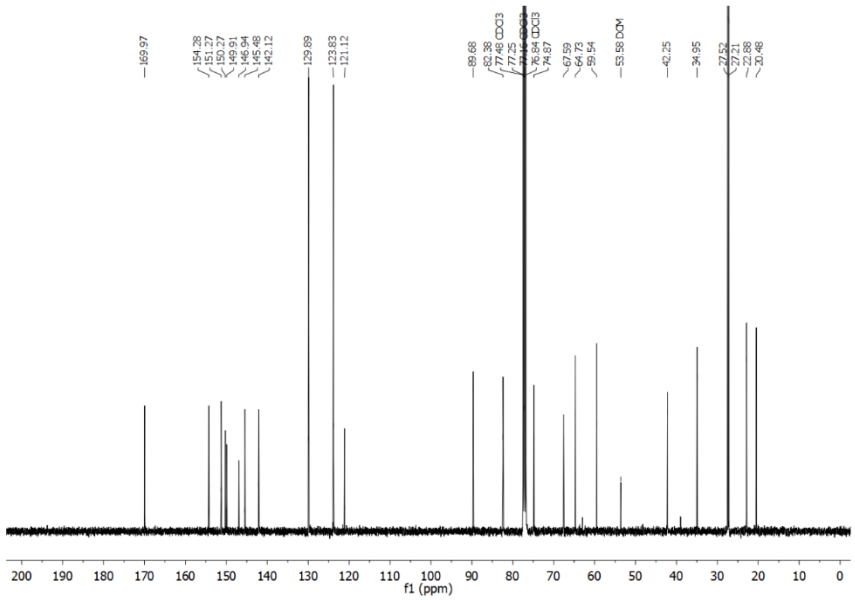
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 28



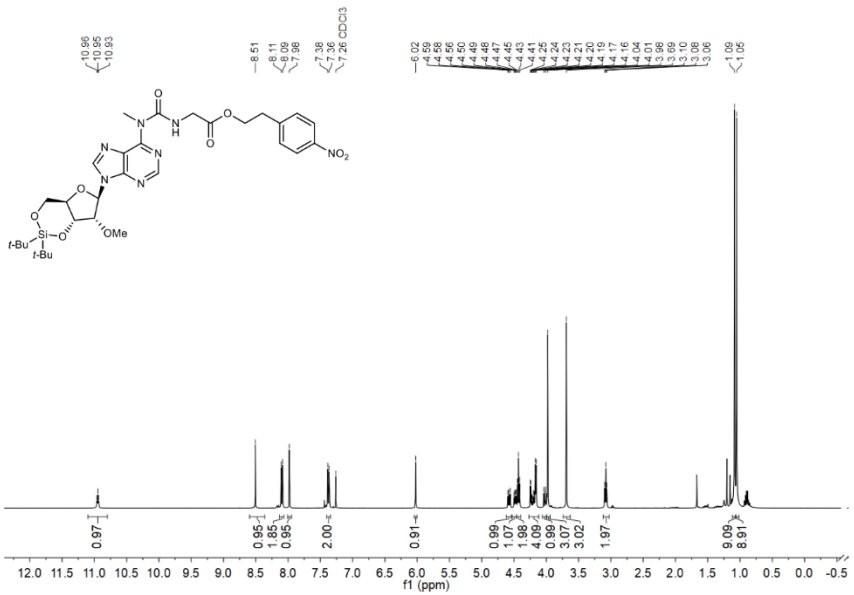


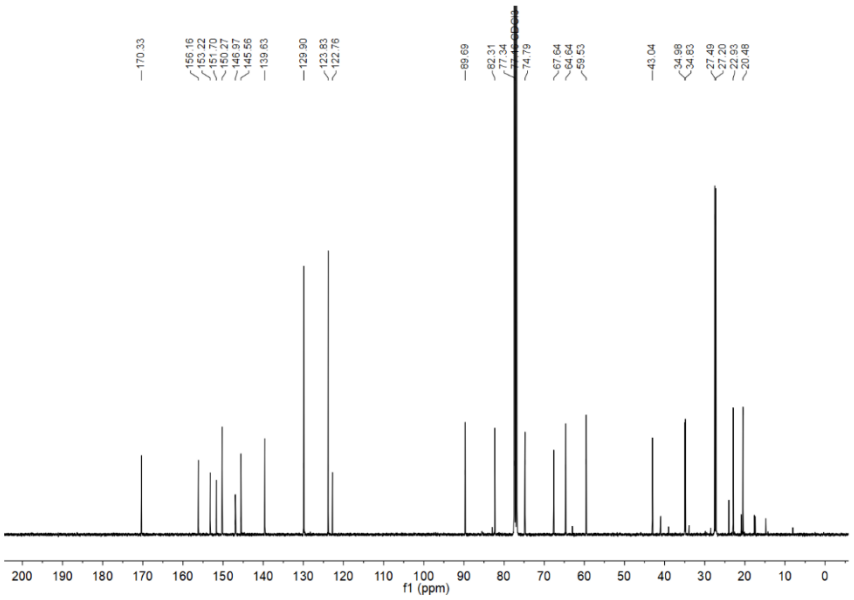
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 29



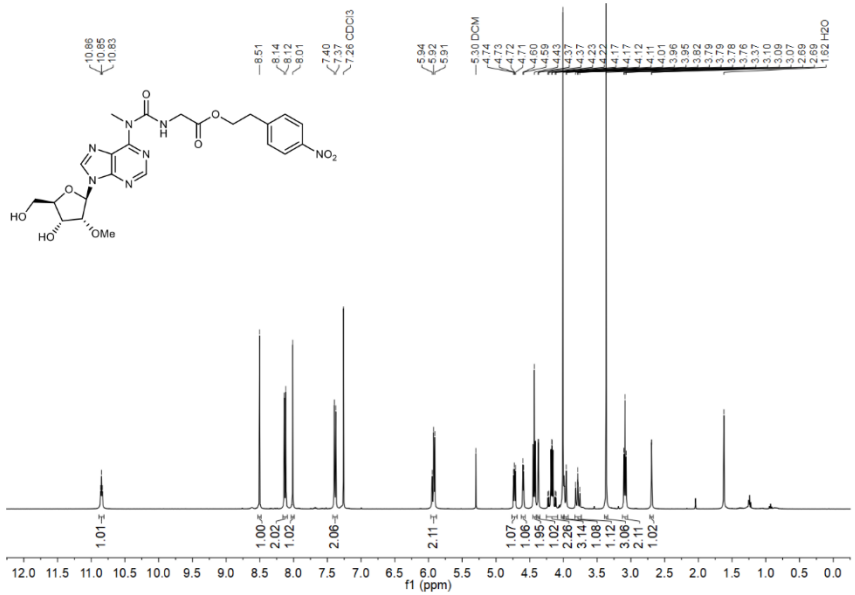


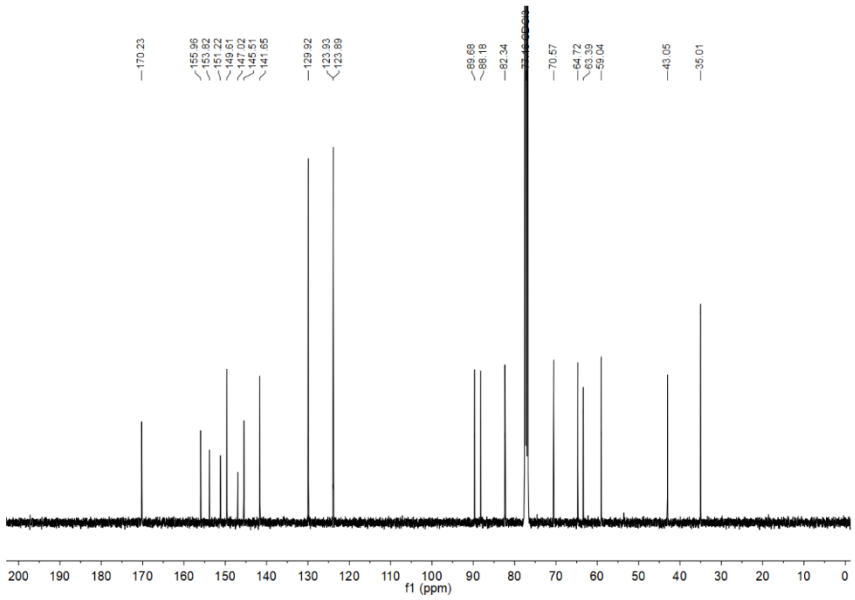
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 30



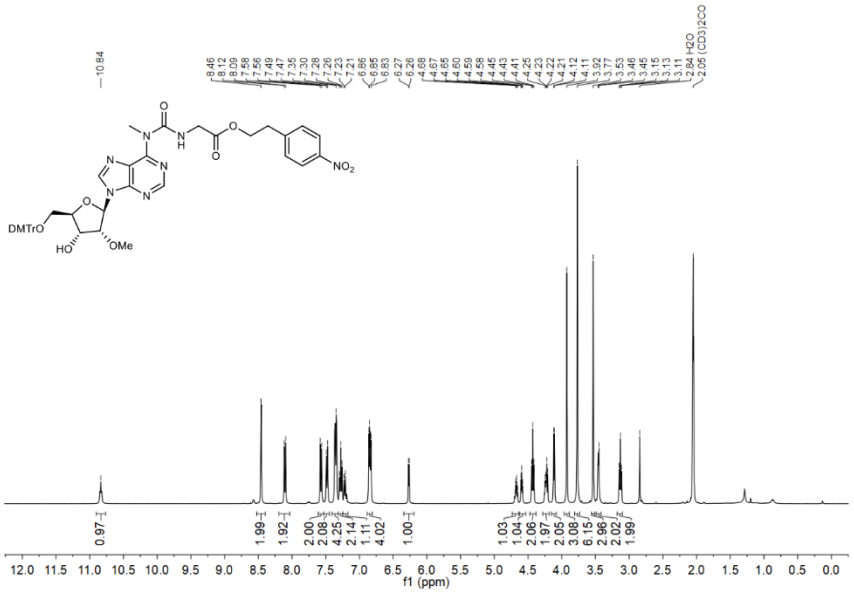


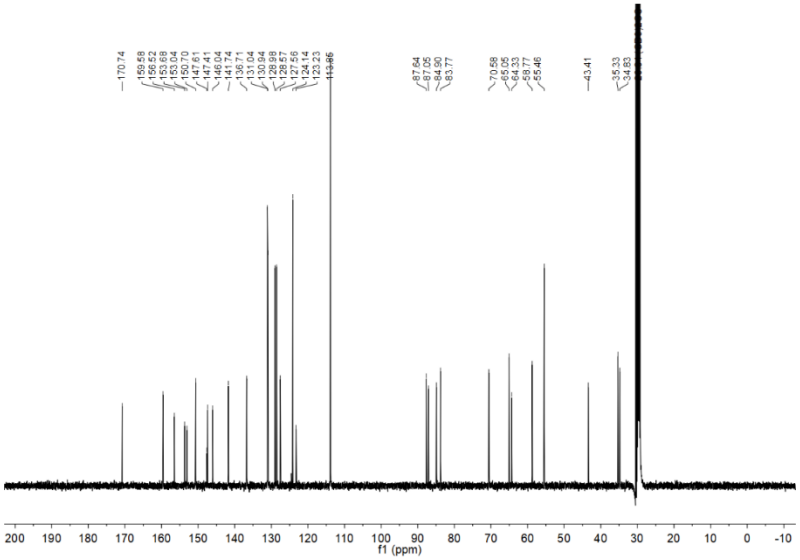
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 31



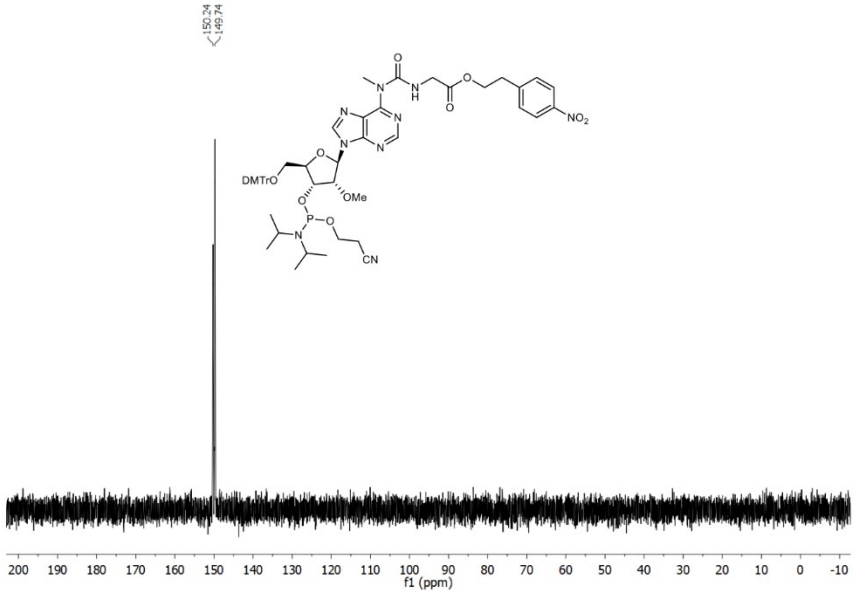


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 32





$^{31}P\{^{1}H\}$ 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 III

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Loading of Amino Acids onto RNA in a Putative RNA-Peptide World

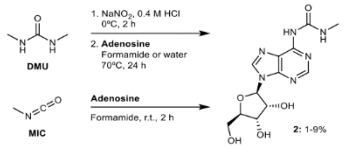
J. N. Singer, F. M. Müller, E. Węgrzyn, C. Hölzl, H. Hurmiz, C. Liu, L. Escobar, T. Carell**

Table of Contents

1. General information and instruments for nucleosides and phosphoramidites

Reagents were purchased from commercial suppliers and used without further purification unless otherwise stated. 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, $^{13}C\{^{1}H\}$ NMR and $^{31}P\{^{1}H\}$ NMR spectra 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), Bruker Ascend 500 spectrometer (500 MHz for ^{1}H NMR, 125 MHz for ^{13}C NMR and 202 MHz for ^{31}P NMR) or 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, HSQC 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 or Shimadzu IRSpirit FT-IR instrument. Both equipped with an ATR accessory. Column chromatography was performed with silica gel technical grade, 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. Prebiotic synthesis of N^6-methylurea adenosine



Scheme S1. Prebiotic synthesis of N^6-methylurea adenosine **2**.

Procedure for the prebiotic synthesis of 2 using DMU:

Step 1: The nitrosation reaction of 1,3-dimethylurea (**DMU**) was carried out following a procedure previously reported in the literature.[2] Step 2: Adenosine (2.67 mg, 10 µmol, 1 equiv.) and 1,3-dimethyl-1-nitrosourea (58.6 mg, 0.5 mmol, 50 equiv.) were dissolved in formamide (1 mL) or water (40 µL). The reaction was stirred at 70°C for 24 h. For the reaction in water solution, the crude was diluted with water (up to 1.0 mL) after 24 h. An aliquot (10 µL) of the crude reaction mixture was taken, diluted with water (up to 1.0 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 with B: 0-15% from 0 to 15 min and 15-20% from 15 to 30 min; Temperature = 40°C; Flow rate = 0.15 mL·min^{-1} and Injection volume = 5 µL).

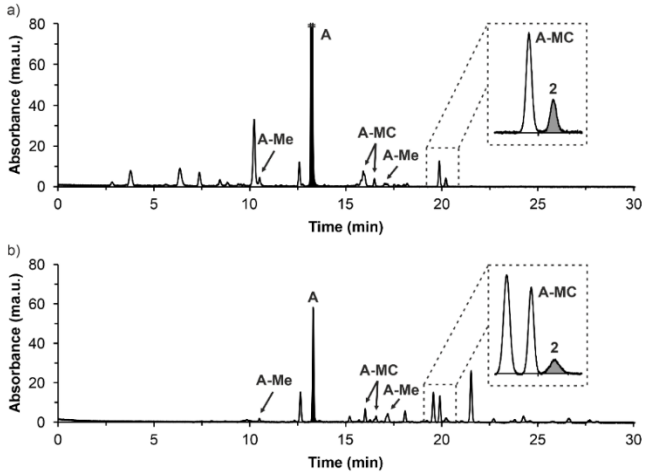


Figure S1. HPLC chromatograms of the crude reaction mixtures for the prebiotic synthesis of **2** using 1,3-dimethylurea (**DMU**) in: a) formamide and b) water. The chromatographic peaks assigned as **A-Me** corresponded to methylated Adenosine derivatives. In turn, the chromatographic peaks assigned as **A-MC** corresponded to Adenosine derivatives bearing a *N*-methylcarbamoyl substituent at the OH groups of the ribose. Structural assignments for **A-Me** and **A-MC** were not performed.

Procedure for the prebiotic synthesis of 2 using MIC:

Adenosine (10.0 mg, 37.4 µmol, 1 equiv.) and methylisocyanate (**MIC**, 12.2 µL, 206 µmol, 5.5 equiv.) were dissolved in formamide (1 mL). The reaction was stirred at r.t. for 2 h. An aliquot (25 µL) of the crude reaction mixture was taken, diluted with water (up to 1.0 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 with B: 0-25% from 0 to 45 min; Temperature = 40°C; Flow rate = 0.15 mL·min^{-1} and Injection volume = 5 µL).

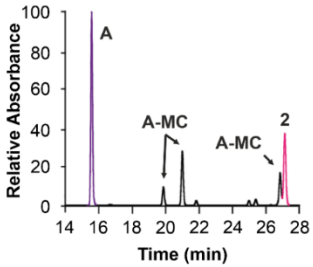


Figure S2. HPLC chromatogram of the crude reaction mixture for the prebiotic synthesis of **2** using methyl isocyanate (**MIC**) in formamide. The chromatographic peaks assigned as **A-MC** corresponded to Adenosine derivatives bearing a *N*-methylcarbamoyl substituent at the OH groups of the ribose. Structural assignments for **A-MC** were not performed.

Procedure for the hydrolysis reaction of the crude reaction mixture obtained from the prebiotic synthesis of 2 using DMU:

The crude reaction mixture in formamide solution (100 µL) was diluted with 50 mM borate buffer pH 9.5 (up to 1.0 mL). The reaction was stirred at 70°C for 24 h. An aliquot (100 µL) of the crude reaction mixture was taken, diluted with water (up to 1.0 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 with B: 0-15% from 0 to 15 min and 15-20% from 15 to 30 min; Temperature = 40°C; Flow rate = 0.15 mL·min^{-1} and Injection volume = 5 µL).

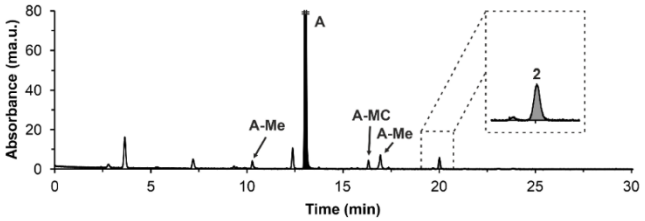
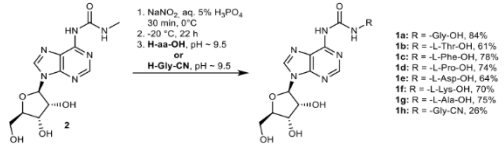


Figure S3. HPLC chromatogram of the crude reaction mixture after the hydrolysis reaction with 50 mM borate buffer pH 9.5 at 70°C for 24 h. The chromatographic peaks assigned as **A-Me** corresponded to methylated Adenosine derivatives. In turn, the chromatographic peak assigned as **A-MC** corresponded to an Adenosine derivative bearing a *N*-methylcarbamoyl substituent at one of the OH groups of the ribose. Structural assignments for **A-Me** and **A-MC** were not performed.

3. Prebiotic synthesis of amino acid-modified carbamoyl nucleosides

3.1 Prebiotic synthesis of a series of amino acid-modified N^6-carbamoyl adenosine nucleosides



Scheme S2. Prebiotic synthesis of amino acid-modified N^6-carbamoyl adenosine nucleosides **1a-h** from N^6-methylurea adenosine **2** (optimized reaction conditions are shown).

Optimization of the reaction conditions using 2 and H-Gly-OH:

Step 1: N^6-methylurea adenosine **2** (1.00 mg, 3.08 µmol, 1.0 equiv.) and $NaNO_2$ (2.66 mg, 38.54 µmol, 12.5 equiv.) were dissolved in an acidic aqueous solution (150 µL, see Table S1 for acids used). The reaction was stirred at 0°C (see Table S2 for time). Step 2: After that, the solution was kept in a freezer at -20°C (see Table S3 for time). Step 3: The thawed adenosine's solution was transferred to a 30 mM borate-buffered solution (3.00 mL) containing the amino acid **H-Gly-OH** (2.31 mg, 30.84 µmol, 10.0 equiv.). The pH was adjusted using a 4 M aqueous NaOH solution (see Table S4 for pH). The reaction was stirred at r.t. for 1 h. Finally, an aliquot (25 µL) of the crude reaction mixture was taken, diluted with water (up to 1.0 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 25 min; Flow rate = 0.15 mL·min^{-1} and Injection volume = 5 µL).

Table S1. Screening of acidic aqueous solutions in step 1. Reaction conditions in other steps were kept constant: step 1) 30 min at 0ºC; step 2) 22 h at -20ºC and step 3) pH ~ 9.5.

Acidic aqueous solution in step 1	Yield (%) of 1a
1% H_3PO_4	3
5% H_3PO_4	**84**
1 M HCl	78
1 M H_2SO_4	64
10% acetic acid (AcOH)	Not detected
neat acetic acid (AcOH)	Not detected
50% formic acid (FA)	3
neat formic acid (FA)	18

Table S2. Screening of reaction time in step 1. Reaction conditions in other steps were kept constant: step 1) 5% H_3PO_4 at 0ºC; step 2) 22 h at -20ºC and step 3) pH ~ 9.5.

Time (min) in step 1	Yield (%) of 1a
30	**84**
60	76
120	57

Table S3. Screening of reaction time in step 2. Reaction conditions in other steps were kept constant: step 1) 5% H_3PO_4 at 0ºC for 30 min; step 2) -20ºC and step 3) pH ~ 9.5.

Time (h) in step 2	Yield (%) of 1a
0	4
1	16
3	42
22	**84**

Table S4. Screening of pH values in step 3. Reaction conditions in other steps were kept constant: step 1) 5% H_3PO_4 at 0ºC for 30 min and step 2) 22 h at -20ºC.

pH of aqueous solution in step 3	Yield (%) of 1a
~6.2	5
~6.9	24
~7.4	34
~8.6	60
~9.5	**84**

General procedure for the prebiotic synthesis of 1a-h under the optimized reaction conditions:

In the general procedure, the amino acid was added in step 3.

Step 1: N^6-methylurea adenosine **2** (1.00 mg, 3.08 µmol, 1.0 equiv.) and $NaNO_2$ (2.66 mg, 38.54 µmol, 12.5 equiv.) were dissolved in 5% aqueous H_3PO_4 solution (150 µL). The reaction was stirred at 0°C for 30 min. Step 2: After that, the solution was kept in the freezer at -20°C for 22 h. Step 3: The thawed adenosine's solution was transferred to a 30 mM borate-buffered solution (3.00 mL) containing the amino acid **H-aa-OH** or the amino nitrile **H-Gly-CN** (30.84 µmol, 10.0 equiv.). The pH was adjusted to *ca.* 9.5 using a 4 M aqueous NaOH solution (60 µL). The reaction was stirred at r.t. for 1 h. For **1a-g**, an aliquot (25 µL) of the crude reaction mixture was taken, diluted with water (up to 1.0 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 25 min; Flow rate = 0.15 mL·min^{-1} and Injection volume = 5 µL). For **1h**, an aliquot (10 µL) of the crude reaction mixture was taken, diluted with water (up to 40 µL) an analyzed by HPLC (A: H_2O and B: 20:80 H_2O/MeCN; Gradient: 0-30% of B in 30 min; Flow rate = 1 mL·min^{-1} and Injection volume = 20 µL).

Modified procedure for the prebiotic synthesis of 1c under the optimized reaction conditions:

In the modified procedure, the amino acid was added in step 1 (nitrosation reaction).

Step 1: N^6-methylurea adenosine **2** (1.00 mg, 3.08 µmol, 1.0 equiv.), the amino acid **H-Phe-OH** (30.84 µmol, 10.0 equiv.) and $NaNO_2$ (2.66 mg, 38.54 µmol, 12.5 equiv.) were dissolved in 5% aqueous H_3PO_4 solution (150 µL). The reaction was stirred at 0°C for 30 min. Step 2: After that, the solution was kept in the freezer at -20°C for 22 h. Step 3: The thawed adenosine's solution was transferred to a 30 mM borate-buffered solution (3.00 mL) and the pH was adjusted to *ca.* 9.5 using a 4 M aqueous NaOH solution (60 µL). The reaction was stirred at r.t. for 1 h. After that, an aliquot (10 µL) of the crude reaction mixture was taken, diluted with water (up to 40 µL) an analyzed by HPLC (A: H_2O and B: 20:80 H_2O/MeCN; Gradient: 0-30% of B in 50 min; Flow rate = 1 mL·min^{-1} and Injection volume = 20 µL).

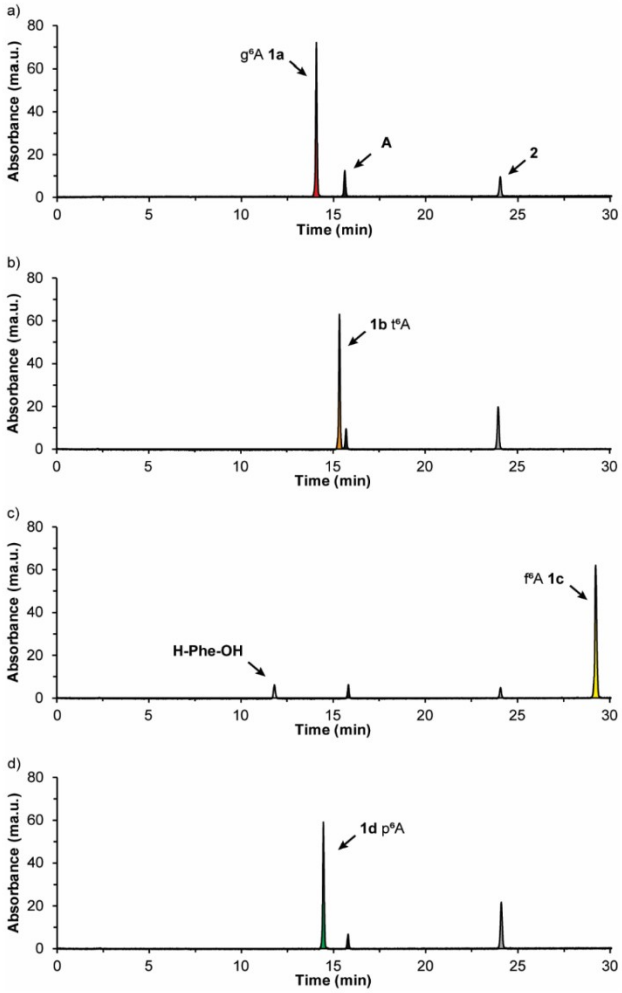


Figure S4. HPLC chromatograms of the crude reaction mixtures for the prebiotic synthesis of: a) g^{6}A **1a**; b) t^{6}A **1b**; c) f^{6}A **1c** and d) p^{6}A **1d**. **A** = Adenosine.

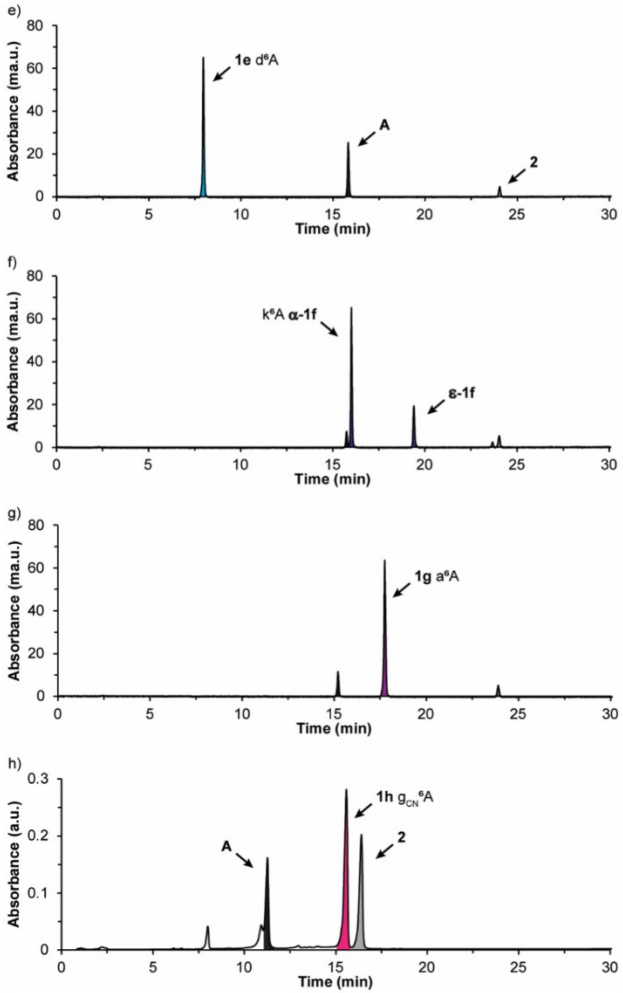


Figure S5. HPLC chromatograms of the crude reaction mixtures for the prebiotic synthesis of: e) d^{6}A **1e**; f) k^{6}A ***α*-1f** (***ε*-1f** is the product of the reaction between the amino group adjacent to the carbon at the *ε*-position of **H-Lys-OH** with the nitrosated derivative of **1**); g) a^{6}A **1g** and h) g_{CN}^{6}A **1h**. **A** = Adenosine.

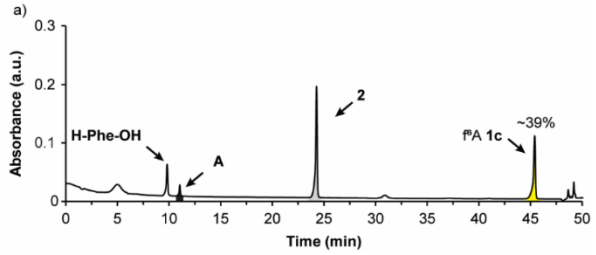


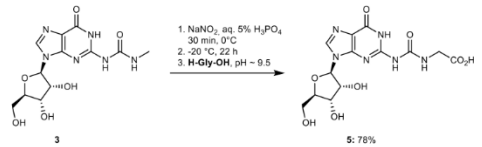
Figure S6. HPLC chromatogram of the crude reaction mixture for the prebiotic synthesis of f^{6}A **1c** following the modified procedure, *i.e.* addition of the amino acid in step 1 (nitrosation reaction).

Table S5. Results obtained for the synthesis of **1a-h** under prebiotically plausible reaction conditions.

Compound	Yield (%)
1a; g^{6}A	84
1b; f^{6}A	61
1c; f^{6}A	78 (39)[a]
1d; p^{6}A	74
1e; d^{6}A	64
***α*-1f;** k^{6}A	70
***ε*-1f**	16
1g; a^{6}A	75
1h; g_{CN}^{6}A	26

[a] Result obtained following the modified procedure, *i.e.* addition of the amino acid in step 1 (nitrosation reaction).

3.2 Prebiotic synthesis of amino acid-modified *N*^{2}-carbamoyl guanosine nucleoside



Scheme S3. Prebiotic synthesis of amino acid-modified *N*^{2}-carbamoyl guanosine nucleoside **5** from *N*^{2}-methylurea guanosine **3** under the optimized reaction conditions.

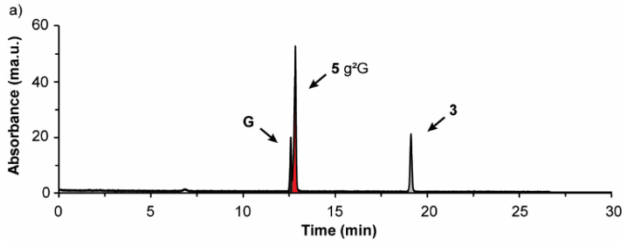
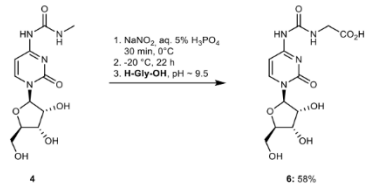


Figure S7. HPLC chromatogram of the crude reaction mixture for the prebiotic synthesis of g^{2}G **5**. **G** = guanosine.

3.3 Prebiotic synthesis of amino acid-modified *N*^{4}-carbamoyl cytidine nucleoside



Scheme S4. Prebiotic synthesis of amino acid-modified *N*^{4}-carbamoyl cytidine nucleoside **6** from *N*^{4}-methylurea cytidine **4** under the optimized reaction conditions.

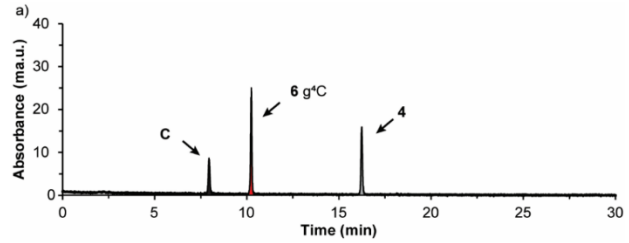
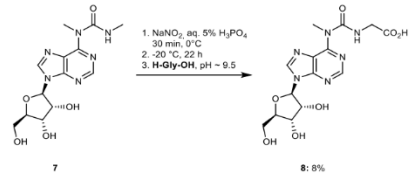


Figure S8. HPLC chromatogram of the crude reaction mixture for the prebiotic synthesis of g^{4}C **6**. **C** = cytidine.

3.4 Prebiotic synthesis of amino acid-modified *N*^{6}-methyl *N*^{6}-carbamoyl adenosine nucleoside



Scheme S5. Prebiotic synthesis of amino acid-modified *N*^{6}-methyl *N*^{6}-carbamoyl adenosine nucleoside **8** from *N*^{6}-methyl *N*^{6}-methylurea adenosine **7** under the optimized reaction conditions.

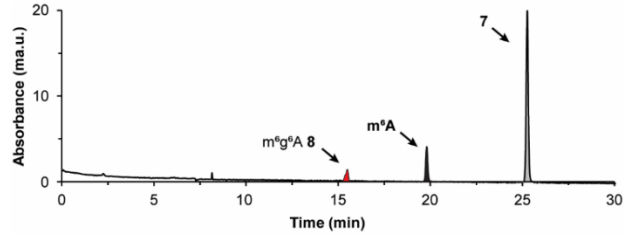
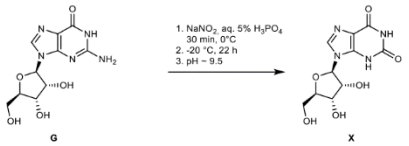


Figure S9. HPLC chromatogram of the crude reaction mixture for the prebiotic synthesis of m^{6}g^{6}A **8**. **m^{6}A** = *N*^{6}-methyl adenosine.

3.5 Control experiment of guanosine under the nitrosation conditions



Scheme S6. Control experiment of guanosine **G** under the nitrosation conditions. **X** = xanthosine.

The control experiment indicated that **G** gives **X** under the nitrosation conditions used in Section 3.1.

4. Calibration curves of amino acid-modified carbamoyl nucleosides

Amino acid-modified carbamoyl nucleosides **1a-h**, **5** and **6** were used for the development of calibration curves. Separate stock solutions of the modified nucleosides were prepared in water (100 µM). Dilute standard solutions of the modified nucleosides (1; 2; 4; 5; 6; 8 µM) were prepared in a final volume of 1 mL. The standard solutions were injected in an analytical UHPLC equipped with a C18 column (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 25 min; Flow rate = 0.15 mL·min^{-1} and Injection volume = 20 µL). For **1h**, a stock solution was prepared in water (1 mM). Dilute standard solutions of **1h** (50; 100; 200; 300; 400; 500 µM) were prepared in a final volume of 20 µL. The standard solutions were injected in an analytical HPLC equipped with a C18 column (A: H_2O and B: 20:80 H_2O/MeCN; Gradient: 0-30% of B in 30 min; Flow rate = 1 mL·min^{-1} and Injection volume = 20 µL). The absorbance was monitored at 260 nm and the areas of the chromatographic peaks were determined by integration. The plot of the chromatographic area (a.u.) versus the amount of each nucleoside followed a linear relationship.

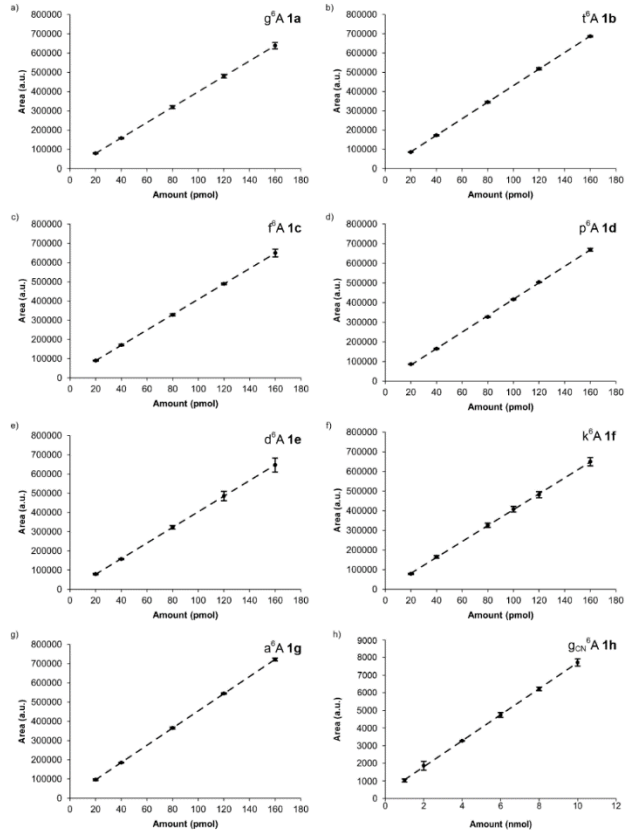


Figure S10. Chromatographic area (a.u.) *vs.* amount (pmol or nmol) of: a) g^6A **1a**; b) t^6A **1b**; c) f^6A **1c**; d) p^6A **1d**; e) d^6A **1e**; f) k^6A **1f**; g) a^6A **1g** and h) $g_{CN}{}^6A$ **1h**. Lines show the fit of the data to a linear regression equation. Error bars are standard deviations from two independent experiments.

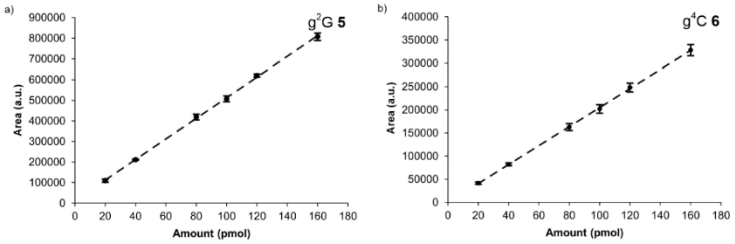


Figure S11. Chromatographic area (a.u.) *vs.* amount (pmol) of: a) g^2G **5** and b) g^4C **6**. Lines show the fit of the data to a linear regression equation. Error bars are standard deviations from two independent experiments.

Table S6. Calibration curves (y = mx + n) obtained by analysis of the chromatographic peaks of **1a-h**, **5** and **6**.

Compound	Slope, m ($pmol^{-1}$)	Intercept, n	r^2
g^6A **1a**	3999.8	-1195.1	0.99
t^6A **1b**	4300.7	-137.4	0.99
f^6A **1c**	3990.0	10177.5	0.99
p^6A **1d**	4183.8	-1542.8	0.99
d^6A **1e**	4062.7	-3010.2	0.99
k^6A ***α*-1f**	4046.2	552.5	0.99
a^6A **1g**	4472.7	6578.1	0.99
$g_{CN}{}^6A$ **1h**[a]	738.7	318.5	0.99
g^2G **5**	5000.9	12316.0	0.99
g^4C **6**	2050.4	-321.0	0.99

[a] Slope (m) in $nmol^{-1}$.

5. General information and instruments for oligonucleotides

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 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 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 nucleosides, the solid support beads were suspended in a 9:1 THF/DBU solution mixture (1 mL) and incubated at r.t. for 2 h.[3] 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-methylaminomethyl uridine nucleosides, 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.[4] 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 (buffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O and buffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-25% of B in 45 min; Flow rate = 5 $mL{\cdot}min^{-1}$). 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% of B in 45 min; Flow rate = 1 $mL{\cdot}min^{-1}$). 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 ON 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.

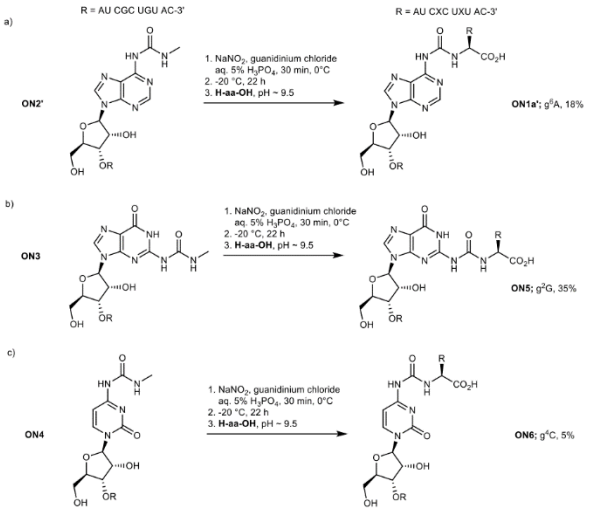
The synthesized ON (2-3 µL) was desalted on a 0.025 µm VSWP filter (Millipore), co-crystallized in a 3-hydroxypicolinic acid matrix (HPA, 1 µL) and analyzed by MALDI-TOF mass spectrometry (negative mode).

Enzymatic digestion of ONs into nucleosides

The ON (~100 pmol) was diluted with the nucleoside digestion mix reaction buffer (10X, 2 µL) and water (up to 19 µL). The nucleoside digestion mix enzyme (1 µL) was added to the ON's solution. The reaction was incubated at 37ºC for 2 h. Finally, the crude reaction mixture was 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 25 min; Flow rate = 0.15 $mL{\cdot}min^{-1}$ and Injection volume = 20 µL).

6. Prebiotic synthesis of amino acid-modified carbamoyl oligonucleotides

6.1 Prebiotic synthesis of amino acid-modified carbamoyl oligonucleotides containing the four canonical bases



Scheme S7. Prebiotic synthesis of amino acid-modified carbamoyl oligonucleotides from those containing: a) N^6-methylurea adenosine **ON2'**; b) N^2-methylurea guanosine **ON3** and c) N^4-methylurea cytidine **ON4**.

General procedure for the nitrosation reaction with oligonucleotides

Step 1: The oligonucleotide **ON** (3 nmol, 1.0 equiv.) and $NaNO_2$ (1.5 µmol, 500 equiv.) were dissolved in 5% aqueous H_3PO_4 solution (60 µL) containing 100 mM guanidinium chloride (GdmCl). The reaction was stirred at 0°C for 30 min. Step 2: After that, the solution was kept in the freezer at -20°C for 22 h. Step 3: **H-aa-OH** (3 µmol, 1000 equiv.) in 30 mM borate-buffered solution (140 µL) was added to the thawed oligonucleotide's solution. The pH was adjusted to *ca.* 9.5 using a 4 M aqueous NaOH solution (*ca.* 19 µL). The reaction was stirred at r.t. for 1 h. Finally, the reaction was quenched with 1 M aqueous HCl solution (5 µL). An aliquot (75 µL) of the crude reaction mixture was taken, diluted with water (up to 100 µL) and analyzed by HPLC (buffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O and buffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-30% of B in 45 min; Flow rate = 1 mL·min^{-1}; Injection volume = 100 µL).

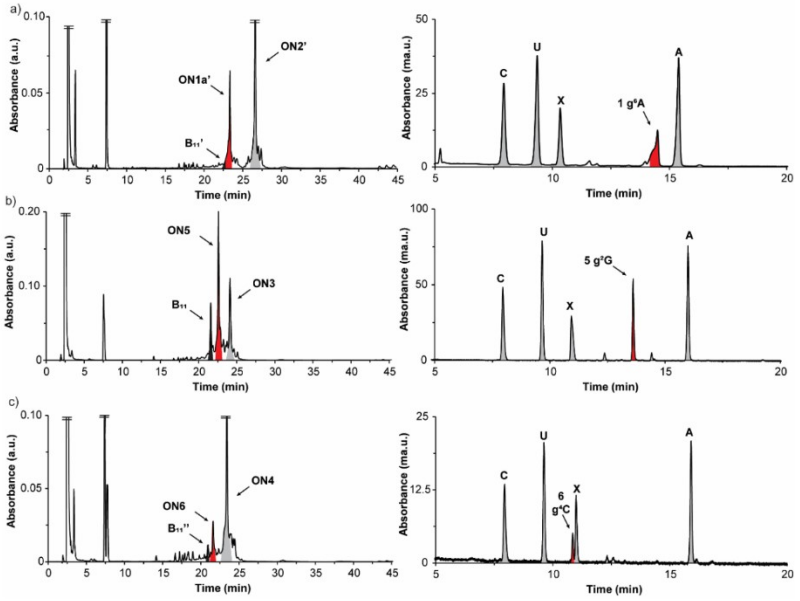
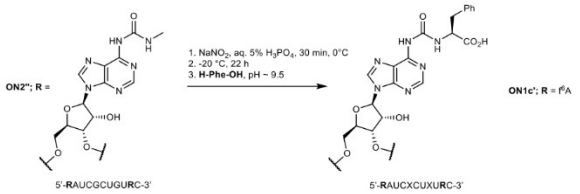


Figure S12. HPLC chromatograms of: left) the crude reaction mixtures for the prebiotic synthesis of a) **ON1a'**; b) **ON5** and c) **ON6**, and right) the enzymatic digestions of the corresponding products. **B_{11}**, **B_{11}'** and **B_{11}''** are the oligonucleotides that do not contain the *N*-methylcarbamoyl substituent at the terminal nucleotide and have xanthosine instead of guanosine.

Table S7. HPLC retention times (0-30% of B in 45 min) and reaction yields of **ON1a'**, **ON5** and **ON6**.

Strand	t_R (min)	Yield%
ON1a'; g^6A	23.4	18
ON5; g^2G	22.6	35
ON6; g^4C	21.5	5

6.2 Prebiotic synthesis of an oligonucleotide containing two amino acid-modified carbamoyl nucleotides at the terminal and internal positions



Scheme S8. Prebiotic synthesis of **ON1c'** from **ON2''** containing two N^6-methylurea adenosine nucleotides. In this case, guanidinium chloride (GdmCl) was not added.

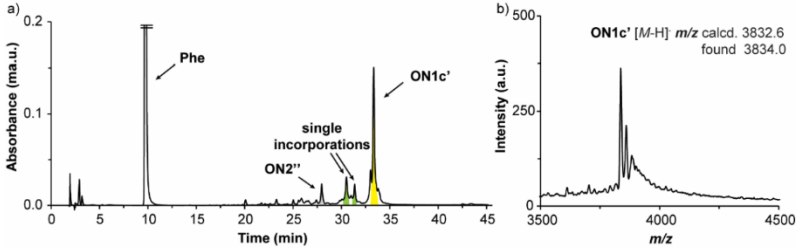
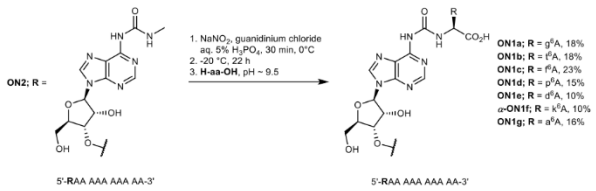


Figure S13. a) HPLC chromatogram of the crude reaction mixture for the prebiotic synthesis of **1c'**. b) MALDI-TOF mass spectrum of the isolated product.

6.3 Prebiotic synthesis of a series of amino acid-modified N^6-carbamoyl adenosine oligonucleotides



Scheme S9. Prebiotic synthesis of amino acid-modified N^6-carbamoyl adenosine oligonucleotides **ON1a**-**g** from N^6-methylurea adenosine oligonucleotide **ON2** (optimized reaction conditions are shown).

Effect of the addition of salts in the prebiotic synthesis of ON1g:

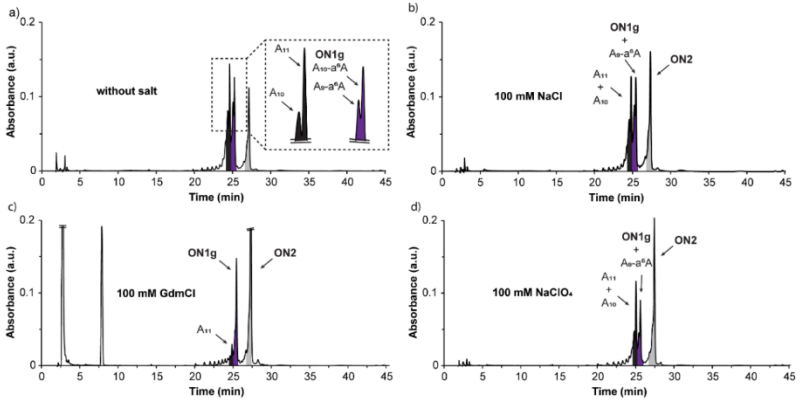


Figure S14. HPLC chromatograms of the crude reaction mixtures for the prebiotic synthesis of **ON1g**: a) without salt and with b) 100 mM NaCl; c) 100 mM guanidinium chloride (GdmCl) and d) 100 mM $NaClO_4$.

Table S8. Results obtained for the addition of salts in the prebiotic synthesis of **ON1g**.

Salt	A_{11} (%)	ON1g (%)	ON2 (%)	Degradation (%)
-	18	15	15	~52
100 mM NaCl	19	15	28	~38
100 mM GdmCl	4	15	53	~28
100 mM $NaClO_4$	15	11	25	~49

Prebiotic synthesis of ON1a-g under the nitrosation conditions using guanidinium chloride:

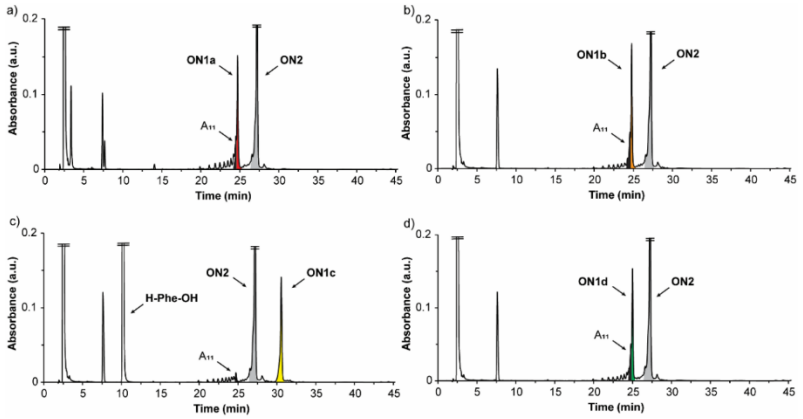


Figure S15. HPLC chromatograms of the crude reaction mixtures for the prebiotic synthesis of: a) **ON1a**; b) **ON1b**; c) **ON1c** and d) **ON1d**.

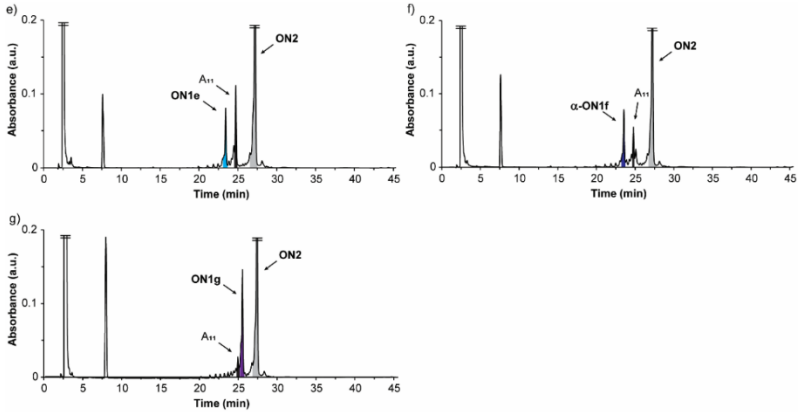
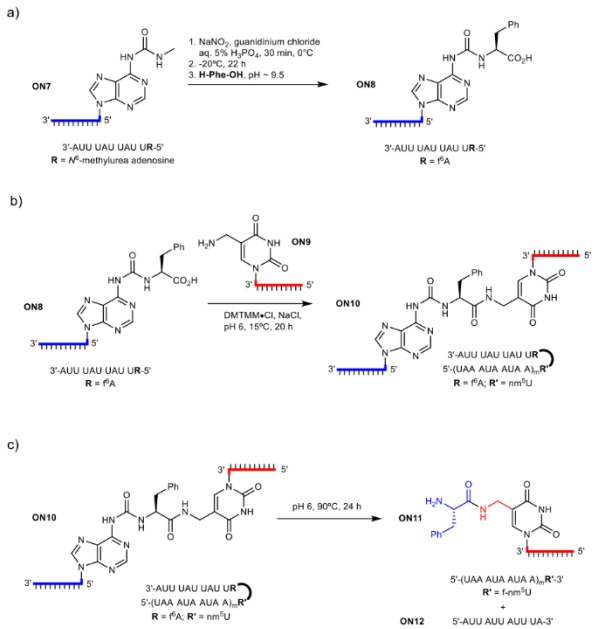


Figure S16. HPLC chromatograms of the crude reaction mixtures for the prebiotic synthesis of: e) **ON1e**; f) ***α*-ON1f** and g) **ON1g**.

Table S9. HPLC retention times (0-30% of B in 45 min), MALDI-TOF mass spectrometric analysis (negative mode) and reaction yields of **ON1a**-**g**.

Strand	t_R (min)	*m/z* calcd. for $[M\text{-H}]^-$	found	Yield%
ON1a; **R** = g^6A	24.7	3657.6	3660.3	18
ON1b; **R** = t^6A	24.8	3701.7	3702.7	18
ON1c; **R** = f^6A	30.6	3747.7	3747.0	23
ON1d; **R** = p^6A	25.0	3697.7	3699.7	15
ON1e; **R** = d^6A	23.4	3715.6	3717.2	10
***α*-ON1f**; **R** = k^6A	23.6	3728.7	3730.3	10
ON1g; **R** = a^6A	25.4	3671.6	3673.6	16

7. Consecutive reactions using N^6-methylurea adenosine oligonucleotide



Scheme S10. Consecutive reactions: a) prebiotic synthesis of amino acid-modified carbamoyl oligonucleotide **ON8**; b) peptide coupling reaction between **ON8** and **ON9** and c) urea cleavage reaction of **ON10**.

Procedures for the stepwise reactions:

a) Prebiotic synthesis of amino acid-modified carbamoyl oligonucleotide ON8:

Step 1: The oligonucleotide **ON7** (20 nmol, 1.0 equiv.) and $NaNO_2$ (10 µmol, 500 equiv.) were dissolved in 5% aqueous H_3PO_4 solution (200 µL) containing 100 mM guanidinium chloride (GdmCl). The reaction was stirred at 0°C for 30 min. Step 2: After that, the solution was kept in the freezer at -20°C for 22 h. Step 3: **H-Phe-OH** (20 µmol, 1000 equiv.) in 30 mM borate-buffered solution (150 µL) was added to the thawed oligonucleotide's solution. The pH was adjusted to *ca.* 9.5 using a 4 M aqueous NaOH solution (*ca.* 50 µL). The reaction was stirred at r.t. for 1 h. Finally, the reaction was quenched with 1 M aqueous HCl solution (50 µL). An aliquot (22.5 µL) of the crude reaction mixture was taken, diluted with water (up to 100 µL) and analyzed by HPLC (buffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O and buffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-30% of B in 45 min; Flow rate = 1 mL·min^{-1}; Injection volume = 100 µL) and MALDI-TOF mass spectrometry. The remaining crude was purified by semi-preparative HPLC. The purified ON was lyophilized, desalted, and redissolved in water. Subsequently, the concentration of **ON8** was determined.

b) Peptide coupling reaction between ON8 and ON9:

An equimolar solution mixture of **ON8** (50 µM) and **ON9** (50 µM) containing MES buffer pH 6 (100 mM), NaCl (1 M) and DMTMM•Cl (50 mM) was incubated at 15°C for 20 h. After that, an aliquot (20 µL) of the crude reaction mixture was taken and analyzed by HPLC (buffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O and buffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-30% of B in 45 min; Flow rate = 1 mL·min^{-1}; Injection volume = 20 µL) and MALDI-TOF mass spectrometry. The remaining crude was purified by semi-preparative HPLC. The purified ON was lyophilized, desalted, and redissolved in water. Subsequently, the concentration of **ON10** was determined.

c) Urea cleavage reaction of ON10:

The oligonucleotide **ON10** (20 µM) was diluted with MES buffer pH 6 (100 mM) containing NaCl (100 mM). The reaction mixture was heated at 90°C for 24 h. After that, an aliquot (100 µL) of the crude reaction mixture was taken and analyzed by HPLC (buffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O and buffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-30% of B in 45 min; Flow rate = 1 mL·min^{-1}; Injection volume = 20 µL) and MALDI-TOF mass spectrometry.

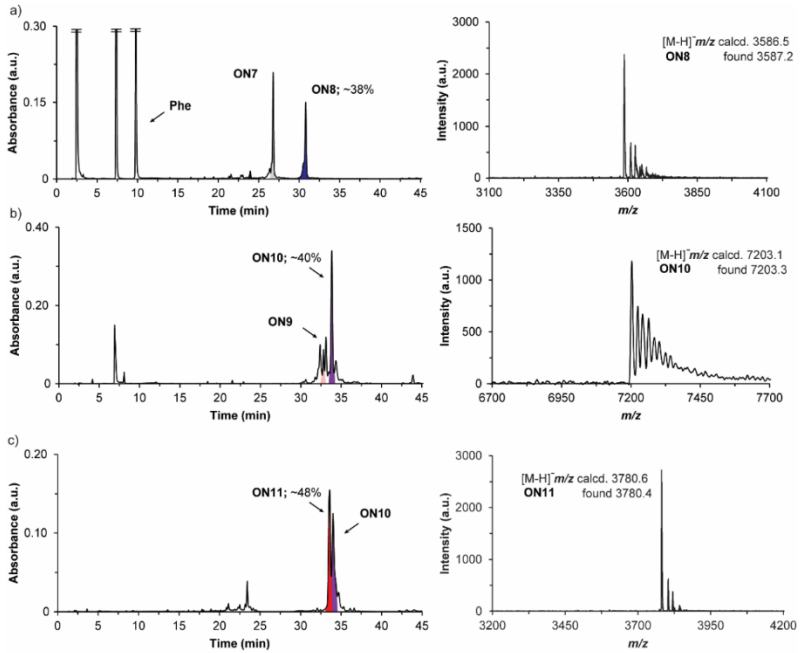


Figure S17. left) HPLC chromatograms of the crude reaction mixtures and right) MALDI-TOF mass spectra of the isolated oligonucleotides in the stepwise reactions: a) prebiotic synthesis of amino acid-modified carbamoyl oligonucleotide **ON8**; b) peptide coupling reaction between **ON8** and **ON9** and c) urea cleavage reaction of **ON10**. Yields of the oligonucleotide products were estimated using the areas of the chromatographic peaks in relation to that determined for reference compounds.

Procedure for the consecutive reactions (a → c):

The reactions a, b and c were performed using the conditions indicated above, but without performing a chromatographic purification after each reaction step. Under these conditions, only a filtration, using an Amicon® ultra centrifugal filter (0.5 mL, 3 kDa), was required after reactions a and b to remove the excess of salts and condensation reagents. For this experiment, 20 nmol of **ON7** were used as starting material and 10 nmol of **ON9** were added for reaction b.

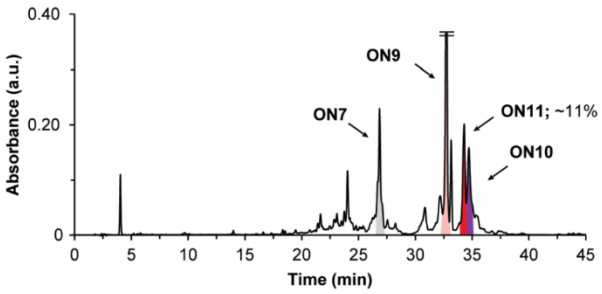


Figure S18: HPLC chromatogram of the crude reaction mixture after the consecutive reactions (a → c). Yield of the oligonucleotide product was estimated using the area of the chromatographic peak in relation to that determined for a reference compound.

8. Calibration curve of N^6-methylurea adenosine oligonucleotide

N^6-methylurea adenosine oligonucleotide **ON2** was used for the development of a HPLC calibration curve. A stock solution of **ON2** was 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 **ON2** were prepared in a final volume of 20 µL. The standard solutions were injected in an analytical HPLC equipped with a C18 column (buffer A: 0.1 M AcOH/Et_3N pH 7 in H_2O and buffer B: 0.1 M AcOH/Et_3N pH 7 in 20:80 H_2O/MeCN; Gradient: 0-30% 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 HPLC chromatograms. The plot of the chromatographic area (a.u.) versus the amount (nmol) of the oligonucleotide followed a linear relationship.

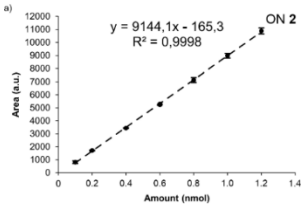


Figure S19. Chromatographic area (a.u.) *vs.* amount (nmol) of **ON2**. Line shows the fit of the data to a linear regression equation. Error bars are standard deviations from three independent experiments.

Table S10. Calibration curve (y = mx + n) obtained by analysis of the chromatographic peaks of **ON2**.

Strand	Slope, m (nmol^{-1})	Intercept, n	r^2
ON2	9144.1	-165.3	0.99

9. Melting curve of double strand

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 1ºC·min^{-1}. The obtained UV spectroscopic data were fit to a two-state melting model, *i.e.* double strand – random coil equilibrium, using a mono-sigmoidal Boltzmann function.[5] The fit of the data returned the melting temperature.

For the experiments, we prepared aqueous solutions containing equimolar amounts of the oligonucleotides (2 µM), MES buffer pH 6 (100 mM) and NaCl (100 mM or 1 M). The oligonucleotides were annealed by heating to 95ºC for 4 min and, subsequently, by cooling down slowly to 5ºC before the variable-temperature UV spectroscopic experiments.

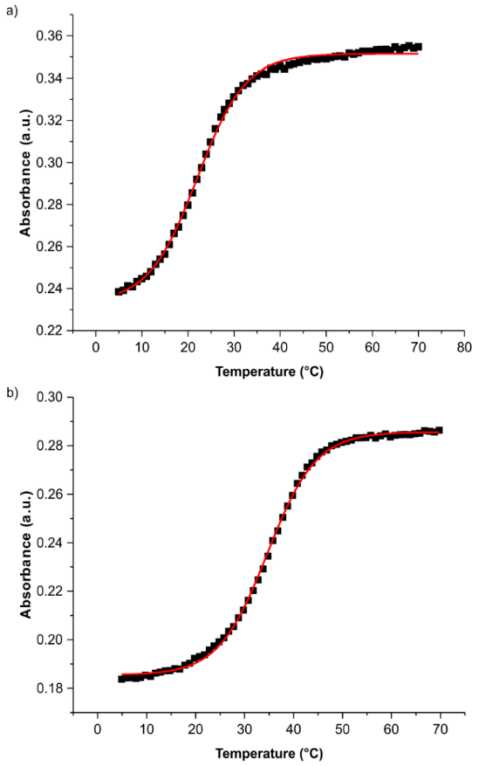
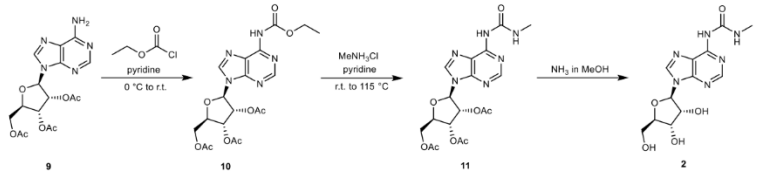


Figure S20. Melting curves of the RNA donor strand **ON8** with the RNA acceptor strand **ON9** containing: a) 100 mM NaCl and b) 1 M NaCl. The fit of the data to a two-state melting model using a mono-sigmoidal Boltzmann function returned the melting temperatures for the double strand: a) T_m = 22.0ºC and b) T_m = 34.6ºC.

10. Synthesis of methylurea and amino acid-modified carbamoyl nucleosides used as reference

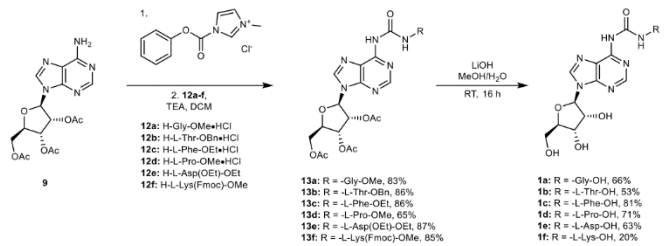
10.1 Synthesis of N^6-methylurea adenosine



Scheme S11. Synthesis of N^6-methylurea adenosine **2**.

Compounds **2**, **10** and **11** were synthesized following procedures previously reported in the literature.[6,7]

10.2 Synthesis of amino acid-modified N^6-carbamoyl adenosine nucleosides



Scheme S12. Synthesis of amino acid-modified N^6-carbamoyl adenosine nucleosides **1a-f** used as reference.

General procedure for the synthesis of 13a-f

Step 1: Acetyl-protected adenosine **9** (1.0 equiv.) and 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (2.0 equiv.) were dissolved in dry DCM. The reaction was stirred at r.t. for 14 h. Step 2: A suspension, in dry DCM, of the protected amino acid **12a-f** (2.0 equiv.), containing TEA (2 equiv.) in the case of **12a-d** (hydrochloride salts), was transferred to the adenosine's mixture. The reaction was stirred at r.t. for 16 h. After that, the crude was washed with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted three times with DCM. The combined organic layers were dried (Na_2SO_4 or $MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (*i*-Hex/EtOAc or DCM/IPA) affording **13a-f** as a white solid.

13a: Yield = 83%. R_f = 0.24 (EtOAc). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3240; 1741; 1698; 1611; 1589; 1538; 1532; 1470; 1366; 1206; 1089; 1043; 1019; 900; 798. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 9.95 (t, J = 5.5 Hz, 1H); 8.59 (s, 1H); 8.34 (s, 1H); 8.22 (s, 1H); 6.21 (d, J = 5.5 Hz, 1H); 5.96 (t, J = 5.5 Hz, 1H); 5.66 (dd, J = 5.5, 4.2 Hz, 1H); 4.49-4.35 (m, 3H); 4.24 (d, J = 5.5 Hz, 2H); 3.80 (s, 3H); 2.16 (s, 3H); 2.13 (s, 3H); 2.08 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 170.6; 170.5; 169.7; 169.5; 154.0; 151.7; 150.4; 150.4; 141.5; 121.1; 86.6; 80.6; 73.2; 70.8; 63.2; 52.5; 42.2; 20.9; 20.7; 20.6. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{20}H_{25}N_6O_{10}$ 509.1627; Found 509.1624.

13b: Yield = 86%. R_f = 0.50 (EtOAc). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3237; 1742; 1694; 1614; 1589; 1551; 1537; 1468; 1366; 1212; 1089; 1044; 1018; 923; 903; 798; 752; 735; 687. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.23 (d, J = 8.5 Hz, 1H); 8.70 (s, 1H); 8.54 (s, 1H); 8.28 (s, 1H); 7.40-7.33 (m, 5H); 6.22 (d, J = 5.5 Hz, 1H); 5.95 (dd, J = 5.5, 5.5 Hz, 1H); 5.64 (dd, J = 5.5, 4.2 Hz, 1H); 5.24 (d, J = 1.9 Hz, 2H); 4.74 (dd, J = 8.5, 2.9 Hz, 1H); 4.50-4.32 (m, 4H); 2.16 (s, 3H); 2.12 (s, 3H); 2.07 (s, 3H); 1.31 (d, J = 6.4 Hz, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 171.0; 170.6; 169.8; 169.5; 154.4; 151.8; 150.5; 150.4; 141.6; 135.5; 128.8; 128.6; 128.3; 121.0; 86.4; 80.6; 73.2; 70.8; 68.5; 67.4; 63.3; 59.0; 20.9; 20.7; 20.5; 20.1. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{28}H_{33}N_6O_{11}$ 629.2202; Found 629.2200.

13c: Yield = 86%. R_f = 0.45 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3231; 1742; 1698; 1614; 1588; 1524; 1468; 1368; 1214; 1131; 1084; 1044; 921; 903; 798; 740; 700. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 9.99 (d, J = 7.7 Hz, 1H); 8.52 (s, 1H); 8.42 (s, 1H); 8.26 (s, 1H); 7.34-7.20 (m, 5H); 6.19 (d, J = 5.3 Hz, 1H); 5.98 (dd, J = 5.3, 5.3 Hz, 1H); 5.66 (dd, J = 5.3, 4.4 Hz, 1H); 4.91 (dt, J = 7.7, 6.1 Hz, 1H); 4.48-4.34 (m, 3H); 4.20 (q, J = 7.1 Hz, 2H); 3.24 (d, J = 6.1 Hz, 2H); 2.15 (s, 3H); 2.12 (s, 3H); 2.07 (s, 3H); 1.25 (t, J = 7.1 Hz, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 171.7; 170.5; 169.7; 169.5; 153.5; 151.4; 150.4; 150.3; 141.8; 136.4; 129.7; 128.6; 127.2; 121.0; 86.6; 80.5; 73.1; 70.7; 63.2; 61.5; 54.8; 38.3; 20.9; 20.7; 20.5; 14.3. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{28}H_{33}N_6O_{10}$ 613.2253; Found 613.2253.

13d: Yield = 65%. R_f = 0.13 (96:4 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2956; 1740; 1690; 1647; 1607; 1403; 1367; 1213; 1043; 898; 727; 643. For major rotamer: ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 9.77 (s, 1H); 9.58 (s, 2H); 6.30 (d, J = 5.6 Hz, 1H); 6.07 (dd, J = 5.6, 5.6 Hz, 1H); 5.64 (dd, J = 5.6, 5.6 Hz, 1H); 4.43-4.38 (m, 3H); 4.27-4.24 (m, 1H); 3.72 (br s, 1H); 3.62 (s, 4H); 2.28-2.21 (m, 1H); 2.13 (s, 3H); 2.04 (s, 3H); 2.01 (s, 3H); 1.92 (br s, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 172.7; 170.1; 169.5; 169.4; 152.2; 151.9; 151.8; 151.1; 142.5; 123.8; 85.6; 79.6; 71.8; 70.1; 62.8; 58.9; 51.9; 46.9; 29.2; 24.5; 20.6; 20.4; 20.2. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{23}H_{29}N_6O_{10}$ 549.1939; Found 549.1942.

13e: Yield = 87%. R_f = 0.50 (97:3 DCM/IPA). ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.40 (d, J = 7.8 Hz, 1H); 8.95 (s, 1H); 8.58 (s, 1H); 8.40 (s, 1H); 6.22 (d, J = 5.5 Hz, 1H); 5.97 (dd, J = 5.5, 5.5 Hz, 1H); 5.65 (dd, J = 5.5, 4.2 Hz, 1H); 4.93 (dt, J = 7.8, 4.8 Hz, 1H); 4.44 (dq, J = 8.0, 3.5 Hz, 2H); 4.36 (dd, J = 12.8, 5.4 Hz, 1H);

4.25 (q, J = 7.1 Hz, 2H); 4.19-4.12 (m, 2H); 3.18-2.94 (m, 2H); 2.14 (s, 3H); 2.10 (s, 3H); 2.06 (s, 3H); 1.31-1.22 (m, 6H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 170.9; 170.7; 170.5; 169.7; 169.4; 153.8; 151.6; 150.5; 150.4; 142.1; 121.0; 86.5; 80.6; 73.1; 70.8; 63.2; 61.9; 61.1; 50.0; 37.1; 20.9; 20.7; 20.5; 14.3; 14.2. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{25}H_{33}N_6O_{12}$ 609.2151; Found 609.2143.

13f: Yield = 85%. R_f = 0.37 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2949; 1743; 1697; 1612; 1532; 1365; 1214; 1043; 741. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.06 (d, J = 7.8 Hz, 1H); 8.64 (s, 1H); 8.60 (s, 1H); 8.30 (s, 1H); 7.74 (d, J = 7.6 Hz, 2H); 7.57 (d, J = 7.4 Hz, 2H); 7.37 (t, J = 7.4 Hz, 2H); 7.28 (td, J = 7.5, 0.8 Hz, 2H); 6.22 (d, J = 5.5 Hz, 1H); 5.96 (t, J = 5.5 Hz, 1H); 5.64 (t, J = 4.6 Hz, 1H); 4.96 (t, J = 6.0 Hz, 1H); 4.70 (td, J = 7.7, 5.3 Hz, 1H); 4.47-4.36 (m, 5H); 4.20 (t, J = 7.0 Hz, 1H); 3.77 (s, 3H); 3.21 (hept, J = 6.8 Hz, 2H); 2.15 (s, 3H); 2.11 (s, 3H); 2.06 (s, 3H); 1.21 (s, 2H); 1.20 (s, 2H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 173.0; 170.5; 169.7; 169.5; 156.5; 153.7; 151.6; 150.5; 150.4; 144.1; 141.7; 141.4; 127.7; 127.1; 125.2; 121.1; 120.0; 86.5; 80.6; 73.1; 70.8; 66.6; 63.2; 53.2; 52.6; 47.4; 40.9; 32.4; 29.5; 22.8; 20.9; 20.7; 20.5. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{39}H_{44}N_7O_{12}$ 802.3042; Found 802.3046.

General procedure for the synthesis of 1a-f

Acetyl-protected adenosine derivative **13a-f** (1.0 equiv.) was dissolved in 3:1 MeOH/H_2O and the solution was cooled to 0°C. LiOH•H_2O (10.0 equiv.) was added to the adenosine's solution and the reaction was stirred at r.t. for 2 h. After that, the solution was neutralized with 1 M aqueous HCl solution. The organic solvent was removed under reduced pressure and the remaining aqueous solution mixture was lyophilized. The crude was purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **1a-f** as a white solid. Note that for **1f**, the Fmoc protecting group was removed with 20% piperidine in DMF before the hydrolysis reaction.

1a: Yield = 66%. t_R = 14.0 min (0-30% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3499; 3377; 1720; 1656; 1610; 1557; 1254; 1227; 1100; 1056; 867; 803; 796; 745; 716. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.70 (br s, 1H); 9.90 (s, 1H); 9.65 (t, J = 5.6 Hz, 1H); 8.67 (s, 1H); 8.56 (s, 1H); 5.99 (d, J = 5.6 Hz, 1H); 5.55 (br s, 1H); 5.25 (br s, 1H); 5.15 (br s, 1H); 4.59 (t, J = 5.3 Hz, 1H); 4.17 (dd, J = 5.0, 3.6 Hz, 1H); 4.01-3.95 (m, 3H), 3.71-3.55 (m, 2H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 171.5; 153.7; 150.8, 150.4; 150.2; 142.3; 120.4; 87.7; 85.7; 73.8; 70.3; 61.3; 41.7. HRMS (ESI) m/z: [M-H]$^-$ Calcd. for $C_{13}H_{15}N_6O_7$ 367.1008; Found 367.1009.

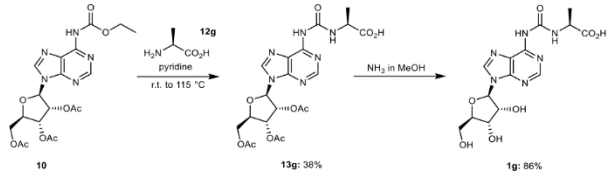
1b: Yield = 53%. t_R = 22.2 min (0-30% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3231; 1684; 1617; 1591; 1553; 1469; 1403; 1364; 1294; 1256; 1234; 1218; 1086; 1055; 896; 868; 848; 797; 753; 718; 695. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.62 (br s, 1H); 9.83 (s, 1H); 9.73 (d, J = 8.3 Hz, 1H); 8.68 (s, 1H); 8.55 (s, 1H); 5.99 (d, J = 5.6 Hz, 1H); 5.55 (d, J = 5.9 Hz, 1H); 5.29-5.17 (m, 1H); 5.14 (t, J = 5.6 Hz, 1H); 4.60 (q, J = 5.0 Hz, 1H); 4.30-4.22 (m, 2H); 4.18 (q, J = 4.1, 3.5 Hz, 1H); 3.97 (q, J = 3.9 Hz, 1H); 3.72-3.54 (m, 2H); 1.14 (d, J = 6.3 Hz, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 172.5; 153.8; 150.9; 150.4; 150.3; 142.3; 120.5; 87.7; 85.7; 73.8; 70.3; 66.2; 61.3; 58.7; 20.9. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{15}H_{21}N_6O_8$ 413.1415; Found 413.1415.

1c: Yield = 81%. t_R = 23.0 min (0-65% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3228; 1714; 1652; 1613; 1590; 1546; 1339; 1246; 1229; 1125; 1082; 986; 896; 868; 822; 797; 741; 699. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.92 (br s, 1H); 9.90 (s, 1H); 9.76 (d, J = 7.3 Hz, 1H); 8.66 (s, 1H); 8.46 (s, 1H); 7.36-7.20 (m, 5H); 5.98 (d, J = 5.6 Hz, 1H); 5.54 (s, 1H); 5.25 (s, 1H); 5.14 (s, 1H); 4.67-4.50 (m, 2H); 4.21-4.12 (m, 1H); 3.97 (q, J = 3.9 Hz, 1H); 3.74-3.51 (m, 2H); 3.21-3.04 (m, 2H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 172.9; 153.2; 150.6; 150.4; 150.1; 142.3; 137.0; 129.5; 128.4; 126.8; 120.3; 87.7; 85.7; 73.8; 70.3; 61.3; 54.4; 37.1. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{20}H_{23}N_6O_7$ 459.1623; Found 459.1624.

1d: Yield = 71%. t_R = 21.8 min (0-20% of B in 30 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3272; 2935; 1654; 1609; 1459; 1401; 1354; 1220; 1078; 1054; 892; 863; 640. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.55 (s, 1H); 9.64 (s, 1H); 8.58 (s, 2H); 5.97 (d, J = 5.3 Hz, 1H); 5.54 (d, J = 6.1 Hz, 1H); 5.24 (d, J = 6.1 Hz, 1H); 5.18 (s, 1H); 4.61 (s, 1H); 4.29 (s, 1H); 4.17-4.15 (m, 1H); 3.97-3.95 (m, 1H); 3.70-3.66 (m, 2H); 3.58-3.54 (m, 2H); 2.25-2.18 (s, 1H); 1.90 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 173.8; 152.3; 151.6; 151.3; 142.0; 123.5; 87.6; 85.7; 73.6; 70.4; 61.4; 59.0; 46.9; 29.3; 24.4 (one carbon signal appeared too broad for an unequivocal assignment). HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{16}H_{21}N_6O_7$ 409.1466; Found 409.1465.

1e: Yield = 63%. t_R = 24.7 min (0-20% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3207; 2937; 1682; 1614; 1592; 1538; 1471; 1399; 1334; 1293; 1219; 1122; 1080; 1053; 985; 895; 866; 797; 704. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.79 (br s, 2H); 9.96 (d, J = 7.9 Hz, 1H); 9.92 (s, 1H); 8.67 (s, 1H); 8.51 (s, 1H); 5.99 (d, J = 5.6 Hz, 1H); 5.55 (d, J = 5.9 Hz, 1H); 5.24 (d, J = 5.0 Hz, 1H); 5.14 (s, 1H); 4.66-4.56 (m, 2H); 4.22-4.11 (m, 1H); 3.97 (q, J = 3.9 Hz, 1H); 3.75-3.53 (m, 2H); 2.93-2.73 (m, 2H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 172.4; 172.1; 153.2; 150.7; 150.5; 150.2; 142.3; 120.4; 87.7; 85.7; 73.8; 70.3, 61.3; 49.4; 36.6. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{15}H_{19}N_6O_9$ 427.1208; Found 427.1207.

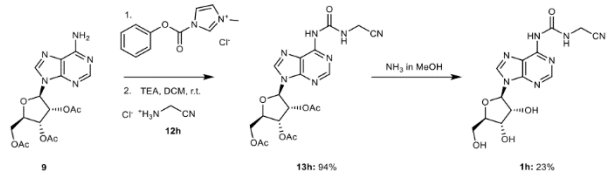
1f: Yield = 20%. t_R = 18.3 min (0-20% of B in 30 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3123; 2934; 1683; 1588; 1526; 1469; 1398; 1252; 1053; 796; 641. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 9.62 (d, J = 6.1 Hz, 1H); 8.65 (s, 1H); 8.51 (s, 1H); 5.97 (d, J = 5.5 Hz, 1H); 4.59 (t, J = 5.4 Hz, 1H); 4.17 (t, J = 4.2 Hz, 1H); 4.00-3.95 (m, 2H); 3.68 (dd, J = 12.0, 4.0 Hz, 1H); 3.56 (dd, J = 12.0, 4.0 Hz, 1H); 2.74 (t, J = 6.7 Hz, 2H); 1.81-1.67 (m, 2H); 1.56 (p, J = 7.6 Hz, 2H); 1.43-1.26 (m, 2H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 174.0; 152.2; 151.0; 150.4; 150.1; 142.1; 120.3; 87.7; 85.7; 73.8; 70.3; 61.3; 55.1; 38.8; 32.3; 27.2; 21.9. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{17}H_{26}N_7O_7$ 440.1888; Found 440.1888.



Scheme S13. Synthesis of alanine-modified N^6-carbamoyl adenosine nucleoside **1g** used as reference.

Acetyl-protected N^6-alanine modified adenosine 13g: Carbamate **10** (501 mg, 1.08 mmol, 1.0 equiv.) was dissolved in dry pyridine. Subsequently, H-L-Ala-OH **12g** (192 mg, 2.15 mmol, 2.0 equiv.) was added to the carbamate's solution at r.t. The reaction was heated under reflux conditions for 16 h. After that, the reaction mixture was cooled to r.t. The suspension was filtered, and the residue was washed with EtOAc. The filtrate was concentrated under reduced pressure and co-evaporated with toluene. The crude was purified by recrystallisation from EtOH affording **13g** as a white solid (210 mg, 0.41 mmol, 38% yield). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3270; 2456; 1744; 1700; 1652; 1607; 1592; 1543; 1485; 1295; 1246; 1236; 1221; 1112; 1102; 1044; 1025; 994; 916; 802; 755. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.83 (br s, 1H); 9.95 (s, 1H); 9.69 (d, J = 7.0 Hz, 1H); 8.65 (s, 1H); 8.60 (s, 1H); 6.30 (d, J = 5.4 Hz, 1H); 6.03 (dd, J = 5.4, 5.4 Hz, 1H); 5.62 (dd, J = 5.4, 4.5 Hz, 1H); 4.46-4.23 (m, 4H); 2.12 (s, 3H); 2.03 (s, 3H); 2.02 (s, 3H); 1.42 (d, J = 7.2 Hz, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 174.2; 170.1; 169.5; 169.3; 152.8; 151.2; 150.4; 150.2; 142.7; 120.5; 85.7; 79.7; 72.0; 70.0; 62.8; 48.6; 20.6; 20.4; 20.2; 18.1. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{20}H_{25}N_6O_{10}$ 509.1627; Found 509.1627.

N^6-alanine modified adenosine 1g: Protected N^6-alanine modified adenosine **13g** (200 mg, 0.39 mmol, 1.0 equiv.) was dissolved in 7 N NH_3 in MeOH. The reaction was stirred at r.t. overnight. After that, the crude was concentrated and purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **1g** as a white solid (130 mg, 0.34 mmol, 86% yield). t_R = 20.7 min (0-40% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3234; 2926; 1684; 1612; 1603; 1540; 1476; 1343; 1313; 1267; 1250; 1226; 1122; 1080; 861; 835; 782; 764; 746; 680. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.81 (br s, 1H); 9.87 (s, 1H); 9.74 (d, J = 7.0 Hz, 1H); 8.67 (s, 1H); 8.58 (s, 1H); 5.99 (d, J = 5.6 Hz, 1H); 5.55 (d, J = 5.6 Hz, 1H); 5.25 (s, 1H); 5.14 (t, J = 5.6 Hz, 1H); 4.59 (q, J = 4.6 Hz, 1H); 4.33 (p, J = 7.1 Hz, 1H); 4.17 (t, J = 4.5 Hz, 1H); 3.97 (q, J = 3.9 Hz, 1H); 3.70-3.58 (m, 2H); 1.42 (d, J = 7.2 Hz, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 174.2; 152.9; 150.9; 150.4; 150.2; 142.2, 120.4; 87.6; 85.7; 73.8; 70.3; 61.3; 48.6; 18.1. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{14}H_{19}N_8O_7$ 383.1310; Found 383.1311.

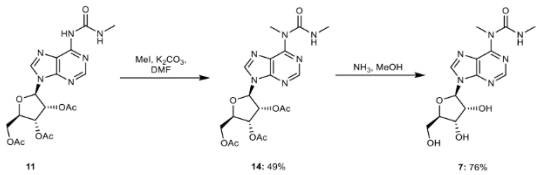


Scheme S14. Synthesis of amino nitrile-modified N^6-carbamoyl adenosine nucleoside **1h** used as reference.

Acetyl-protected N^6-amino nitrile modified adenosine 13h: Step 1: Acetyl-protected adenosine **9** (500 mg, 1.27 mmol, 1.0 equiv.) and 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (607 mg, 2.54 mmol, 2.0 equiv.) were dissolved in dry DCM. The reaction was stirred at r.t. for 14 h. Step 2: A suspension, in dry DCM, of the amino nitrile hydrochloride **12h** (235 mg, 2.54 mmol, 2.0 equiv.), containing TEA (2 equiv.), was transferred to the adenosine's mixture. The reaction was stirred at r.t. for 16 h. After that, the reaction was quenched with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (95:5 DCM/IPA) affording **13h** as a white solid (570 mg, 1.20 mmol, 94% yield). R_f = 0.50 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 1742; 1701; 1612; 1589; 1517; 1470; 1365; 1211; 1043; 904; 798; 642; 579; 507; 407. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.19 (t, J = 5.7 Hz, 1H); 9.17 (s, 1H); 8.58 (s, 1H); 8.48 (s, 1H); 6.23 (d, J = 5.2 Hz, 1H); 5.99 (dd, J = 5.2, 5.2 Hz, 1H); 5.74-5.64 (m, 1H); 4.51-4.45 (m, 2H); 4.42-4.35 (m, 3H); 2.16 (s, 3H); 2.12 (s, 3H); 2.09 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 170.6; 169.8; 169.5; 154.1; 151.2; 150.6; 149.9; 142.8; 121.1; 116.4; 86.8; 80.6; 73.2; 70.8; 63.3; 28.4; 21.0; 20.7; 20.6. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{19}H_{22}N_7O_8$ 476.1524; Found 476.1524.

N^6-amino nitrile modified adenosine 1h: Protected N^6-amino nitrile modified adenosine **13h** (100 mg, 0.21 mmol, 1.0 equiv.) was dissolved in 7 N NH_3 in MeOH. The reaction was stirred at r.t. for 3 h. After that, the crude was concentrated and purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN) affording **1h** as a white solid (17 mg, 0.05 mmol, 23% yield). t_R = 14.1 min (0-30% of B in 30 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 1614; 1589; 1469; 1401; 1297; 1253; 1079; 1049; 1016; 984; 900; 865; 796; 644; 578; 511; 457; 444; 433; 414; 407. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 10.31 (s, 1H); 9.81 (t, J = 5.7 Hz, 1H); 8.69 (s, 1H); 8.57 (s, 1H); 5.99 (d, J = 5.7 Hz, 1H); 5.55 (d, J = 6.0 Hz, 1H); 5.26 (d, J = 5.0 Hz, 1H); 5.14 (t, J = 5.6 Hz, 1H); 4.59 (dd, J = 5.6, 5.6 Hz, 1H); 4.34 (d, J = 5.7 Hz, 2H); 4.17 (dd, J = 4.8 Hz, 1H); 3.97 (dd, J = 3.8, 3.8 Hz, 1H); 3.69 (ddd, J = 12.0, 5.2, 4.0 Hz, 1H); 3.57 (ddd, J = 12.0, 6.1, 3.8 Hz, 1H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 153.7; 150.7; 150.6; 149.9; 142.4; 120.5; 118.1; 87.7; 85.7; 73.8; 70.3; 61.3; 28.4. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{13}H_{16}N_7O_5$ 350.1207; Found 350.1208.

10.3 Synthesis of N^6-methyl N^6-methylurea adenosine

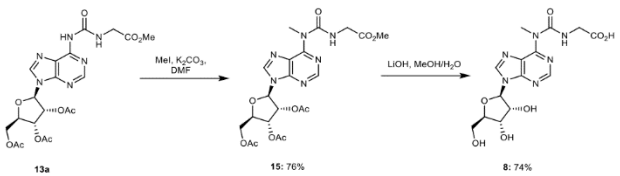


Scheme S15. Synthesis of N^6-methyl N^6-methylurea adenosine **7**.

Acetyl-protected N^6-methyl N^6-methylurea adenosine 14: Acetyl-protected N^6-methylurea adenosine **11** (300 mg, 0.67 mmol, 1.0 equiv.) was dissolved in dry DMF. The solution was cooled to 0°C and K_2CO_3 (276 mg, 2.00 mmol, 3.0 equiv.) was added. Finally, methyl iodide (82.9 µL, 1.33 mmol, 2.0 equiv.) was added dropwise to the mixture. The reaction was stirred at r.t. for 16 h. After that, the crude reaction mixture was diluted with Et_2O and washed with sat. aq. NH_4Cl solution and water. The aqueous phase was re-extracted with Et_2O and the combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (98:2 DCM/IPA) affording **14** as a white foam (151 mg, 0.32 mmol, 49% yield). R_f = 0.55 (9:1 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3483; 3198; 2950; 1742; 1678; 1583; 1566; 1532; 1467; 1423; 1367; 1327; 1211; 1093; 1042; 1000; 920; 901; 795. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.31 (q, J = 4.9 Hz, 1H); 8.52 (s, 1H); 8.08 (s, 1H); 6.23 (d, J = 5.3 Hz, 1H); 5.93 (dd, J = 5.3, 5.3 Hz, 1H); 5.64 (dd, J = 5.3, 4.6 Hz, 1H); 4.54-4.32 (m, 3H); 3.99 (s, 3H); 2.99 (d, J = 4.9 Hz, 3H); 2.15 (s, 3H); 2.13 (s, 3H); 2.08 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 170.5; 169.8; 169.5; 156.6; 153.6; 152.0; 150.5; 139.2; 122.6; 86.5; 80.5; 73.2; 70.7; 63.2; 34.8; 27.5; 21.0; 20.7; 20.6. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{19}H_{25}N_6O_8$ 465.1728; Found 465.1726.

N^6-methyl N^6-methylurea adenosine 7: Acetyl-protected adenosine **14** (119 mg, 0.26 mmol, 1.0 equiv.) was dissolved in 7 N NH_3 in MeOH. The reaction was stirred at r.t. overnight. After that, the crude was concentrated and purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **7** as a white solid (66 mg, 0.20 mmol, 76% yield). t_R = 26.9 min (0-40% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3362; 3225; 3146; 1689; 1586; 1561; 1514; 1463; 1258; 1129; 1102; 1055; 1039; 983; 817; 793, 768, 740; 696. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 9.72 (q, J = 4.5 Hz, 1H); 8.69 (s, 1H); 8.57 (s, 1H); 6.01 (d, J = 5.6 Hz, 1H); 5.54 (d, J = 5.6 Hz, 1H); 5.26 (d, J = 4.9 Hz, 1H); 5.22-5.10 (m, 1H); 4.57 (q, J = 5.3 Hz, 1H); 4.17 (q, J = 4.5 Hz, 1H); 3.97 (q, J = 3.9 Hz, 1H); 3.78 (s, 3H); 3.67-3.59 (m, 2H); 2.81 (d, J = 4.5 Hz, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 155.8; 152.7; 151.8; 150.3; 141.3; 121.9; 87.6; 85.7; 73.8; 70.3; 61.2; 34.2; 27.2. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{13}H_{19}N_6O_5$ 339.1411; Found 339.1413.

10.4 Synthesis of amino acid-modified N^6-methyl N^6-carbamoyl adenosine nucleoside

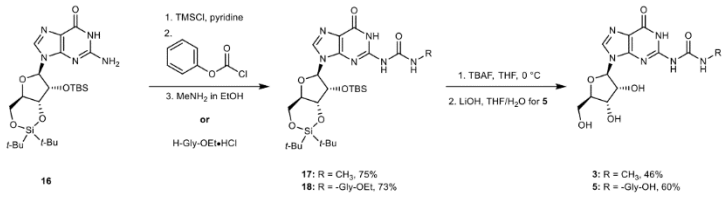


Scheme S16. Synthesis of glycine-modified N^6-methyl N^6-carbamoyl adenosine **8** used as reference.

Acetyl-protected N^6-glycine modified N^6-methyl adenosine 15: Acetyl-protected N^6-methylurea adenosine derivative **13a** (0.98 g, 1.94 mmol, 1.0 equiv.) was dissolved in dry DMF. The solution was cooled to 0°C and K_2CO_3 (0.80 g, 5.81 mmol, 3.0 equiv.) was added. Finally, methyl iodide (0.30 mL, 4.80 mmol, 2.5 equiv.) was added dropwise to the mixture. The reaction was stirred at r.t. for 16 h. After that, the crude reaction mixture was diluted with Et_2O, and washed with sat. aq. NH_4Cl solution and water. The aqueous phase was re-extracted with Et_2O and the combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (98:2 DCM/IPA) affording **15** as a white foam (0.77 g, 1.47 mmol, 76% yield). R_f = 0.69 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3196; 2955; 1740; 1681; 1565; 1518; 1463; 1370; 1223; 1206; 1029; 922; 796. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.97 (t, J = 5.2 Hz, 1H); 8.57 (s, 1H); 8.10 (s, 1H); 6.25 (d, J = 5.4 Hz, 1H); 5.92 (dd, J = 5.5, 5.5 Hz, 1H); 5.63 (dd, J = 5.5, 4.5 Hz, 1H); 4.49-4.34 (m, 3H); 4.21 (d, J = 5.4 Hz, 2H); 4.00 (s, 3H); 3.78 (s, 3H); 2.15 (s, 3H); 2.14 (s, 3H); 2.08 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 171.0; 170.4; 169.8; 169.5; 156.0; 153.4; 152.2; 150.5; 139.4; 122.7; 86.4; 80.5; 73.2; 70.8; 63.2; 52.4; 43.0; 34.9; 21.0; 20.7; 20.5. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{21}H_{27}N_6O_{10}$ 523.1782; Found 523.1784.

N^6-glycine modified N^6-methyl adenosine 8: Acetyl-protected adenosine derivative **15** (103 mg, 0.20 mmol, 1.0 equiv.) was dissolved in 3:1 MeOH/H_2O. The solution was cooled to 0°C. LiOH•H_2O (83 mg, 2.0 mmol, 10.0 equiv.) was added to the solution and the reaction was stirred at r.t. for 2 h. After that, the solution was neutralized with 1 M aqueous HCl solution. The organic solvent was removed under reduced pressure and the remaining aqueous solution mixture was lyophilized. The crude was purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **8** as a white solid (56 mg, 0.15 mmol, 74% yield). t_R = 27.1 min (0-30% of B in 45 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3389; 3319; 3220; 3122; 2946; 1718; 1668; 1566; 1538; 1468; 1231; 1209; 1145; 1122; 1086; 1073; 1024; 815; 749. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.68 (s, 1H); 10.36 (t, J = 5.5 Hz, 1H); 8.74 (s, 1H); 8.60 (s, 1H); 6.03 (d, J = 5.3 Hz, 1H); 5.56 (br s, 1H); 5.25 (br s, 1H); 5.14 (br s, 1H); 4.57 (dd, J = 5.3, 5.3 Hz, 1H); 4.18 (dd, J = 4.3, 4.3 Hz, 1H); 4.00-3.90 (m, 3H); 3.82 (s, 3H); 3.74-3.51 (m, 2H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 171.5; 155.3; 152.4; 152.1; 150.1; 141.6; 121.9; 87.6; 85.6; 73.9; 70.2; 61.2; 42.6; 34.3. HRMS (ESI) m/z: [M-H]$^-$ Calcd. for $C_{14}H_{17}N_6O_7$ 381.1164; Found 381.1165.

10.5 Synthesis of N^2-methylurea guanosine and amino acid-modified N^2-carbamoyl guanosine nucleoside used as reference



Scheme S17. Synthesis of N^2-methylurea guanosine **3** and amino acid-modified N^2-carbamoyl guanosine **5** used as reference.

Silyl-protected guanosine **16** was synthesized following a procedure previously reported in the literature.[8]

Silyl-protected N^2-methyurea guanosine 17: Step 1: Silyl-protected guanosine **16** (4.00 g, 7.44 mmol, 1.0 equiv.) was suspended in dry pyridine and TMSCl (1.50 mL, 11.9 mmol, 1.6 equiv.) was added. The reaction was stirred at r.t. for 1 h. After that, phenyl chloroformate (1.50 mL, 11.9 mmol, 1.6 equiv.) was added and the reaction was stirred at r.t. for 5 h. Step 2: 33% $MeNH_2$ in EtOH (4.70 mL, 37.8 mmol, 5.1 equiv.) was added dropwise and the

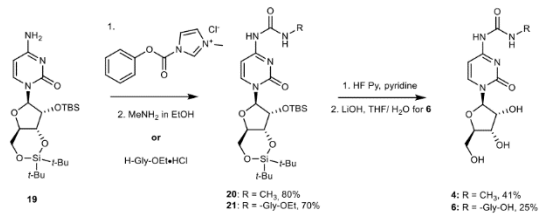
reaction was stirred at r.t. for 16 h. After that, the reaction was quenched with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (95:5 DCM/MeOH) affording **17** as a white solid (3.32 g, 7.44 mmol, 75% yield). R_f = 0.47 (9:1 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3204; 2900; 2859; 1662; 1602; 1550; 1475; 1399; 1345; 1256; 1170; 1120; 1090; 1058; 1020; 970; 840; 781; 755; 733. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.01 (s, 1H); 9.84 (s, 1H); 8.15 (s, 1H); 7.00 (q, J = 4.5 Hz, 1H); 5.80 (s, 1H); 4.59 (d, J = 5.2 Hz, 1H); 4.35 (dd, J = 8.4, 4.3 Hz, 1H); 4.25 (dd, J = 9.1, 5.2 Hz, 1H); 4.08-3.92 (m, 2H); 2.72 (d, J = 4.5 Hz, 3H); 1.06 (s, 9H); 1.00 (s, 9H); 0.86 (s, 9H); 0.08 (s, 3H); 0.07 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 155.4; 155.1; 149.0; 148.7; 137.2; 118.9; 90.0; 75.7; 74.9; 74.0; 66.9; 27.3; 26.9; 26.2; 25.7; 22.2; 20.0; 18.0; -4.6; -5.2. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{26}H_{47}O_6N_6Si_2$ 595.3090; Found 595.3088.

Silyl-protected N^2-glycine modified guanosine 18: Silyl-protected guanosine **16** (1.00 g, 1.86 mmol, 1.0 equiv.) was suspended in dry pyridine and TMSCl (0.40 mL, 3.15 mmol, 1.7 equiv.) was added. The reaction was stirred at r.t. for 2 h. After that, phenyl chloroformate (0.40 mL, 3.17 mmol, 1.7 equiv.) was added and the mixture was stirred at r.t. for 5 h. **H-Gly-OEt•HCl** (0.52 g, 3.72 mmol, 2.0 equiv.) was added to the solution and the mixture was stirred at r.t. for 16 h. After that, the reaction was quenched with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (DCM/MeOH) affording **18** as a white solid (0.90 g, 1.35 mmol, 73% yield). R_f = 0.62 (9:1 DCM/MeOH). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2932; 2859; 1827; 1735; 1661; 1618; 1550; 1472; 1397; 1366; 1251; 1142; 1054; 999; 940; 894; 832; 781; 747; 744; 687. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 11.83 (br s, 1H); 10.09 (br s, 1H); 8.17 (s, 1H); 7.53 (s, 1H); 5.83 (s, 1H); 4.59 (d, J = 5.1 Hz, 1H); 4.36 (dd, J = 8.3, 4.2 Hz, 1H); 4.26 (dd, J = 9.0, 5.1 Hz, 1H); 4.14 (q, J = 7.1 Hz, 2H); 4.07-3.92 (m, 4H); 1.21 (t, J = 7.1 Hz, 3H); 1.07 (s, 9H); 1.01 (s, 9H); 0.86 (s, 9H); 0.08 (s, 3H); 0.07 (s, 3H). ^{13}C{^{1}H} NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 169.7; 155.0; 148.8; 148.7; 148.5; 137.4; 119.1; 90.1; 75.7; 74.9; 74.1; 66.9; 60.8; 41.4; 27.3; 26.9; 25.7; 22.2; 20.0; 18.0; 14.1; -4.6; -5.1. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{29}H_{51}O_8N_6Si_2$ 667.3301; Found 667.3295.

N^2-methylurea guanosine 3: Silyl-protected cytidine derivative **17** (0.40 g, 0.67 mmol, 1.0 equiv.) was suspended in THF. 1 M tetrabutylammonium fluoride (TBAF) in THF (4.0 mL, 4.0 mmol, 6.0 equiv.) was added and the reaction was stirred at r.t. for 4 h. After that, the reaction was quenched by addition of methoxytrimethylsilane (0.56 mL, 4.0 mmol, 6.0 equiv.). The crude was concentrated under reduced pressure and purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **3** as a white solid (106 mg, 0.31 mmol, 46% yield). t_R = 12.1 min (0-30% of B in 30 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3257; 1703; 1659; 1636; 1554; 1497; 1475; 1406; 1354; 1267; 1233; 1170; 1125; 1090; 1058; 1042; 1020; 972; 862; 821; 781; 750; 720. ^{1}H NMR (500 MHz, DMSO-d_6, 298 K): δ (ppm) = 11.97 (br s, 1H); 10.18 (br s, 1H); 8.16 (s, 1H); 6.82 (q, J = 4.7 Hz, 1H); 5.75 (d, J = 5.5 Hz, 1H); 5.47 (d, J = 5.9 Hz, 1H); 5.17 (d, J = 4.9 Hz, 1H); 5.04 (t, J = 5.4 Hz, 1H); 4.41 (q, J = 5.5 Hz, 1H); 4.11 (q, J = 4.6 Hz, 1H); 3.90 (q, J = 4.0 Hz, 1H); 3.64 (ddd, J = 12.0, 5.5, 4.0 Hz, 1H); 3.54 (ddd, J = 11.2, 5.5, 4.1 Hz, 1H); 2.72 (d, J = 4.7 Hz, 3H). ^{13}C{^{1}H} NMR (126 MHz, DMSO-d_6, 298 K): δ (ppm) = 155.6; 155.1; 149.2; 148.9; 137.1; 119.1; 86.9; 85.3; 74.0; 70.2; 61.1; 26.2. HRMS (ESI) m/z: [M+H]$^+$ Calcd. for $C_{12}H_{17}N_6O_6$ 341.1204; Found 341.1203.

N^2-glycine modified guanosine 5: Step 1: Silyl-protected guanosine derivative **18** (20 mg, 30 µmol, 1.0 equiv.) was suspended in THF. 1 M tetrabutylammonium fluoride (TBAF) in THF (0.18 mL, 0.18 mmol, 6.0 equiv.) was added and the reaction was stirred at r.t. for 4 h. The reaction was quenched by addition of methoxytrimethylsilane (25 µL, 0.18 mmol, 6.0 equiv.). The crude was concentrated under reduced pressure. Step 2: The residue was redissolved in 1:1 THF/H_2O and LiOH (13 mg, 0.30 mmol, 10.0 equiv.) was added. The reaction was stirred at 0°C for 1 h. Subsequently, the reaction was quenched by addition of aqueous 1 M HCl solution. The organic solvent was removed under reduced pressure and the remaining aqueous solution was lyophilized. The crude was purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **5** as a white solid (7.0 mg, 30 µmol, 60% yield). t_R: 11.8 min (0-30% of B in 30 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3370; 1693; 1611; 1532; 1460; 1362; 1209; 1051; 774; 665. ^{1}H NMR (500 MHz, DMSO-d_6, 298 K): δ (ppm) = 8.19 (s, 1H); 7.61 (br s, 1H); 5.80 (d, J = 5.0 Hz, 1H); 5.06 (br s, 1H); 4.37 (t, J = 5.0 Hz, 1H); 4.11 (t, J = 4.6 Hz, 1H); 3.91 (q, J = 4.0 Hz, 1H); 3.80-3.72 (m, 2H); 3.66 (dd, J = 12.0, 4.0 Hz, 1H); 3.56 (dd, J = 12.0, 4.0 Hz, 1H). ^{13}C{^{1}H} NMR (126 MHz, DMSO-d_6, 298 K): δ (ppm) = 170.8; 155.2; 154.6; 148.9; 148.6; 137.2; 119.2; 87.2; 85.1; 74.1; 69.9; 61.0; 42.3. HRMS (ESI) m/z: [M-H]$^-$ Calcd. for $C_{13}H_{15}N_6O_8$ 383.0956; Found 383.0964.

10.6 Synthesis of N^4-methylurea cytidine and amino acid-modified N^4-carbamoyl cytidine nucleoside used as reference



Scheme S18. Synthesis of N^4-methylurea cytidine **4** and amino acid-modified N^4-carbamoyl cytidine **6** used as reference.

Silyl-protected cytidine **19** was synthesized following a procedure previously reported in the literature.[8]

Silyl-protected N^4-methyurea cytidine 20: Silyl-protected cytidine **19** (3.00 g, 6.03 mmol, 1.0 equiv.) and 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (2.88 g, 12.1 mmol, 2.0 equiv.) were dissolved in dry DCM. The reaction was stirred at r.t. for 16 h. 2 M $MeNH_2$ in THF (15.1 mL, 30.1 mmol, 5.0 equiv.) was added to the solution and the mixture was stirred at r.t. for 16 h. After that, the reaction was quenched with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (1:1 DCM/EtOAc) affording **20** as a white solid (2.67 g, 4.81 mmol, 80% yield). R_f = 0.21 (1:1 DCM/EtOAc). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2933; 1720; 1641; 1562; 1509; 1472; 1416; 1382; 1326; 1204; 1166; 1145; 1129; 1115; 1079; 1053; 1023; 1000; 940; 902; 829; 783; 751; 712; 687; 652; 493; 441; 408. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.78 (s, 1H); 9.11 (s, 1H); 7.67 (d, J = 7.8 Hz, 1H); 7.60 (d, J = 7.8 Hz, 1H); 5.69 (s, 1H); 4.55 (dd, J = 9.3, 5.2 Hz, 1H); 4.35-4.24 (m, 2H); 4.02 (t, J = 10.0 Hz, 1H); 3.80 (dd, J = 9.7, 4.2 Hz, 1H); 2.82 (d, J = 4.3 Hz, 3H); 1.03 (s, 9H); 1.02 (s, 9H); 0.95 (s, 9H); 0.24 (s, 3H); 0.17 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 165.1; 156.4; 154.8; 141.3; 97.5; 94.2; 75.8; 75.4; 74.9; 67.9; 27.6; 27.1; 26.7; 26.0; 23.0; 20.5; 18.3; -4.2; -4.6. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{25}H_{47}N_4O_6Si_2$ 555.3028; Found 555.3029.

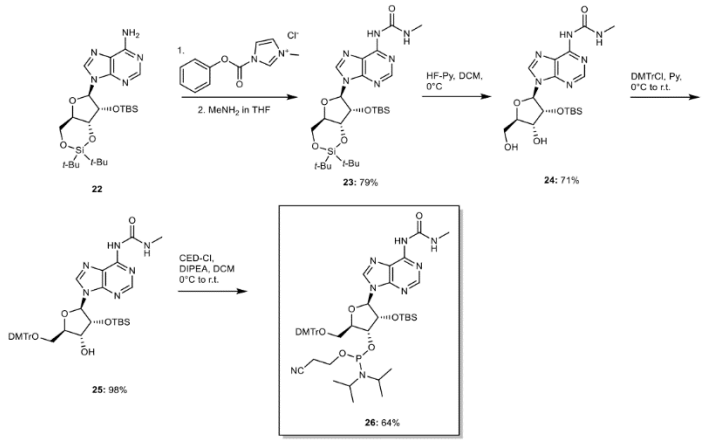
Silyl-protected N^4-glycine modified cytidine 21: Silyl-protected cytidine **19** (1.00 g, 2.01 mmol, 1.0 equiv.) and 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (0.90 g, 4.02 mmol, 2.0 equiv.) were dissolved in dry DCM. The reaction was stirred at r.t. for 14 h. Subsequently, a suspension in dry DCM of **H-Gly-OEt•HCl** (0.56 g, 4.02 mmol, 2.0 equiv.), containing TEA (0.84 mL, 6.03 mmol, 3.0 equiv.) was transferred to the reaction mixture. The reaction was stirred at r.t. for 16 h. After that, the reaction was quenched with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (0 to 50% *i*-Hex/EtOAc) affording **21** as a white solid (0.88 g, 1.40 mmol, 70% yield). R_f = 0.33 (1:1 *i*-Hex/EtOAc); IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3229; 3063; 2932; 2857; 1718; 1643; 1570; 1503; 1471; 1434; 1373; 1263; 1187; 1165; 1126; 1053; 1026; 1010; 996; 937; 897; 826; 783; 752; 713; 685. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 10.88 (s, 1H); 9.68 (s, 1H); 7.70-7.55 (m, 2H); 5.64 (s, 1H); 4.52 (dd, J = 9.2, 5.2 Hz, 1H); 4.29-4.25 (m, 1H); 4.26-4.21 (m, 1H); 4.17 (q, J = 7.2 Hz, 2H); 4.02-3.96 (m, 1H); 3.96-3.90 (m, 2H); 3.75 (dd, J = 9.7, 4.1 Hz, 1H); 1.24 (t, J = 7.2 Hz, 3H); 1.00 (s, 9H); 0.99 (s, 9H); 0.91 (s, 9H); 0.16 (s, 3H); 0.12 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 170.1; 165.0; 156.3; 154.7; 141.3; 97.6; 94.0; 75.6; 75.3; 74.8; 67.8; 61.1; 41.7; 27.0; 25.9; 22.7; 20.4; 18.2; 14.3; -4.3; -4.8. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{28}H_{51}N_4O_8Si_2$ 627.3240; Found 627.3243.

N^4-methylurea cytidine 4: Silyl-protected cytidine derivative **20** (0.37 g, 0.67 mmol, 1.0 equiv.) was dissolved in dry pyridine. 70% HF-pyridine (0.54 mL, 3.33 mmol, 5 equiv.) was added to the solution and the reaction was stirred at r.t. for 5 h. After that, the reaction was quenched by addition of methoxytrimethylsilane (1.38 mL, 4.79 mmol, 15 equiv.). The crude was concentrated under reduced pressure and purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **4** as a white solid (82 mg, 0.67 mmol, 41% yield). t_R = 13.2 min (0-15% of B in 15 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3257; 1712; 1693; 1651; 1602; 1561; 1538; 1483; 1374; 1334; 1308; 1229; 1154; 1138; 1098; 1061; 1032; 983; 944; 871; 851; 811; 792; 727; 704; 650; 601; 586; 438. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 9.92 (s, 1H); 8.24 (d, J = 7.4 Hz, 1H); 6.22 (br s, 1H); 5.76 (d, J = 3.5 Hz, 1H); 5.43 (br s, 1H); 5.12 (br s, 1H); 5.04 (br s, 1H); 4.00-3.91 (m, 2H); 3.87 (dt, J = 5.9, 3.0 Hz, 1H); 3.70 (dd, J = 12.2, 2.8 Hz, 1H); 3.57 (dd, J = 12.2, 3.1 Hz, 1H); 2.75 (d, J = 4.7 Hz, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 162.2; 154.2; 153.7; 143.7; 94.6; 89.8; 84.2; 74.3; 68.9; 60.1; 26.0. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{11}H_{17}N_4O_6$ 301.1142; Found 301.1144.

N^4-glycine modified cytidine 6: Step 1: Silyl-protected cytidine derivative **21** (200 mg, 0.32 mmol, 1.0 equiv.) was dissolved in dry pyridine. 70% HF-pyridine (0.26 mL, 1.60 mmol, 5 equiv.) was added to the solution and the reaction was stirred at r.t. for 5 h. After that, the reaction was quenched by addition of methoxytrimethylsilane (0.66 mL, 4.79 mmol, 15 equiv.). The crude was concentrated under reduced pressure. Step 2: The obtained residue was redissolved in 1:1 THF/H_2O and LiOH (67 mg, 1.6 mmol, 5.0 equiv.) was added to the suspension. The reaction was stirred at r.t. for 1 h. Subsequently, the crude reaction mixture was neutralised by addition of Dowex ion-exchange resin. The resin was filtered off and the organic solvent was removed under reduced pressure. The remaining aqueous solution was lyophilized and the crude was purified by semi-preparative HPLC (A: H_2O and B: 20:80 H_2O/MeCN; both containing 0.1% of formic acid) affording **6** as a white solid (27 mg, 0.08 mmol, 25% yield). t_R = 12.1 min (0-15% of B in 15 min). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3339; 3191; 3096; 2942; 2544; 1703; 1655; 1608; 1572; 1530; 1468; 1446; 1371; 1361; 1307; 1252; 1219; 1171; 1135; 1107; 1085; 1059; 1035; 994; 983; 964; 915; 905; 849; 840; 805; 791; 755; 722; 696; 667. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 12.67 (br s, 1H); 10.04 (br s, 1H); 9.10 (br s, 1H); 8.28 (d, J = 7.4 Hz, 1H); 6.26 (s, 1H); 5.77 (d, J = 3.0 Hz, 1H); 5.46 (d, J = 4.7 Hz, 1H); 5.15 (t, J = 5.1 Hz, 1H); 5.07 (d, J = 5.3 Hz, 1H); 4.02-3.90 (m, 4H); 3.90-3.85 (m, 1H); 3.71 (dd, J = 12.1, 2.9 Hz, 1H); 3.58 (dd, J = 12.1, 3.1 Hz, 1H). $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 171.3; 162.2; 153.9; 153.7; 143.9; 94.6; 89.9; 84.2; 74.4; 68.8; 60.0; 41.4. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{12}H_{17}N_4O_8$ 345.1040; Found 345.1043.

11. Synthesis of methylurea nucleoside phosphoramidites

11.1 Synthesis of N^6-methylurea adenosine phosphoramidite



Scheme S19. Synthesis of N^6-methylurea adenosine phosphoramidite **26**.

Silyl-protected adenosine **22** was synthesized following a procedure previously reported in the literature.[8]

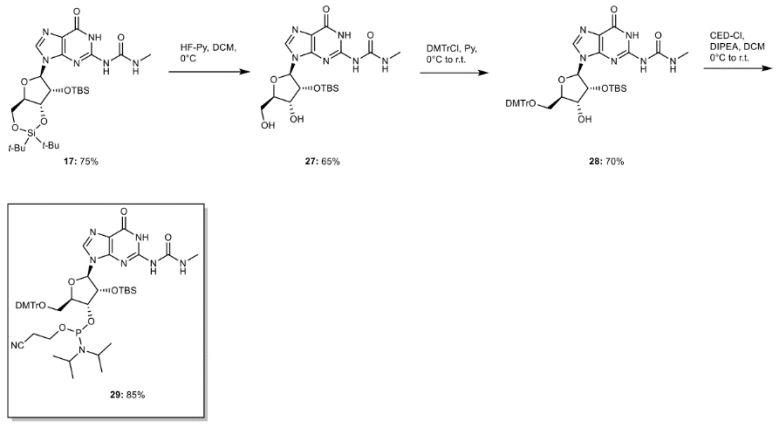
Silyl-protected N^6-methylurea adenosine 23: Silyl-protected adenosine **22** (4.00 g, 7.67 mmol, 1.0 equiv.) and 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (3.66 g, 15.3 mmol, 2.0 equiv.) were dissolved in dry DCM. The reaction was stirred at r.t. for 16 h. 2 M $MeNH_2$ in THF (15.3 mL, 30.7 mmol, 4.0 equiv.) was added to the solution and the mixture was stirred at r.t. for 16 h. After that, the reaction was quenched with sat. aq. $NaHCO_3$ solution and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel (95:5 DCM/IPA) affording **23** as a white foam (3.52 g, 6.08 mmol, 79% yield). R_f = 0.58 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3248; 2932; 2858; 1706; 1615; 1586; 1558; 1471; 1257; 1138; 1107; 1066; 1011; 1001; 893; 825; 783; 754. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 9.33 (br s, 1H); 8.50 (s, 1H); 8.12 (s, 1H); 5.98 (br s, 1H); 4.57 (d, J = 4.5 Hz, 1H); 4.50 (dd, J = 9.2, 5.1 Hz, 1H); 4.48-4.42 (m, 1H); 4.24 (td, J = 10.0, 5.1 Hz, 1H); 4.06 (dd, J = 10.5, 9.2 Hz, 1H); 3.01 (d, J = 4.7 Hz, 3H); 1.08 (s, 9H); 1.04 (s, 9H); 0.93 (s, 9H); 0.16 (s, 3H); 0.15 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 154.8; 149.7; 141.3; 121.0; 92.4; 76.0; 75.8; 74.9; 67.9; 27.6; 27.1; 26.8; 26.0; 22.9; 20.5; 18.5; -4.1; -4.9. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{26}H_{47}N_6O_5Si_2$ 579.3141; Found 579.3142.

5',3'-deprotected N^6-methylurea adenosine 24: Modified adenosine **23** (1.68 g, 2.90 mmol, 1.0 equiv.) was dissolved in 9:1 DCM/pyridine in a plastic flask and the solution was cooled to 0°C. Subsequently, 70% HF-pyridine (5.0 equiv.) was added slowly to the adenosine's solution and the reaction was stirred at 0°C for 2 h. After that, the crude reaction mixture was diluted with sat. aq. $NaHCO_3$ solution and the crude was extracted three times with DCM. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by silica gel column chromatography (96:4 DCM/IPA) affording **24** as a white solid (0.90 g, 2.04 mmol, 71% yield). R_f = 0.48 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3244; 2928; 2857; 1694; 1611; 1590; 1548; 1470; 1360; 1331; 1297; 1252; 1220; 1131; 1085; 1062; 1024; 865; 835; 779. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 9.26 (q, J = 4.7 Hz, 1H); 8.51 (s, 1H); 8.41 (s, 1H); 8.13 (s, 1H); 5.97 (dd, J = 11.9, 2.2 Hz, 1H); 5.84 (d, J = 7.3 Hz, 1H); 5.08 (dd, J = 7.3, 4.8 Hz, 1H); 4.40-4.30 (m, 2H); 3.97 (dt, J = 12.9, 2.0 Hz, 1H); 3.77 (ddd, J = 13.2, 11.9, 1.6 Hz, 1H); 3.02 (d, J = 4.7 Hz, 3H); 2.88 (d, J = 0.7 Hz, 1H); 0.79 (s, 9H); -0.19 (s, 3H); -0.39 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 154.4; 151.1; 151.0; 149.2; 143.1; 122.0; 91.2; 87.7; 74.7; 72.9; 63.4; 26.9; 25.6; 18.0; -5.2; -5.3. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{18}H_{31}N_6O_5Si_2$ 439.2120; Found 439.2121.

DMTr-protected N^6-methylurea adenosine 25: 3',5'-Deprotected adenosine **24** (0.80 g, 1.82 mmol, 1.0 equiv.) was dissolved in dry pyridine. DMTrCl (0.87 g, 2.55 mmol, 1.4 equiv.) was added. The reaction was stirred at r.t. for 16 h. After that, the crude was concentrated under reduced pressure and purified by silica gel column chromatography (97:3 DCM/IPA, containing 0.1% pyridine). The DMTr-protected compound **25** was isolated as a white foam (1.34 g, 1.80 mmol, 98% yield). R_f = 0.70 (95:5 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3553; 2927; 1702; 1608; 1589; 1548; 1507; 1468; 1299; 1248; 1174; 1066; 1031; 833; 780; 700. ^{1}H NMR (400 MHz, $CDCl_3$, 298 K): δ (ppm) = 9.35 (s, 1H); 8.72-8.61 (m, 1H); 8.54 (s, 1H); 8.45 (s, 1H); 7.50 (d, J = 7.5 Hz, 2H); 7.39-7.34 (m, 4H); 7.29-7.25 (m, 2H); 7.20 (t, J = 7.3 Hz, 1H); 6.86-6.82 (m, 4H); 6.15 (d, J = 4.7 Hz, 1H); 5.10 (t, J = 4.7 Hz, 1H); 4.51 (q, J = 5.0 Hz, 1H); 4.29 (q, J = 3.7, 3.3 Hz, 1H); 4.02 (d, J = 5.6 Hz, 1H); 3.76 (s, 3H); 3.76 (s, 3H); 3.49-3.42 (m, 2H); 2.91 (d, J = 4.6 Hz, 3H); 0.84 (s, 9H); 0.04 (s, 3H); -0.07 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$, 298 K): δ (ppm) = 159.6; 154.8; 151.8; 151.4; 151.2; 146.1; 143.3; 136.7; 131.0; 130.9; 129.0; 128.6; 127.5; 121.5; 113.8; 89.8; 87.1; 84.8; 76.4; 72.0; 64.4; 55.4; 26.6; 26.1; 18.6; -4.7; -4.9. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{39}H_{49}N_6O_7Si$ 741.3427; Found 741.3428.

N^6-methylurea adenosine phosphoramidite 26: 5'-DMTr-protected adenosine **25** (300 mg, 0.41 mmol, 1.0 equiv.) was dissolved in dry DCM and *N,N*-diisopropylethylamine (DIPEA) (0.28 mL, 1.62 mmol, 4.0 equiv.) was added to the solution. The mixture was cooled to 0°C, followed by the addition of 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (0.23 mL, 1.01 mmol, 2.5 equiv.). The reaction was stirred at r.t. for 5 h. After that, sat. aq. $NaHCO_3$ solution was added to the crude reaction mixture and the aqueous phase was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by silica gel column chromatography (1:1 to 4:6 Hex/EtOAc, containing 0.1% pyridine). Finally, the product was lyophilized from benzene affording **26** as a mixture of diastereoisomers (245 mg, 0.26 mmol, 64% yield). R_f = 0.17 (1:1 Hex/EtOAc). $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K): δ (ppm) = 150.4; 149.1. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{48}H_{66}N_8O_8PSi$ 941.4505; Found 941.4488.

11.2 Synthesis of N^2-methylurea guanosine phosphoramidite



Scheme S20. Synthesis of N^2-methylurea guanosine phosphoramidite **29**.

The synthetic procedure of silyl-protected N^2-methylurea guanosine **18** is described in Section 10.5.

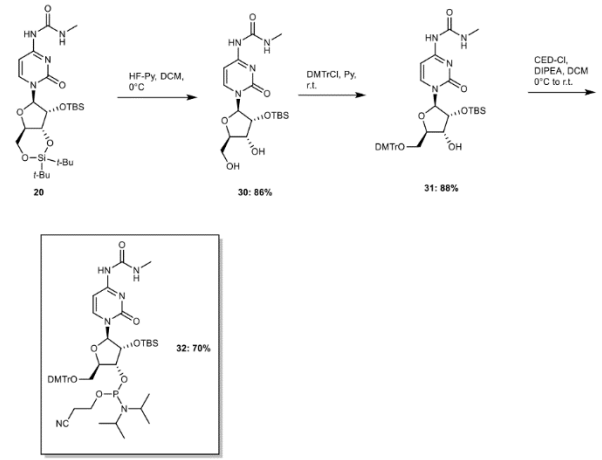
5',3'-deprotected N^2-methylurea guanosine 27: Protected N^2-methylurea guanosine **17** (1.6 g, 2.7 mmol, 1.0 equiv.) was dissolved in 9:1 DCM/pyridine in a plastic flask and the solution was cooled to 0°C. Subsequently, 70% HF-pyridine (0.35 mL, 13 mmol, 5.0 equiv.) was added slowly to the solution and the reaction was stirred at 0°C for 2 h. After that, the crude reaction mixture was diluted with sat. aq. $NaHCO_3$ solution and the crude was extracted with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by silica gel column chromatography (94:6 DCM/IPA) affording **27** as a white solid (0.80 g, 2,7 mmol, 65% yield). R_f = 0.4 (9:1 DCM/MeOH). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2857; 1666; 1606; 1534; 1403; 1362; 1321; 1145; 1089; 1058; 1020; 908; 835; 779; 750; 641. ^{1}H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 11.98 (br s, 1H); 10.06 (br s, 1H); 8.18 (s, 1H); 6.86 (br s, 1H); 5.79 (d, J = 6.2 Hz, 1H); 5.10 (t, J = 5.4 Hz, 1H); 5.04 (d, J = 5.1 Hz, 1H); 4.51 (dd, J = 6.2, 4.9 Hz, 1H); 4.09 (td, J = 5.0, 2.9 Hz, 1H); 3.95 (q, J = 3.7 Hz, 1H); 3.66 (ddd, J = 12.0, 5.4, 4.1 Hz, 1H); 3.57 (ddd, J = 12.0, 5.4, 3.8 Hz, 1H); 2.71 (d, J = 4.6 Hz, 3H); 0.74 (s, 9H); -0.06 (s, 3H); -0.17 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 155.5; 149.3; 137.1; 119.1; 86.5; 85.9; 76.2; 70.5; 61.3; 26.2; 25.5; 17.8; -4.9; -5.4. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{18}H_{31}O_6N_6Si$ 455.2068; Found 455.2069.

DMTr-protected N^2-methylurea guanosine 28: 3',5'-Deprotected N^2-methylurea guanosine derivative **27** (0.50 g, 1.1 mmol, 1.0 equiv.) was dissolved in dry pyridine and 4,4'-dimethoxytrityl chloride (DMTrCl) (0.56 g, 1.65 mmol, 1.5 equiv.) was added to the solution. The reaction was stirred at r.t. for 16 h. After that, the reaction mixture was concentrated under reduced pressure and the crude was purified by silica gel column chromatography (96:4 DCM/MeOH, containing 0.1% of pyridine). The DMTr-protected compound **28** was isolated as a white foam (0.59 g, 0.78 mmol, 70% yield). R_f = 0.50 (9:1 DCM/MeOH). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2928; 1666; 1607; 1548; 1508; 1484; 1410; 1361; 1298; 1249; 1174; 1128; 1074; 1035; 969; 914; 833; 782; 752; 726; 699. ^{1}H NMR (400 MHz, acetone-d_6, 298 K): δ (ppm) = 12.21 (br s, 1H); 9.61 (br s, 1H); 8.00 (s, 1H); 7.49-7.44 (m, 2H); 7.38-7.17 (m, 8H); 6.88-6.80 (m, 4H); 6.79-6.70 (m, 1H); 5.94 (d, J = 5.1 Hz, 1H); 4.83 (t, J = 5.1 Hz, 1H); 4.50-4.42 (m, 1H); 4.22-4.15 (m, 1H); 3.89 (br s, 1H); 3.77 (s, 6H); 3.50-3.31 (m, 2H); 2.83 (d, J = 4.4 Hz, 3H); 0.84 (s, 9H); 0.05 (s, 3H); -0.07 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K): δ (ppm) = 159.6; 156.6; 150.3; 146.0; 137.5; 136.6; 130.9; 128.9; 128.6; 127.6; 121.0; 113.9; 88.7; 87.1; 84.9; 77.3; 72.2; 64.8; 55.5; 26.7; 26.1; 18.7; -4.7; -4.9. HRMS (ESI) m/z: $[M+H]^+$ Calcd. for $C_{39}H_{49}O_8N_6Si$ 757.3375; Found 757.3375.

N^2-methylurea guanosine phosphoramidite 29: A solution of 5'-DMTr-protected guanosine derivative **28** (250 mg, 0.33 mmol, 1.0 equiv.) and DIPEA (0.23 mL, 1.3 mmol, 4.0 equiv.) in dry DCM was cooled to 0°C. 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (0.18 mL, 0.83 mmol, 2.5 equiv.) was added to the solution and the reaction mixture was stirred at r.t. for 5 h. After that, the reaction was quenched by addition of sat. aq. $NaHCO_3$ solution and the crude was extracted three times with DCM. The combined organic layers were dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by silica gel column chromatography (2:1 *n*-Hex/EtOAc, containing 0.1% of pyridine). Finally, the product was lyophilized from benzene affording **29** as

a mixture of diastereoisomers (270 mg, 0.28 mmol, 85% yield). R_f = 0.36 (9:1 DCM/MeOH). $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K): δ (ppm) = 150.7; 148.8. HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{48}H_{66}O_9N_8PSi$ 957.4453; Found 957.4436.

11.3 Synthesis of N^4-methylurea cytidine phosphoramidite



Scheme S21. Synthesis of N^4-methylurea cytidine phosphoramidite **32**.

The synthetic procedure of silyl-protected N^4-methylurea cytidine **20** is described in Section 10.6.

5',3'-deprotected N^4-methylurea cytidine 30: Protected cytidine **20** (250 mg, 0.45 mmol, 1.0 equiv.) was dissolved in 9:1 DCM/pyridine in a plastic flask and the solution was cooled to 0°C. Subsequently, 70% HF-pyridine (0.06 mL, 2.25 mmol, 5.0 equiv.) was added slowly to the solution and the reaction was stirred at 0°C for 2 h. After that, the crude reaction mixture was diluted with sat. aq. $NaHCO_3$ solution and the crude was extracted three times with DCM. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and concentrated under reduced pressure. The crude was purified by silica gel column chromatography (97:3 DCM/IPA) affording **30** as a white solid (160 mg, 0.39 mmol, 86% yield). R_f = 0.27 (9:1 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 2928; 1704; 1641; 1567; 1504; 1462; 1385; 1362; 1275; 1251; 1112; 1065; 999; 960; 918; 839; 812; 789; 738; 691; 668; 637; 598; 418; 405. ^1H NMR (400 MHz, DMSO-d_6, 298 K): δ (ppm) = 9.94 (s, 1H); 8.31 (d, *J* = 7.5 Hz, 1H); 6.21 (br s, 1H); 5.69 (d, *J* = 2.6 Hz, 1H); 5.18 (t, *J* = 5.1 Hz, 1H); 4.99 (d, *J* = 5.3 Hz, 1H); 4.07 (dd, *J* = 4.0, 2.6 Hz, 1H); 3.99-3.86 (m, 2H); 3.76 (ddd, *J* = 12.1, 5.1, 2.4 Hz, 1H); 3.60 (ddd, *J* = 12.1, 5.1, 2.6 Hz, 1H); 2.75 (d, *J* = 4.6 Hz, 3H); 0.86 (s, 9H); 0.07 (s, 3H), 0.05 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, DMSO-d_6, 298 K): δ (ppm) = 162.2; 154.2; 153.6; 143.3; 94.4; 90.0; 83.8; 76.4; 68.1; 59.5; 26.0; 25.8; 18.0; -4.8; -4.9. HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{17}H_{31}N_4O_6Si$ 415.2007; Found 415.2010.

DMTr-protected N^4-methylurea cytidine 31: 3',5'-Deprotected cytidine derivative **30** (100 mg, 0.24 mmol, 1.0 equiv.) was dissolved in dry pyridine and 4,4'-dimethoxytrityl chloride (DMTrCl) (123 mg, 0.36 mmol, 1.5 equiv.) was added to the solution. The reaction was stirred at r.t. for 16 h. After that, the reaction mixture was concentrated under reduced pressure and purified by silica gel column chromatography (96:4 DCM/IPA, containing 0.1% of pyridine) affording the DMTr-protected compound **31** as a white foam (152 mg, 0.24 mmol, 88% yield). R_f = 0.40 (9:1 DCM/IPA). IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3081; 1717; 1645; 1610; 1568; 1506; 1446; 1416; 1385; 1249; 1176; 1114; 1062; 1035; 1005; 828; 786; 754; 702; 585; 407. ^1H NMR (400 MHz, acetone-d_6, 298 K): δ (ppm) = 9.22 (br s, 1H); 8.48 (br s, 1H); 7.53-7.48 (m, 2H); 7.41-7.33 (m, 6H); 7.27 (t, *J* = 7.7 Hz, 1H); 6.97-6.88 (m, 4H); 5.85 (d, *J* = 1.1 Hz, 1H); 4.52 (td, *J* = 8.1, 4.5 Hz, 1H); 4.42 (d, *J* = 4.0 Hz, 1H); 4.22 (dt, *J* = 8.6, 2.7 Hz, 1H); 3.87 (d, *J* = 7.7 Hz, 1H); 3.82 (s, 6H); 3.54 (d, *J* = 2.4 Hz, 2H); 2.79 (d, *J* = 4.2 Hz, 3H); 0.97 (s, 9H); 0.29 (s, 3H); 0.21 (s, 3H). $^{13}C\{^1H\}$ NMR (100 MHz, acetone-d_6, 298 K): δ (ppm) = 159.7; 154.9; 145.4; 144.1; 136.8; 136.5; 131.0; 130.8; 129.1; 128.8; 127.8; 114.1; 92.3; 87.6; 83.1; 77.6; 69.7; 62.4; 55.5; 26.6; 26.3; 18.8; -4.2; -4.6. HRMS (ESI) *m/z*: $[M+H]^+$ Calcd. for $C_{38}H_{49}N_4O_8Si$ 717.3314; Found 717.3317.

N^4-methylurea cytidine phosphoramidite 32: A solution of 5'-DMTr-protected cytidine derivative **31** (90 mg, 0.13 mmol, 1.0 equiv.) and DIPEA (0.09 mL, 0.50 mmol, 4.0 equiv.) in dry DCM was cooled to 0°C. 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CED-Cl) (0.07 mL, 0.31 mmol, 2.5 equiv.) was added to the solution and the reaction was stirred at r.t. for 5 h. After that, 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. The crude was purified by silica gel column chromatography (2:1 *n*-Hex/EtOAc, containing 0.1% pyridine). Finally, the product was lyophilized from benzene affording **32** as a mixture of diastereoisomers (82 mg, 0.13 mmol, 70% yield). R_f = 0.20 (2:1 *n*-Hex/EtOAc). $^{31}P\{^1H\}$ NMR (162 MHz, acetone-d_6, 298 K): δ (ppm) = 150.4, 148.1. HRMS (ESI) *m/z*: $[M-H]^-$ Calcd. for $C_{47}H_{64}N_6O_9PSi$ 915.4247; Found 915.4263.

12. Synthesized oligonucleotides using a DNA/RNA automated synthesizer

ON2: 5'-**R**AA AAA AAA AA-3'; **R** = N^6-methylurea adenosine

ON2': 5'-**R**AU CGC UGU AC-3'; **R** = N^6-methylurea adenosine

ON2'': 5'-**R**AU CGC UGU **R**C-3'; **R** = N^6-methylurea adenosine

ON3: 5'-**R**AU CGC UGU AC-3'; **R** = N^2-methylurea guanosine

ON4: 5'-**R**AU CGC UGU AC-3'; **R** = N^4-methylurea cytidine

ON7: 5'-**R**UU AUU AUU UA-3'; **R** = N^6-methylurea adenosine

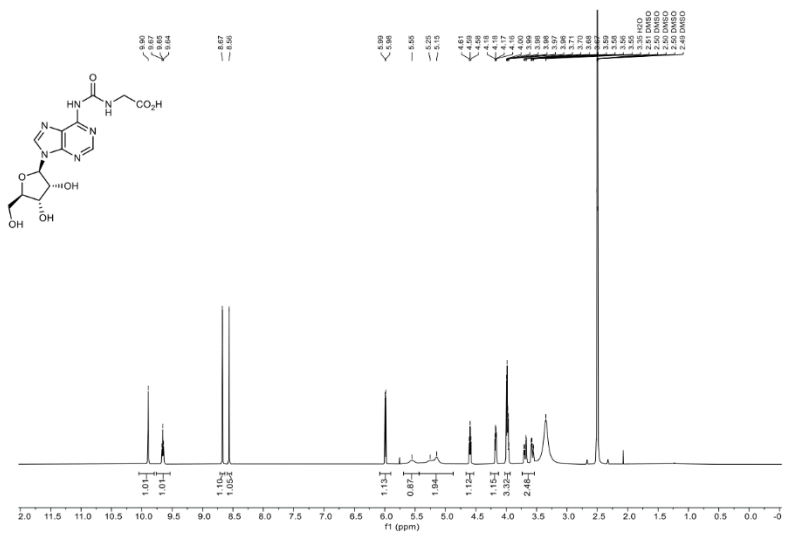
ON9: 5'-(UAA AUA AUA A)$_m$**R'**-3'; **R'** = nm^{5}U

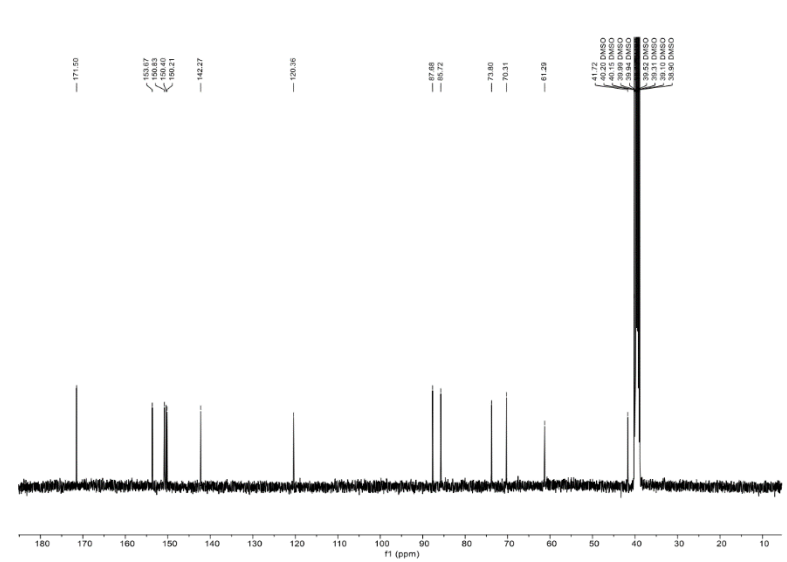
Table S11. HPLC retention times (0-30% of B in 45 min) and MALDI-TOF mass spectrometric analysis (negative mode) of **ON2-4**, **ON7** and **ON9**.

Strand	t_R (min)	*m/z* calcd. for $[M-H]^-$	Found
ON2	27.1	3613.6	3614.0
ON2'	26.7	3505.5	3504.4
ON2''	33.3	3832.6	3833.7
ON3	24.1	3520.5	3520.6
ON4	23.3	3480.5	3481.3
ON7	26.8	3452.4	3452.6
ON9	32.7	3633.5	3634.2

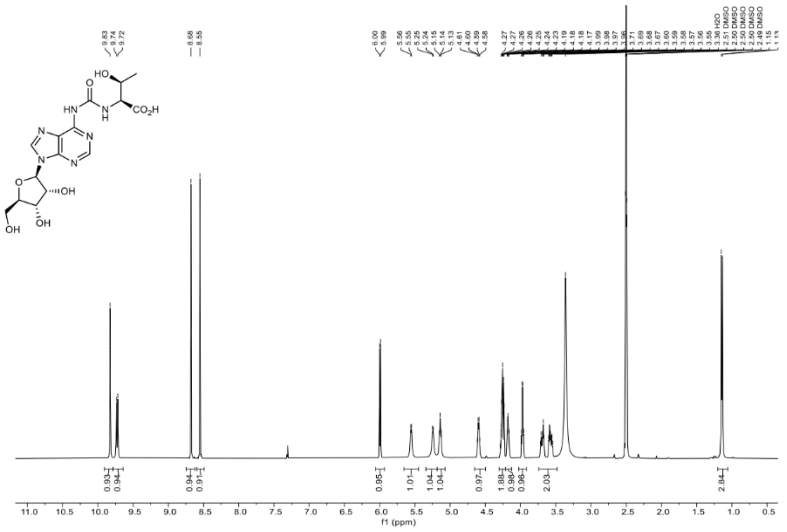
13. NMR spectra of synthesized compounds

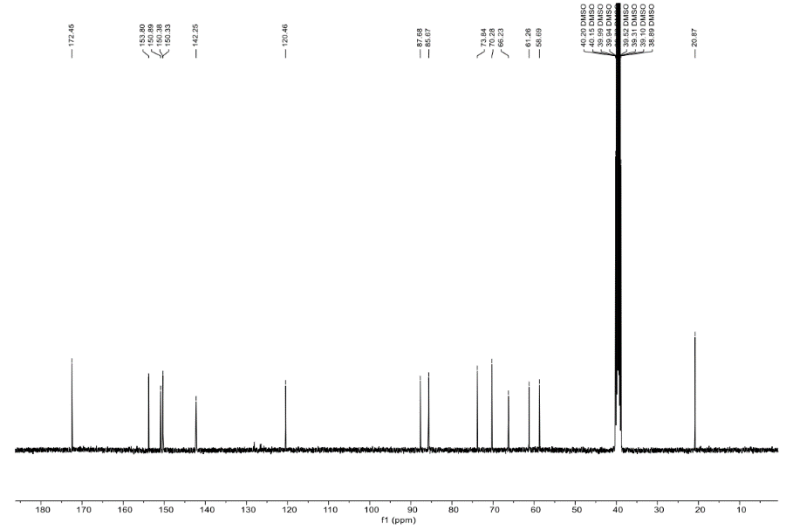
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1a



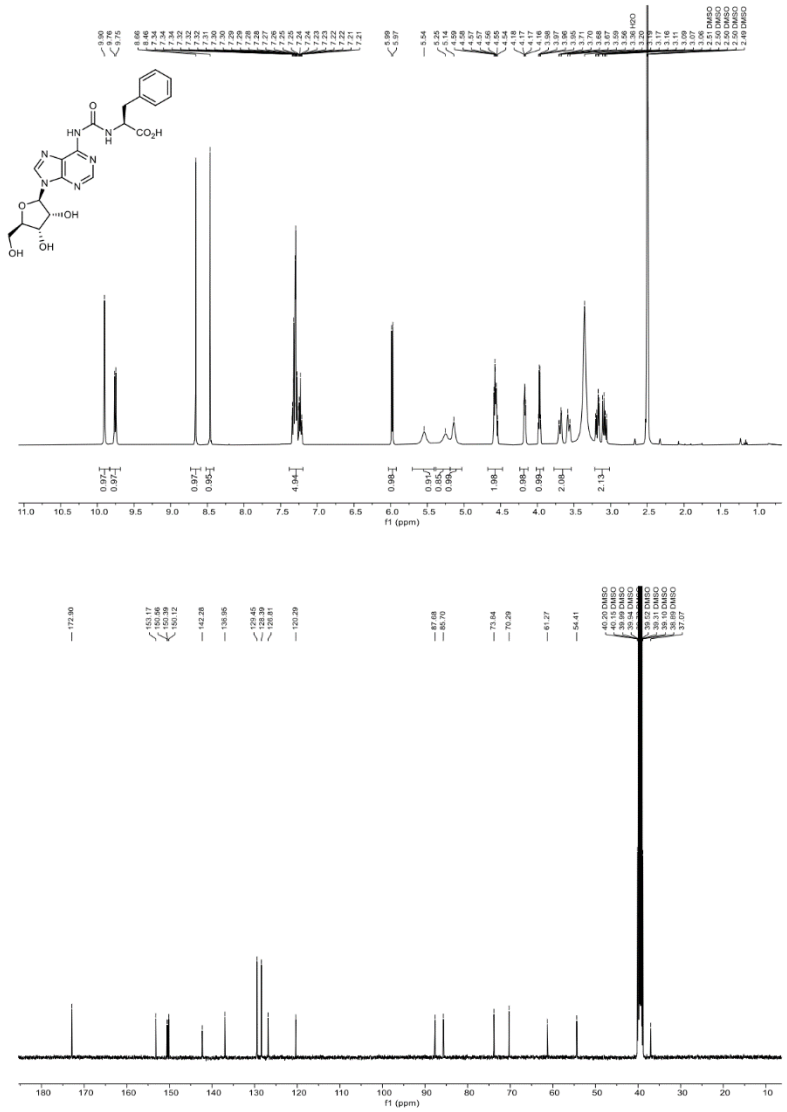


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1b

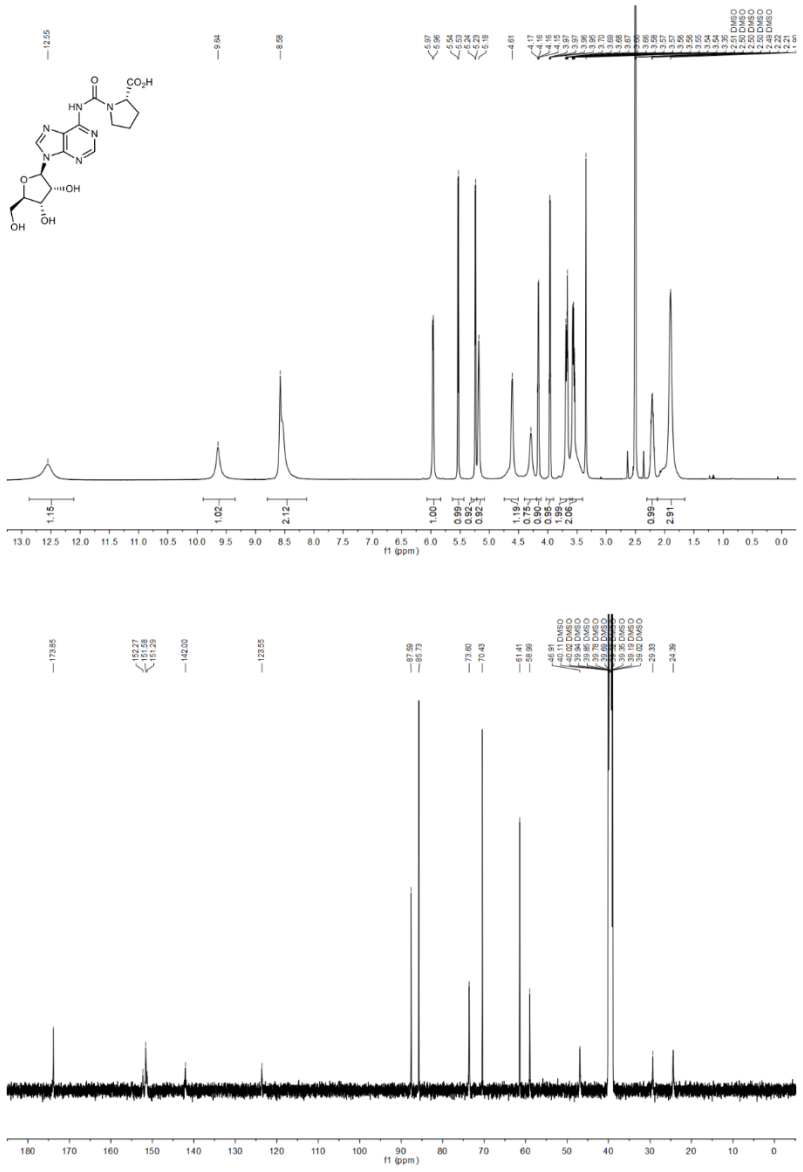




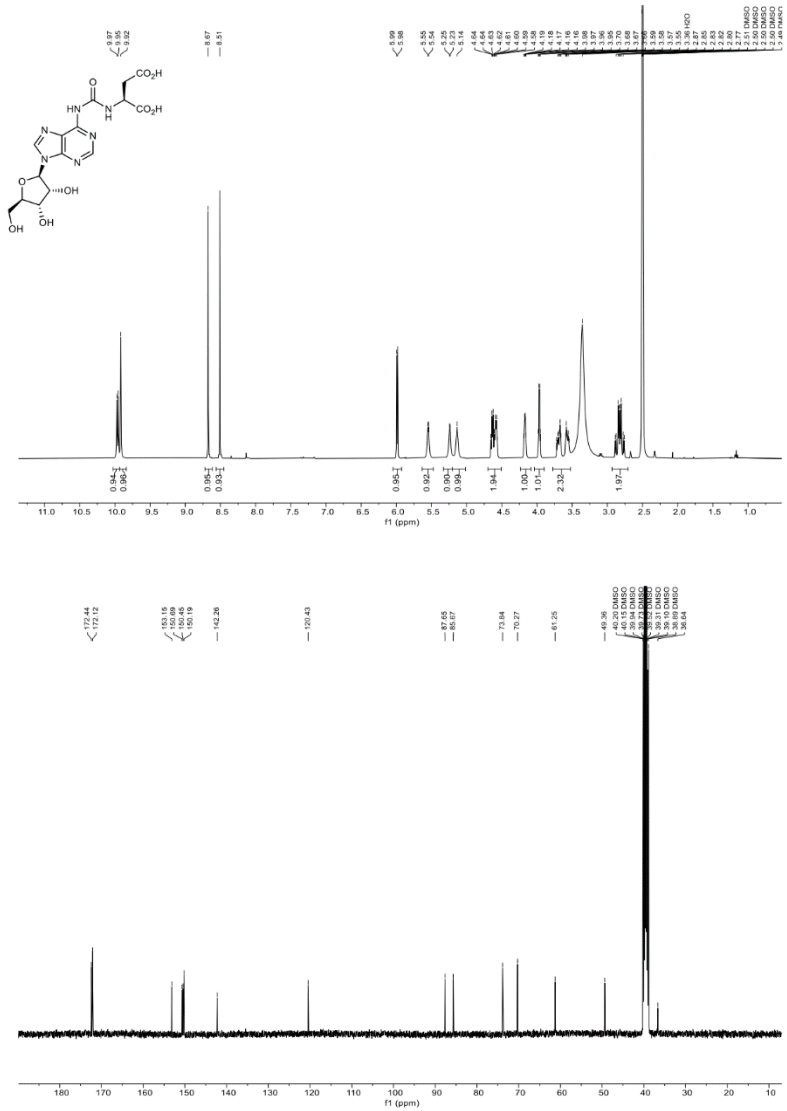
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1c



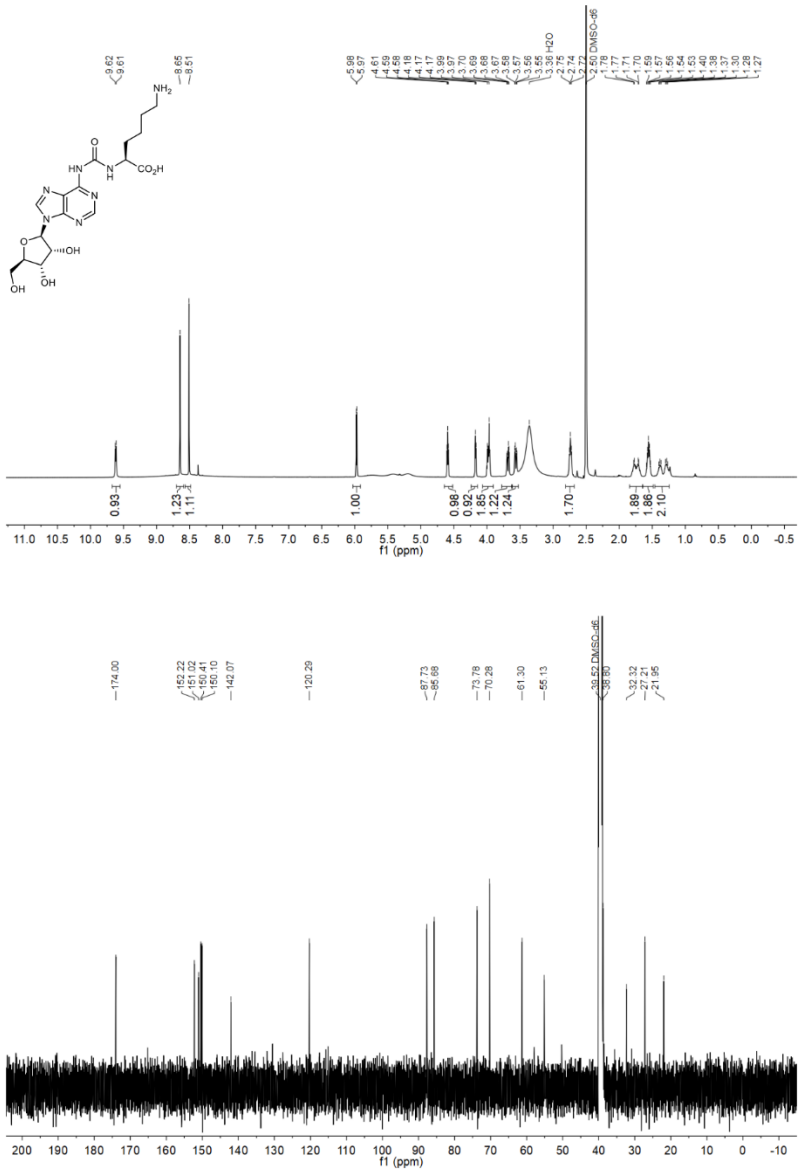
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1d



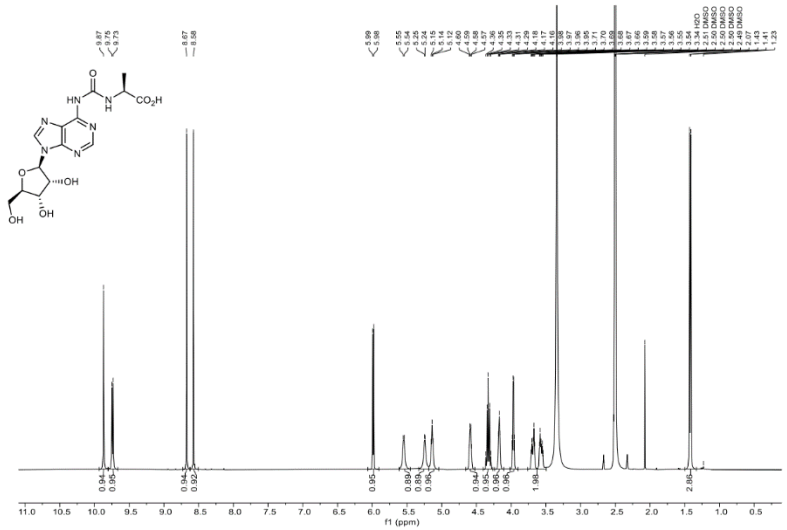
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1e

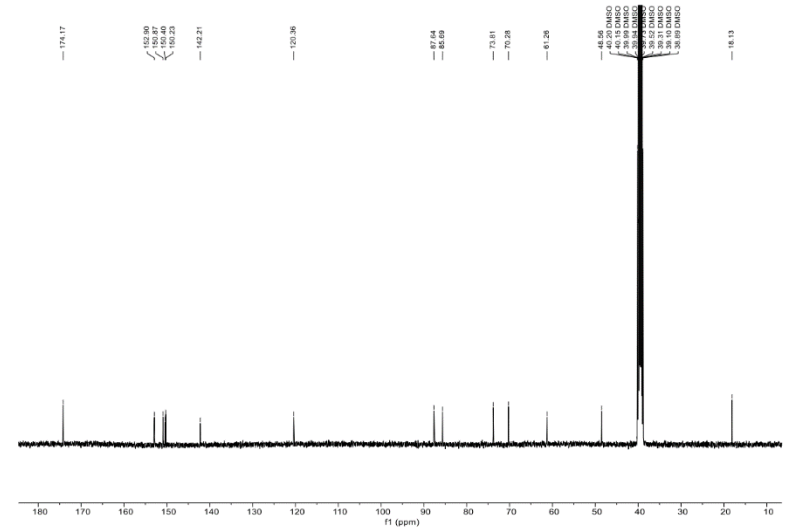


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound *α*-1f

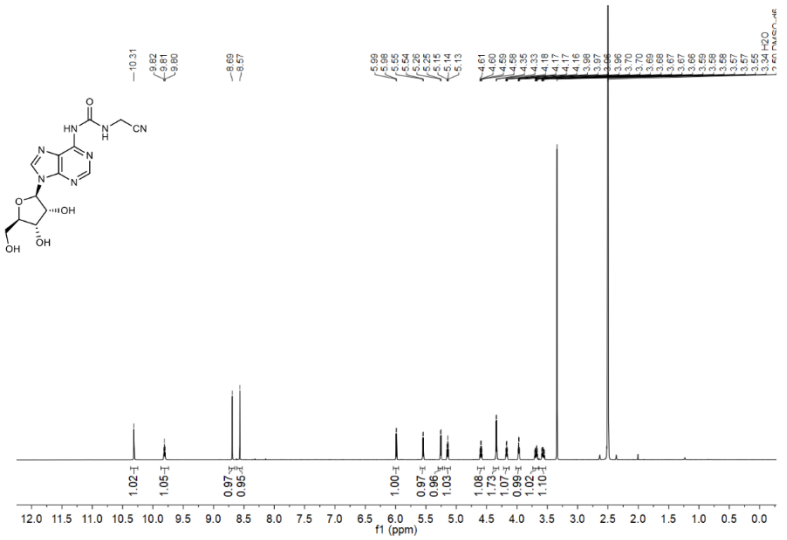


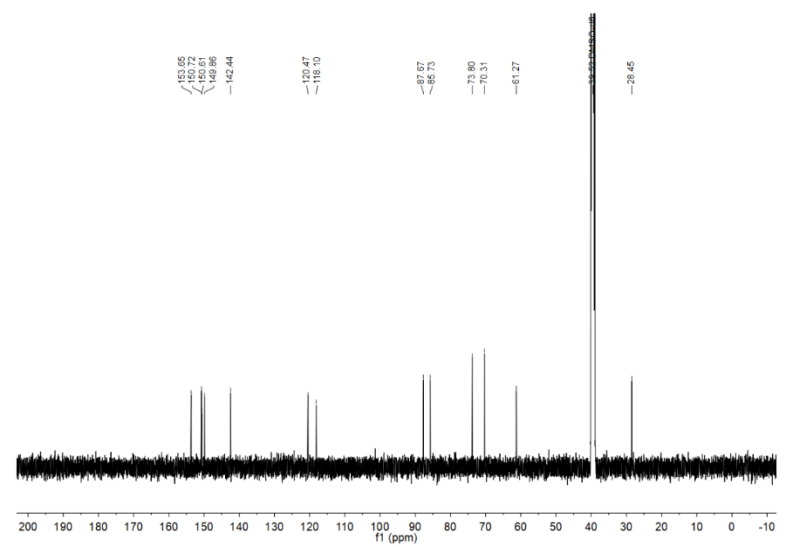
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1g



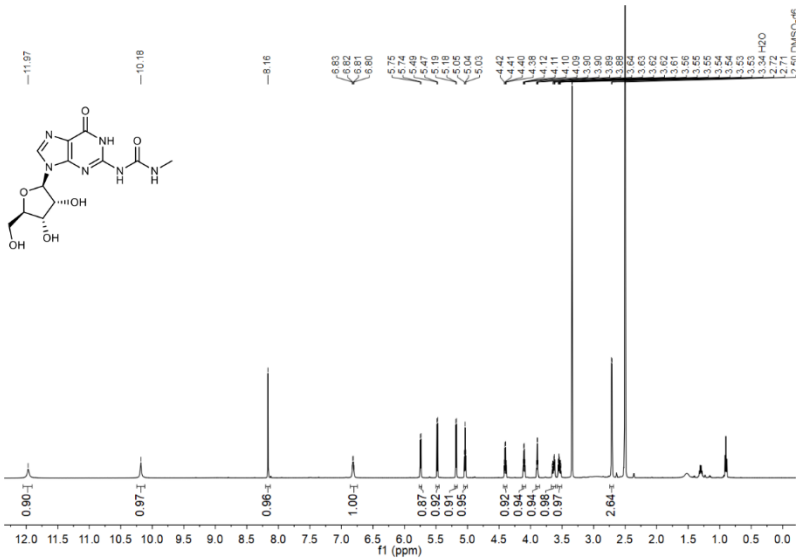


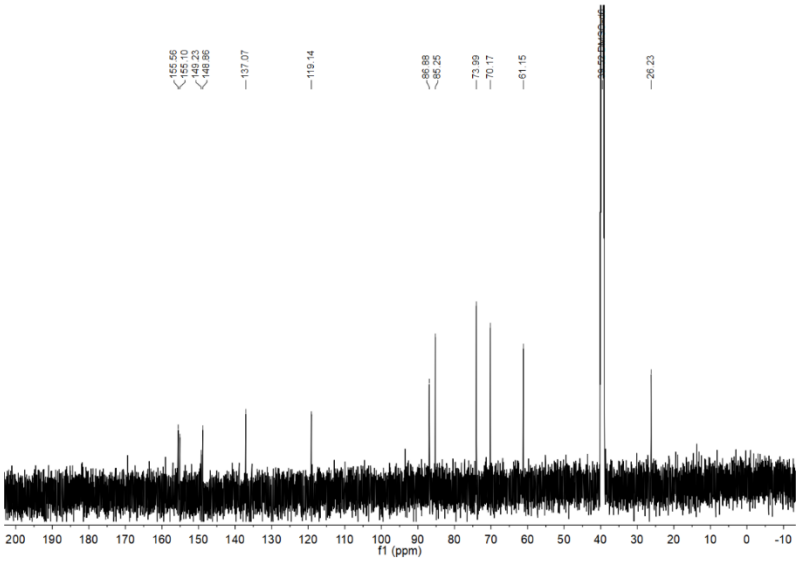
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 1h



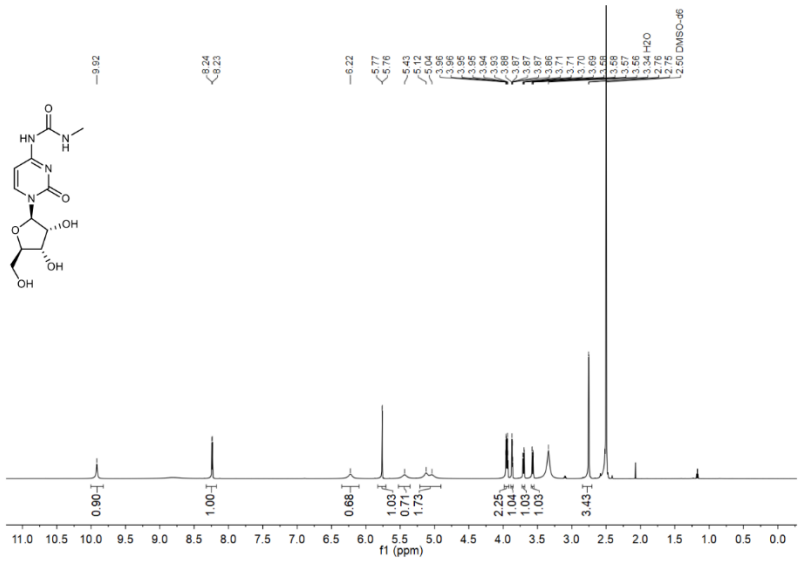


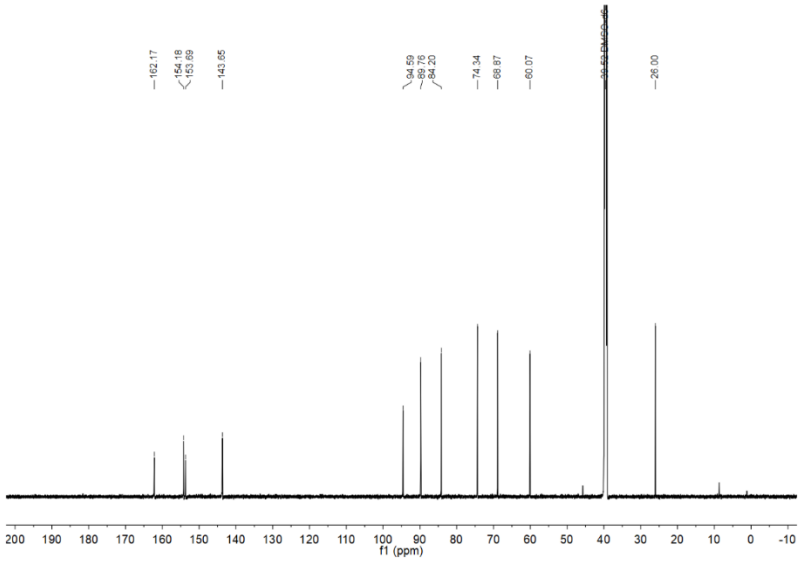
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 3



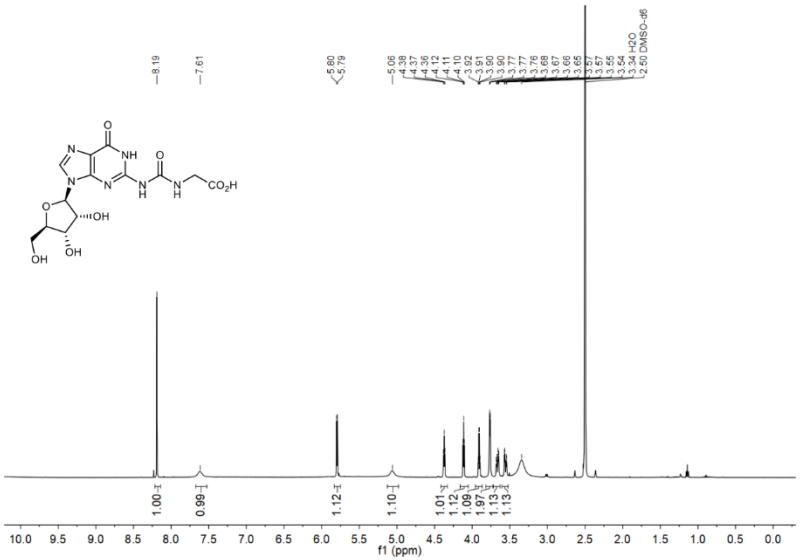


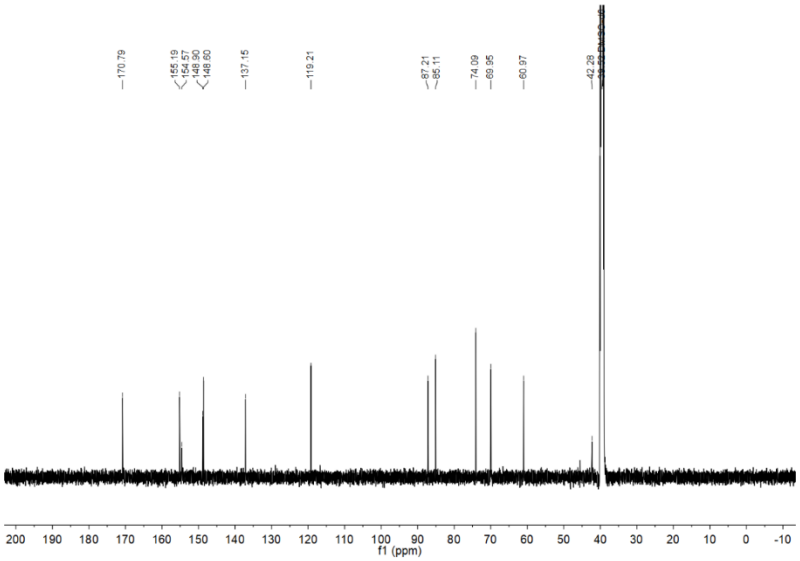
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 4



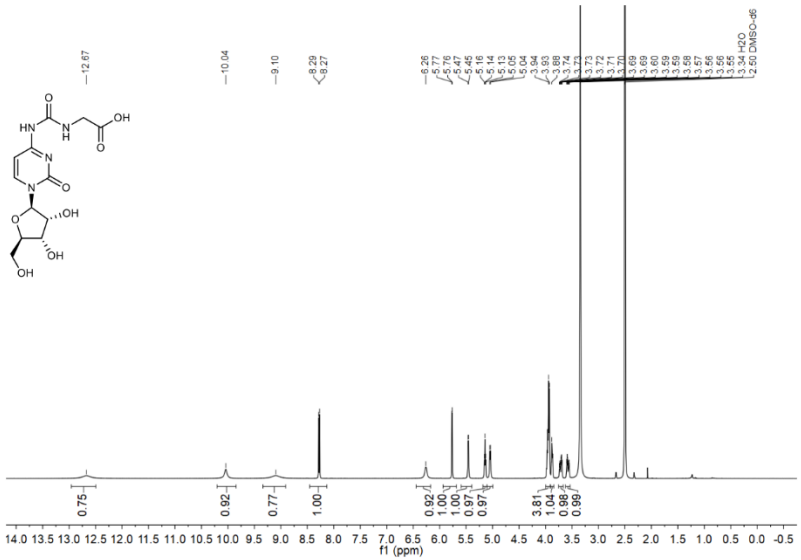


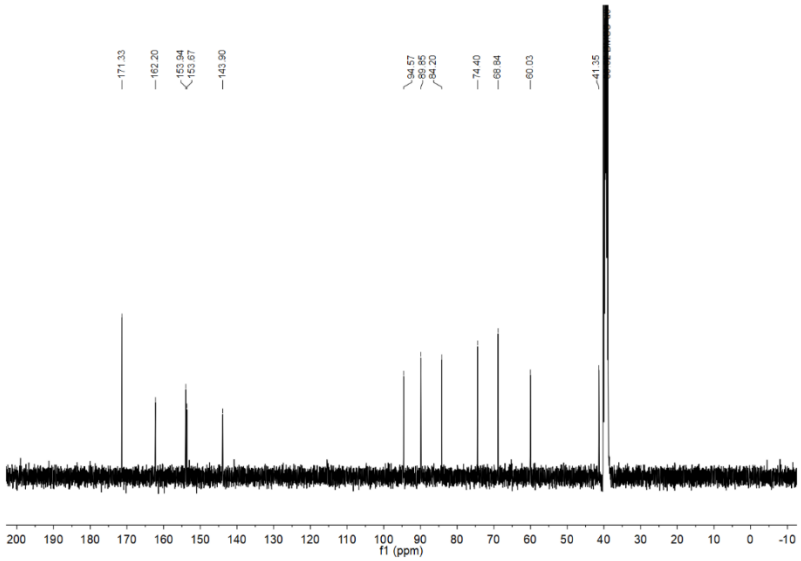
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 5



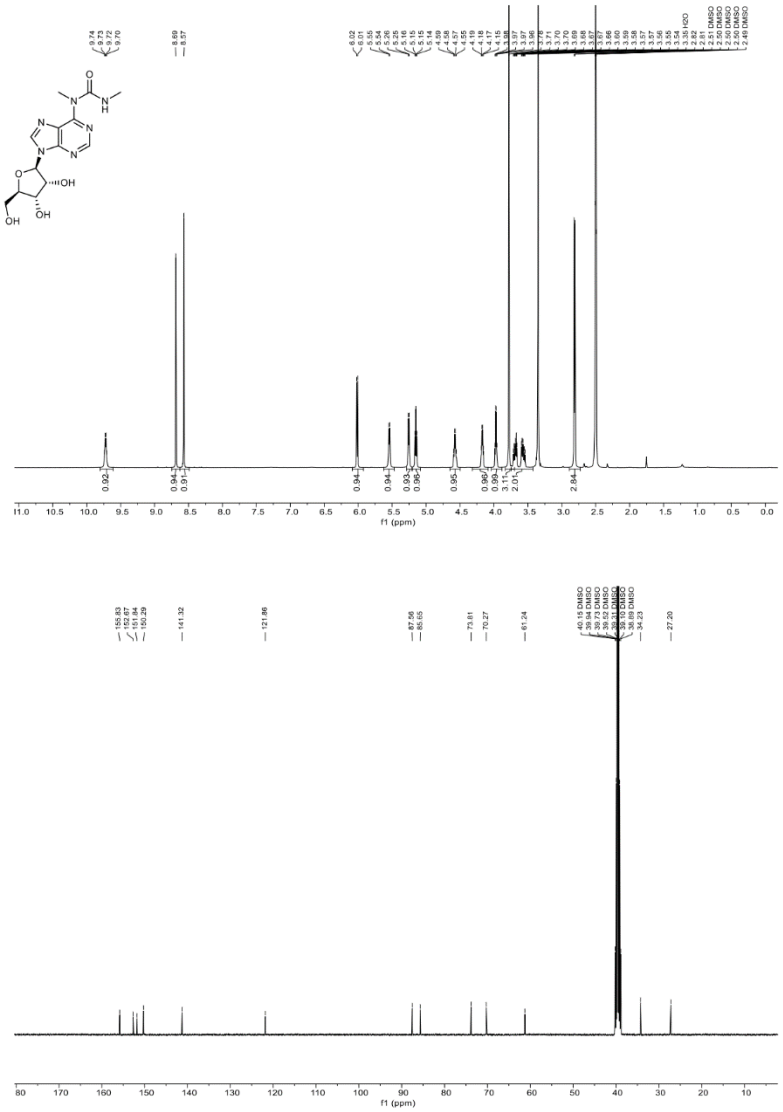


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 6

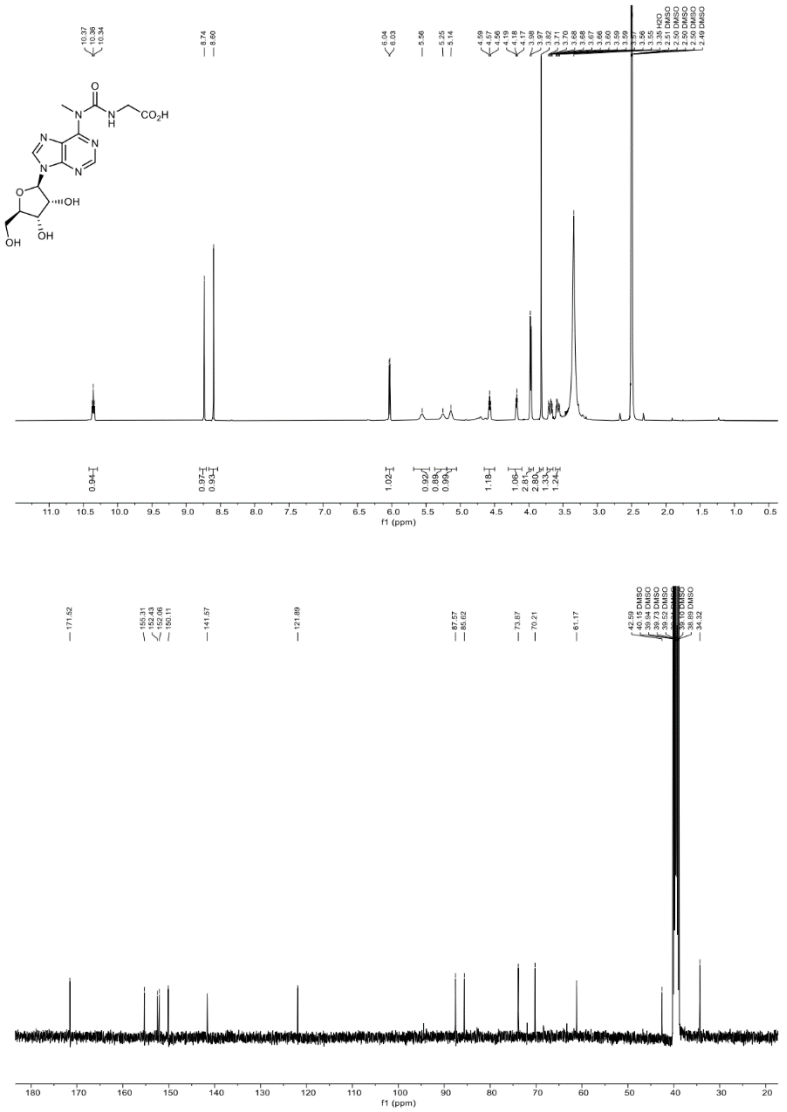




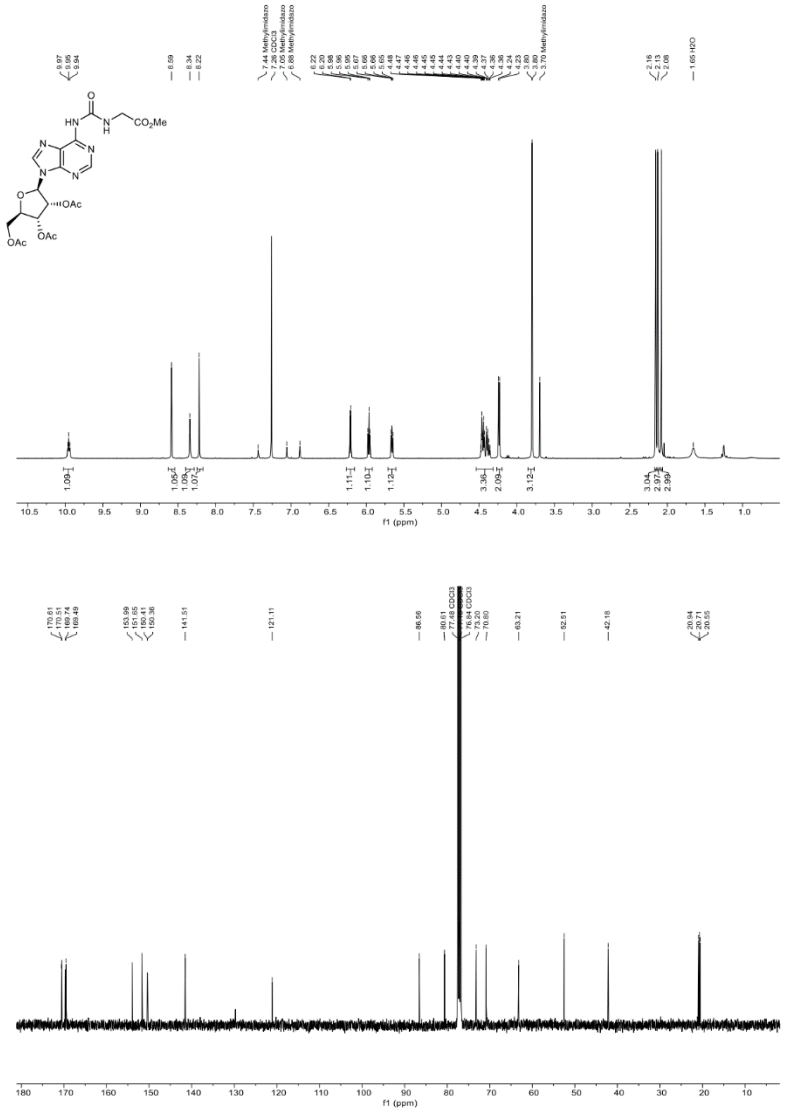
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 7



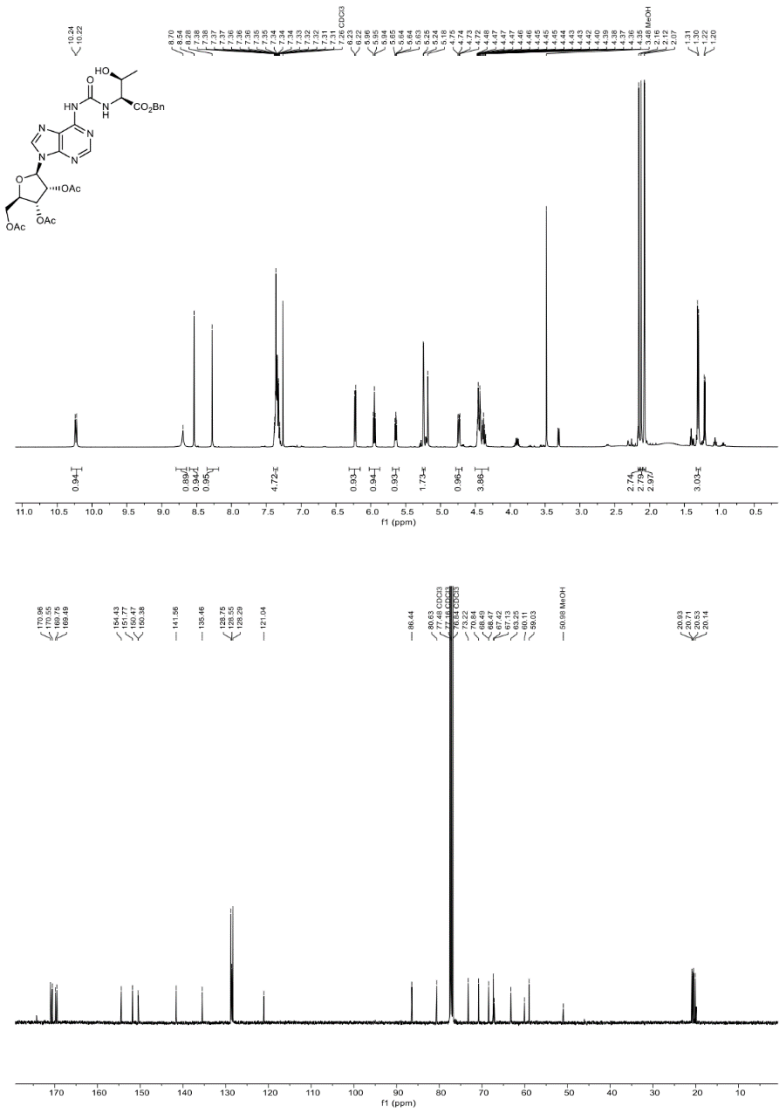
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 8



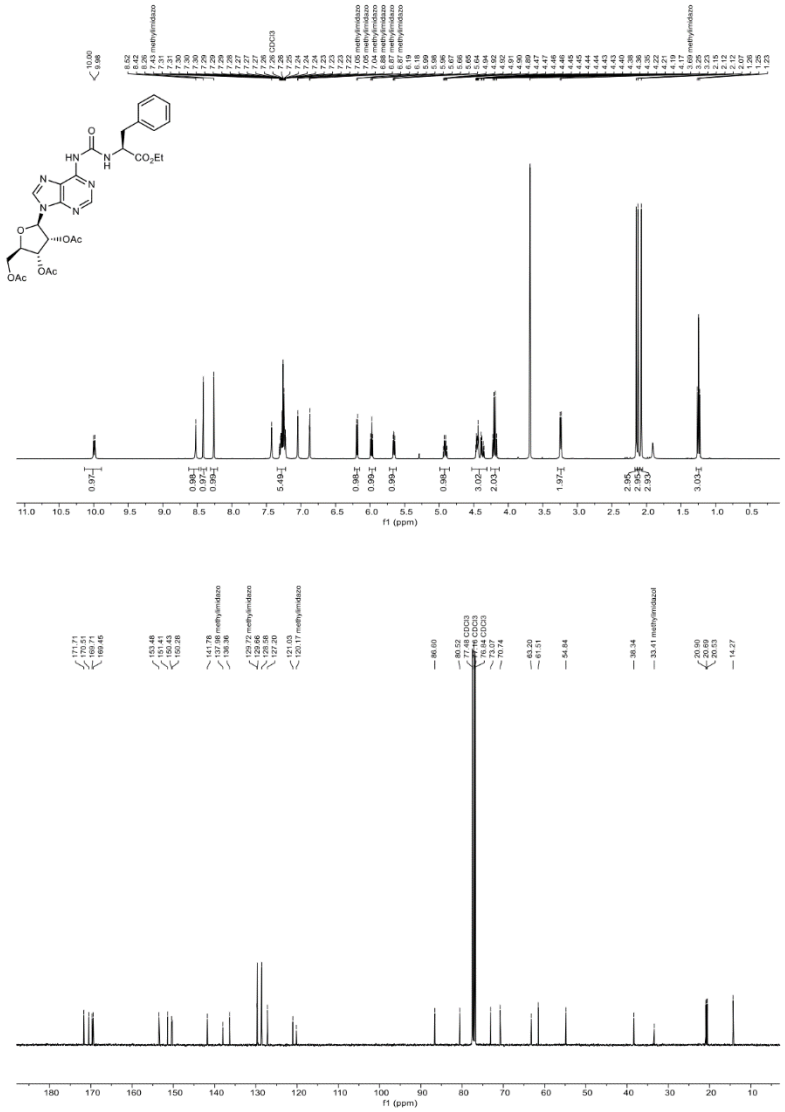
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13a



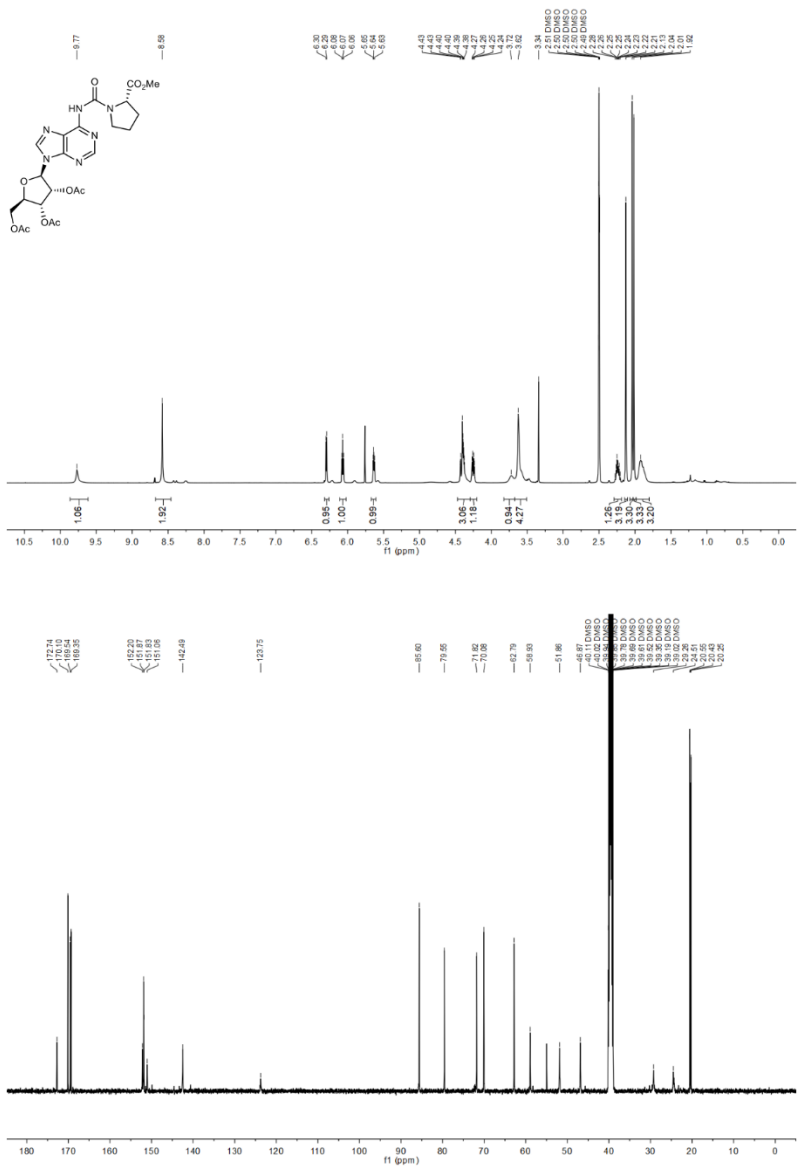
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13b



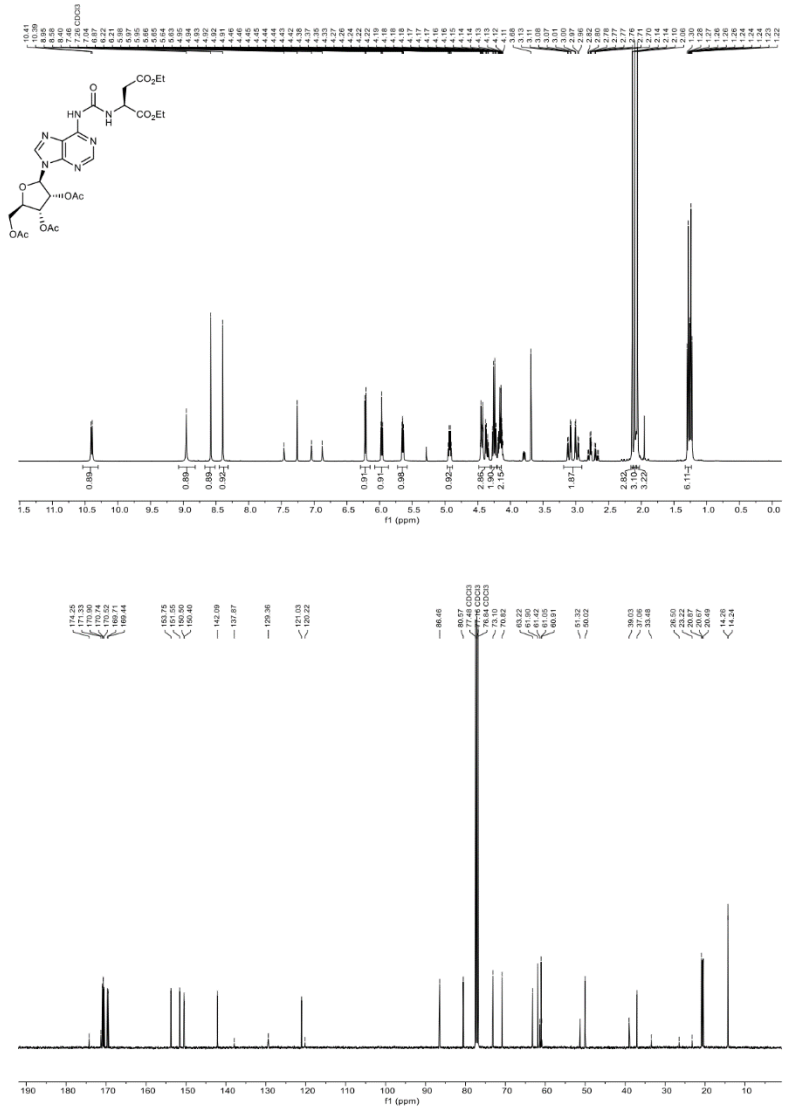
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13c



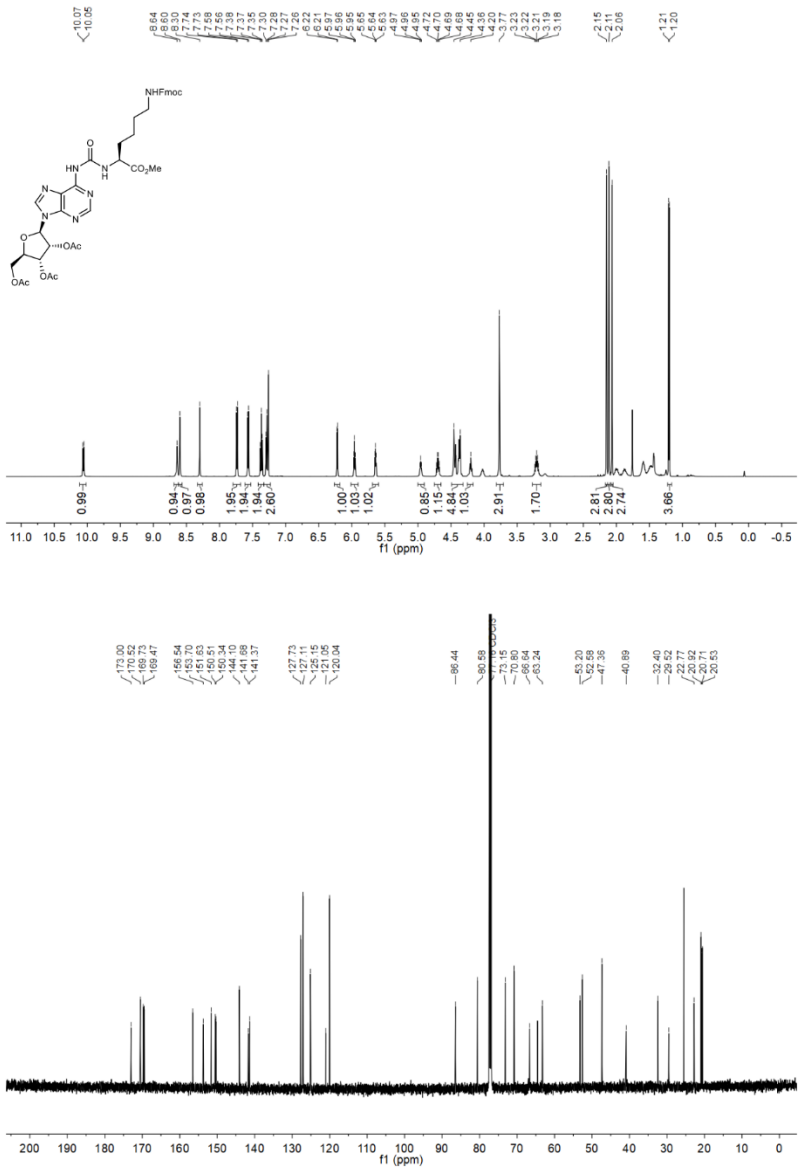
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13d



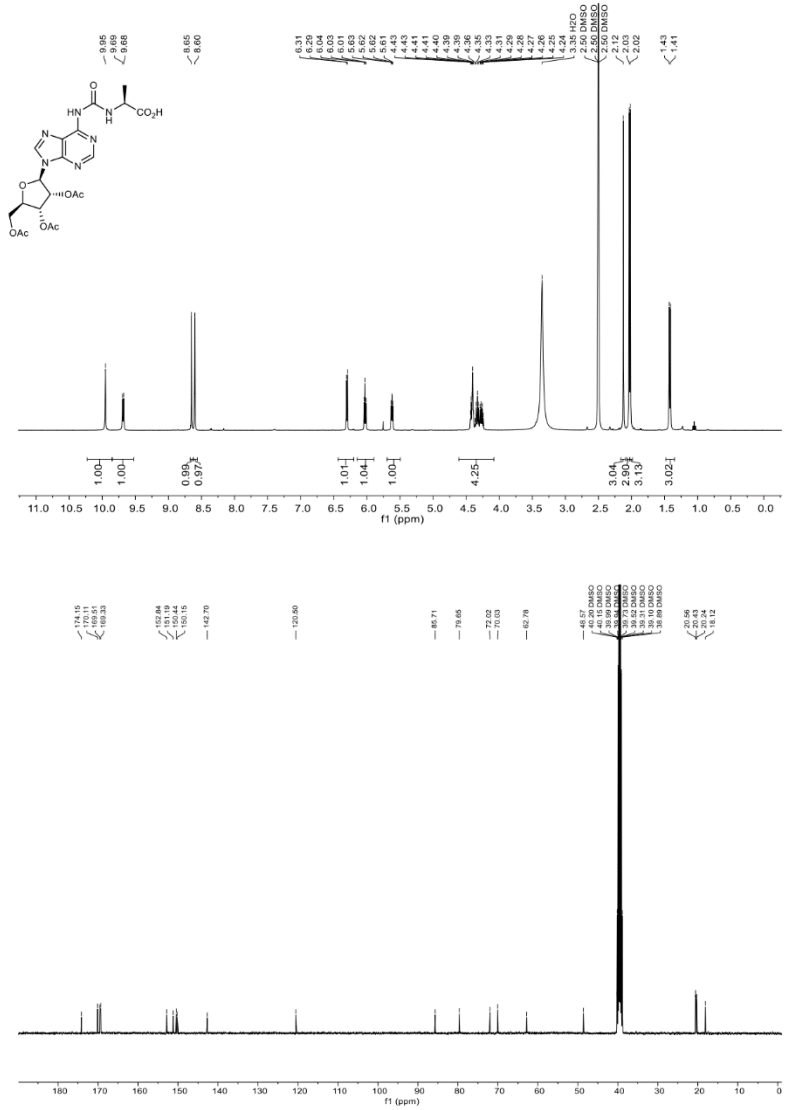
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13e



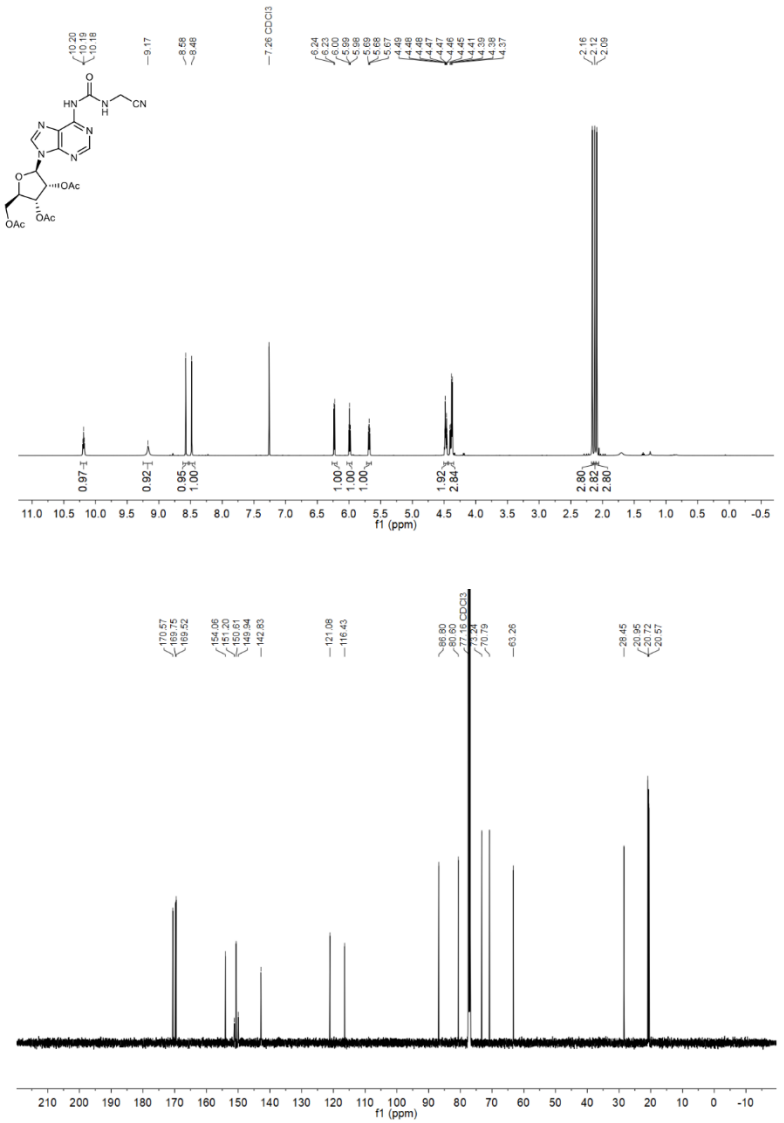
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13f



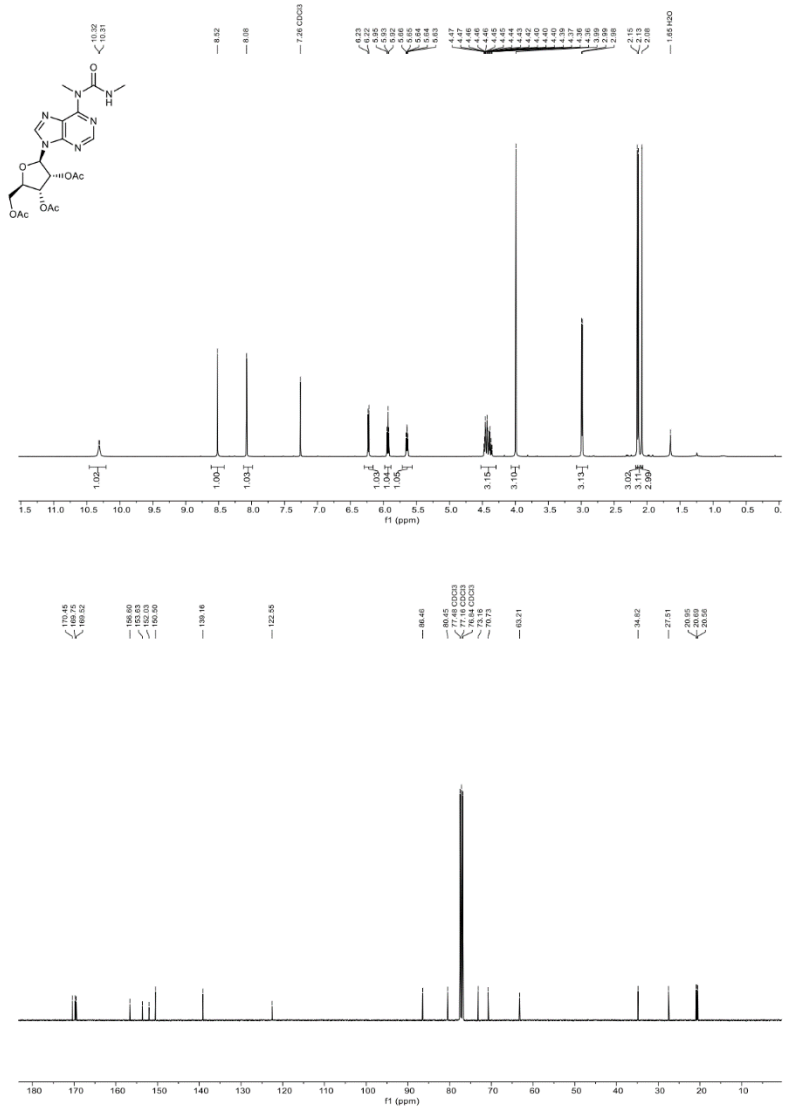
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13g



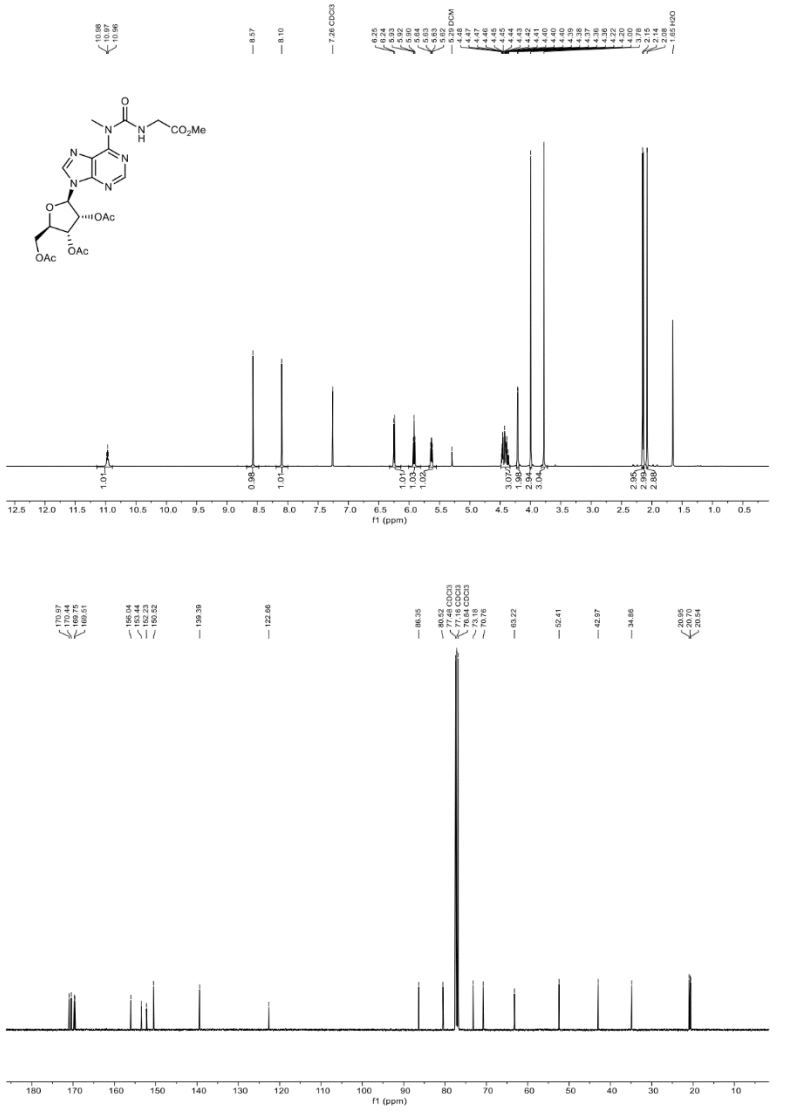
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 13h



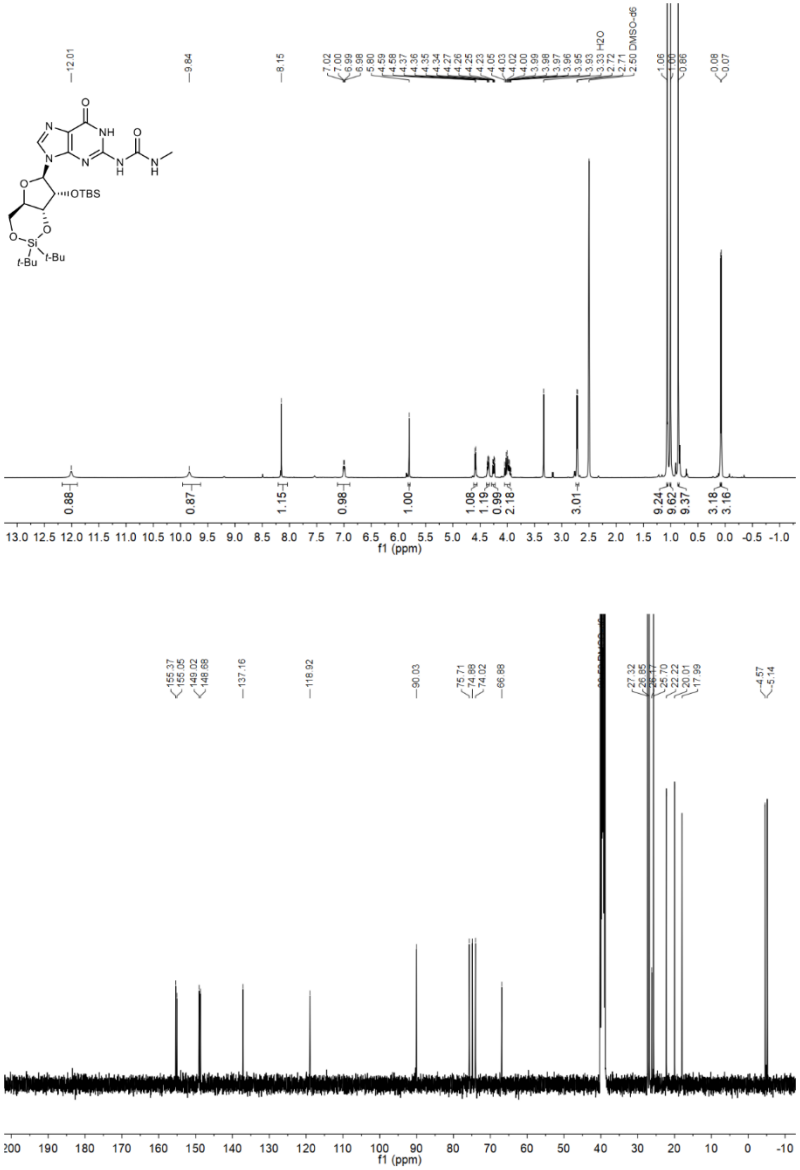
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 14



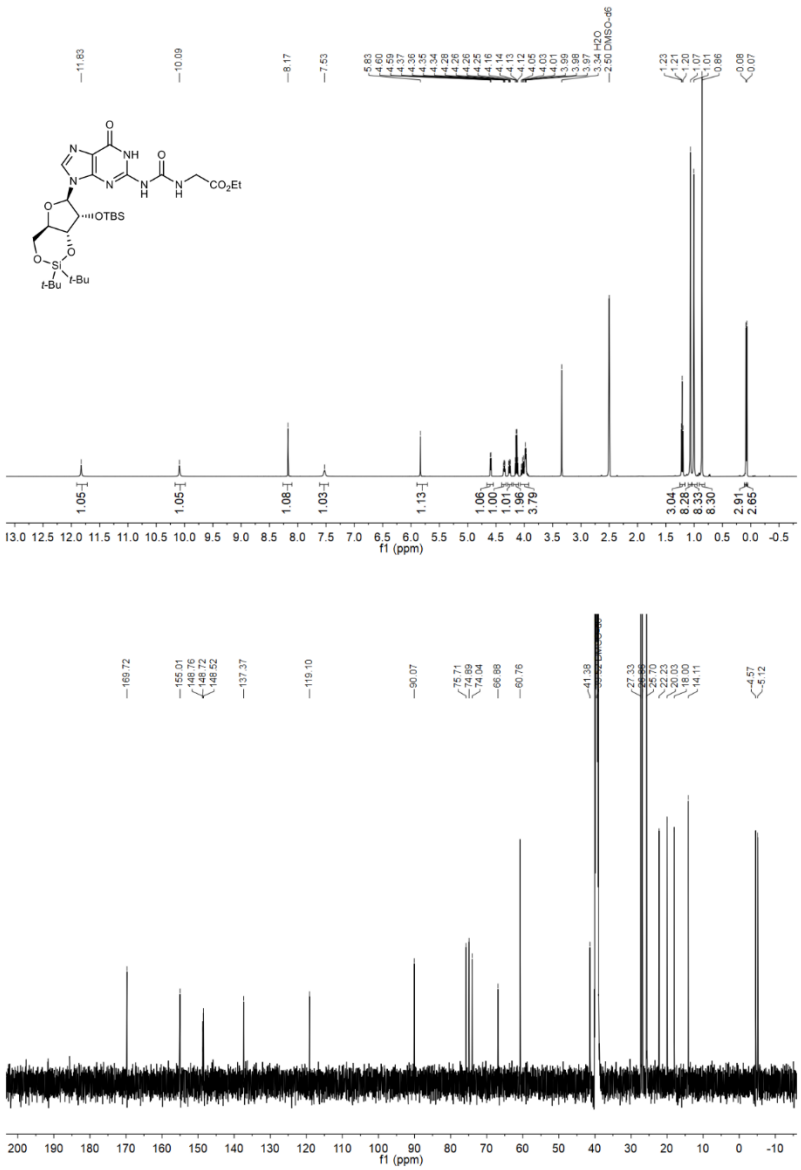
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 15



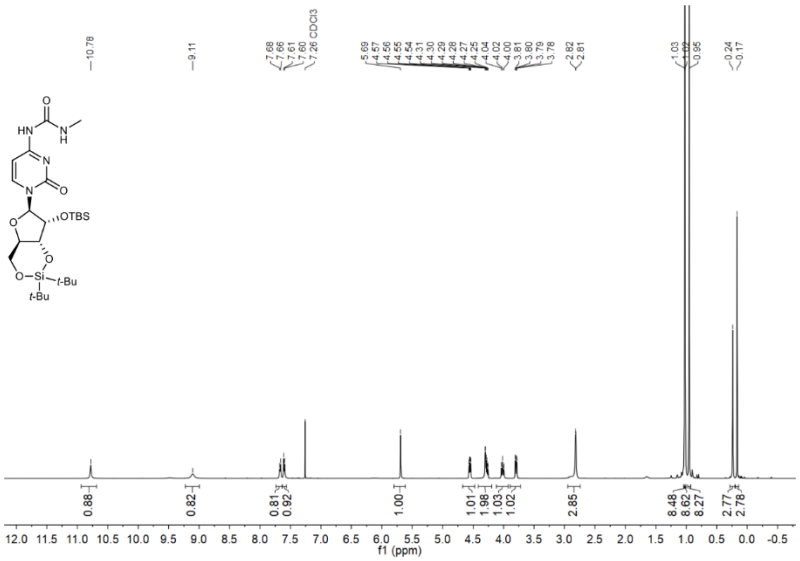
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 17

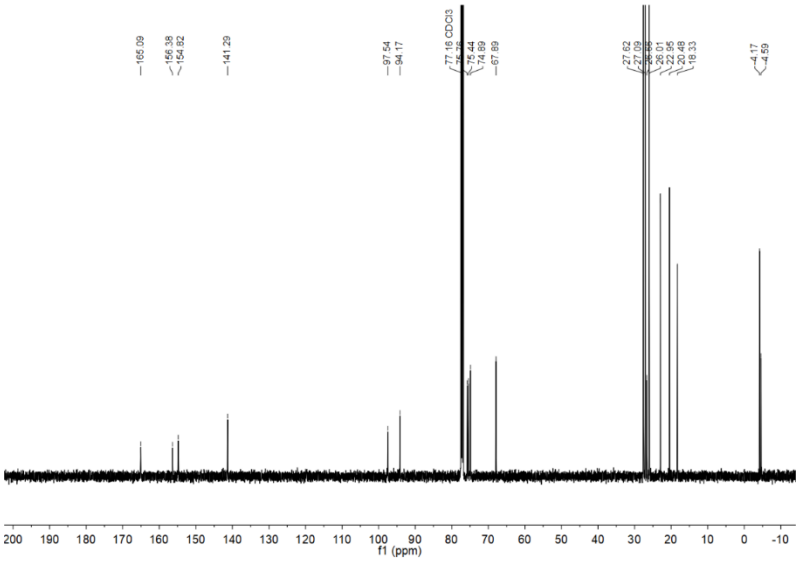


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 18

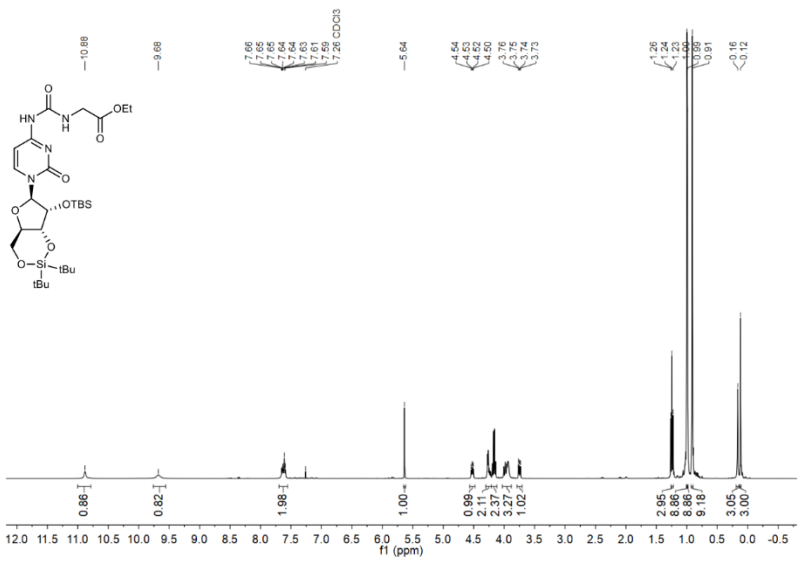


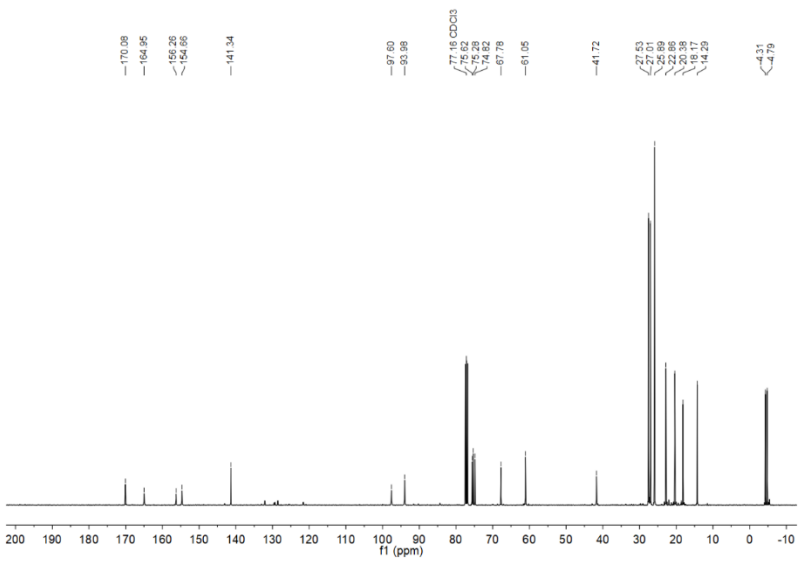
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 20



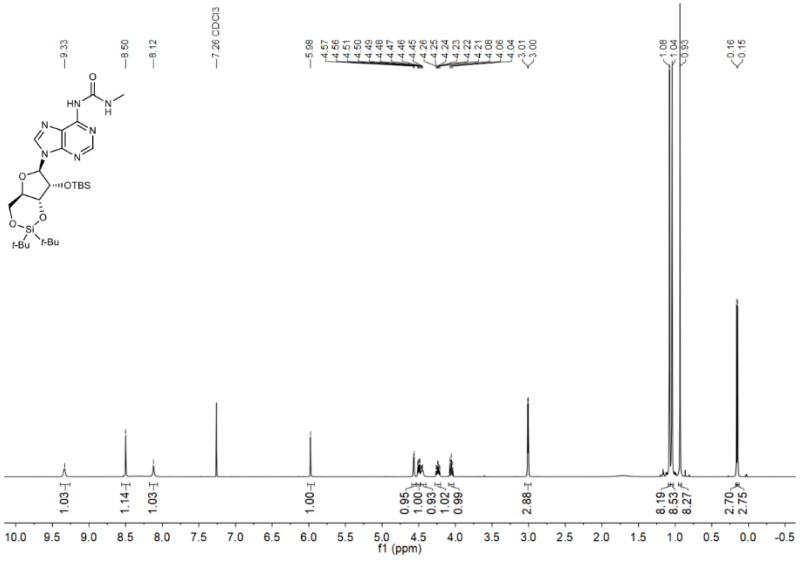


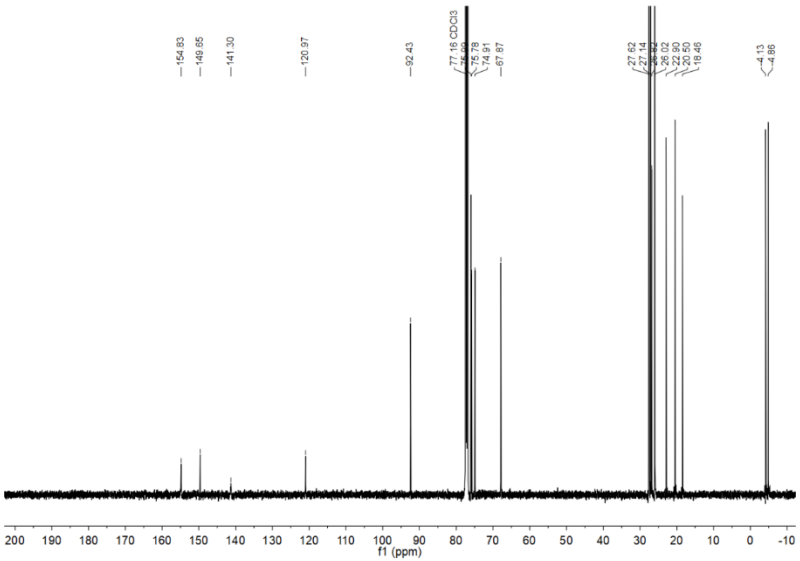
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 21



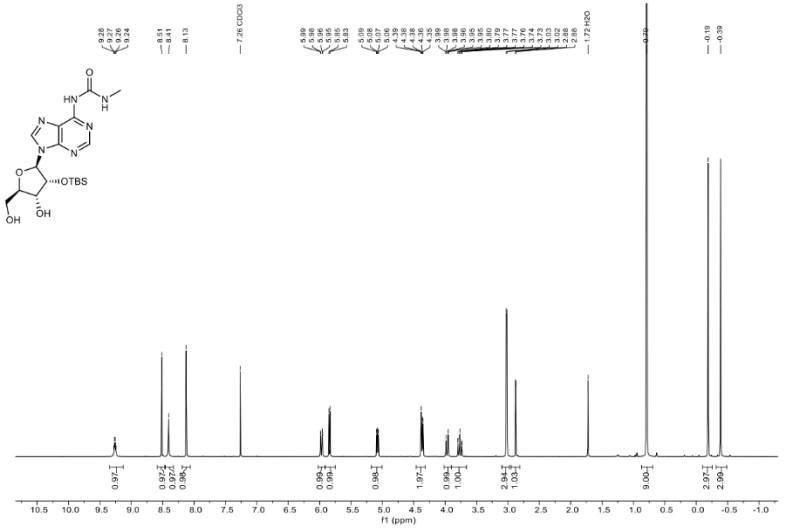


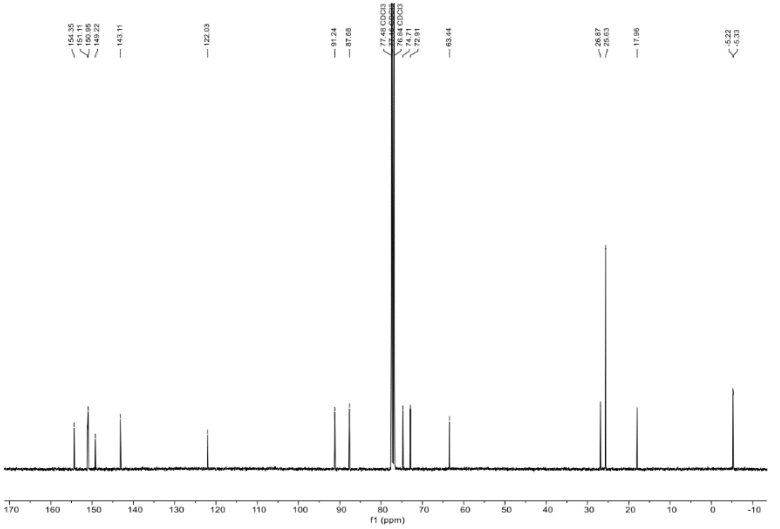
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 23



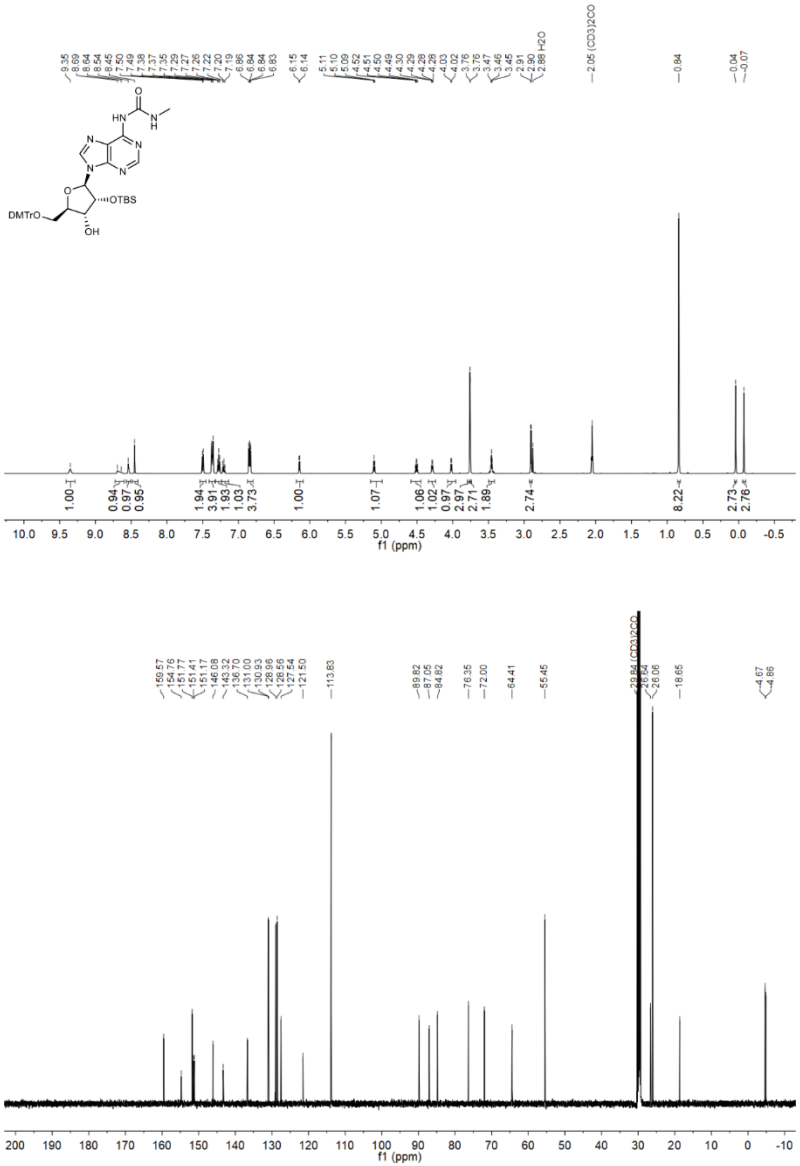


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 24

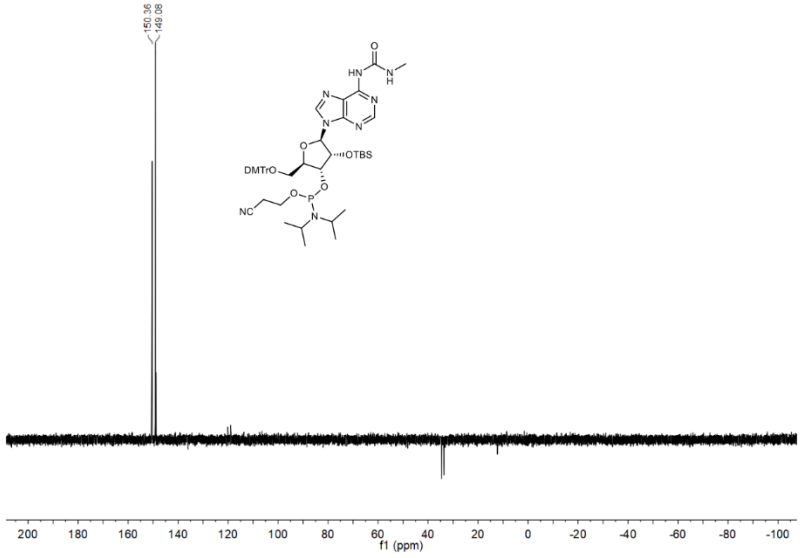




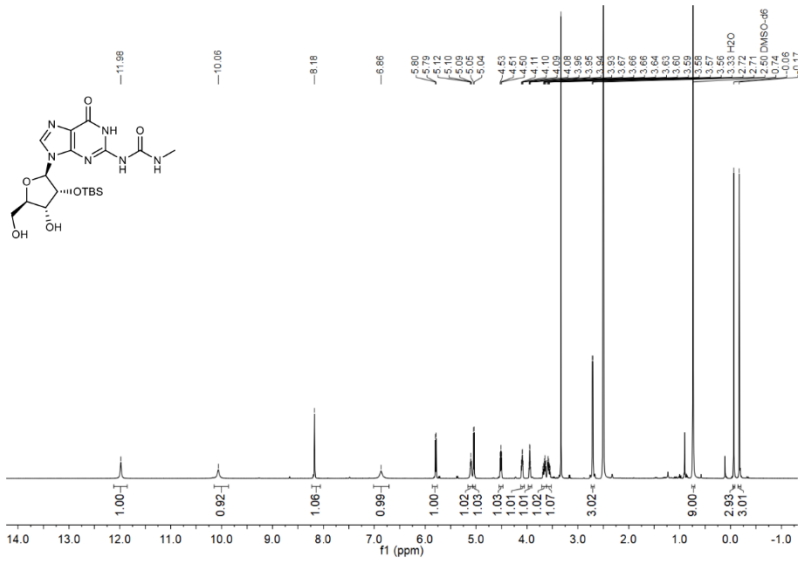
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 25

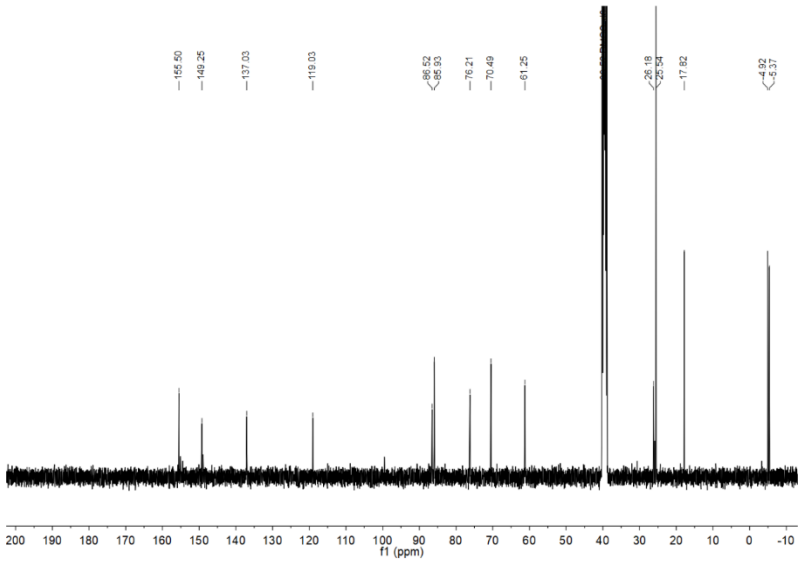


$^{31}P\{H\}$ NMR spectra of compound 26

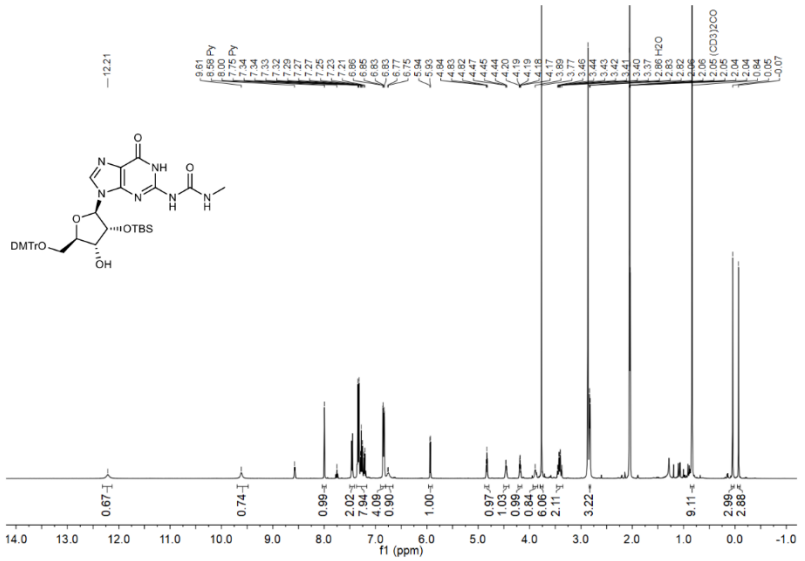


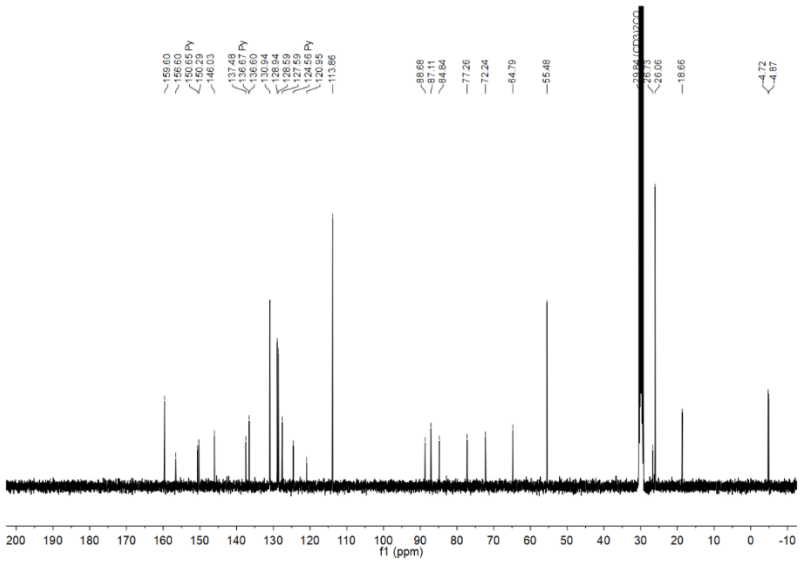
^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 27



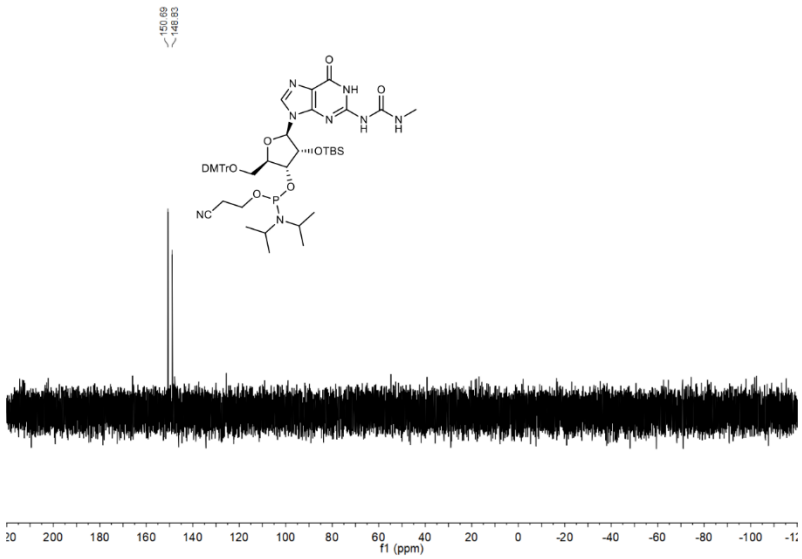


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 28

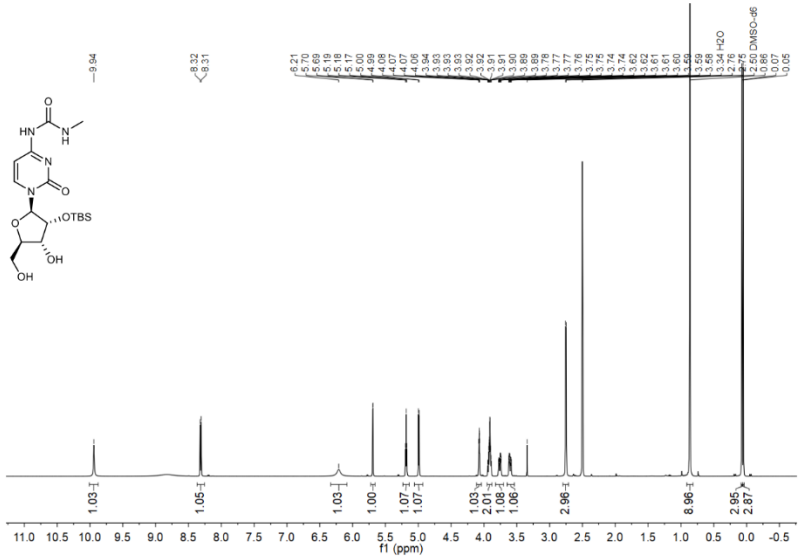


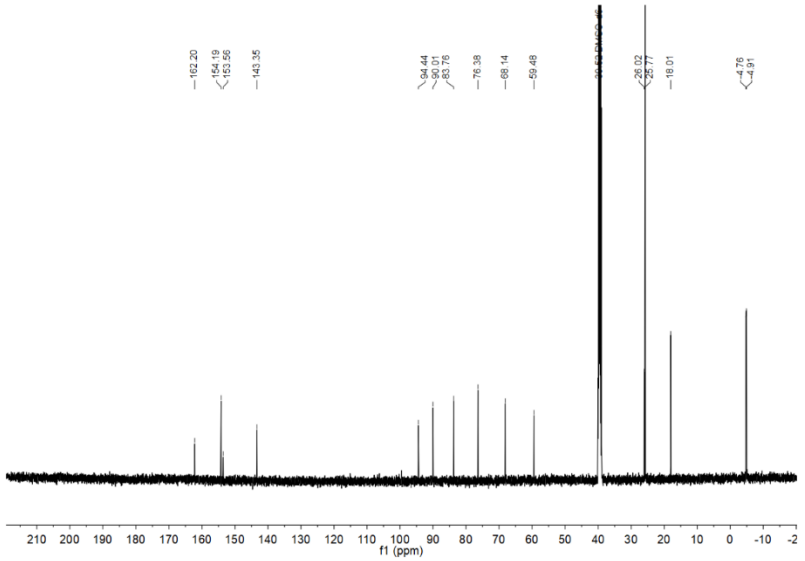


$^{31}P\{^{1}H\}$ NMR spectra of compound 29

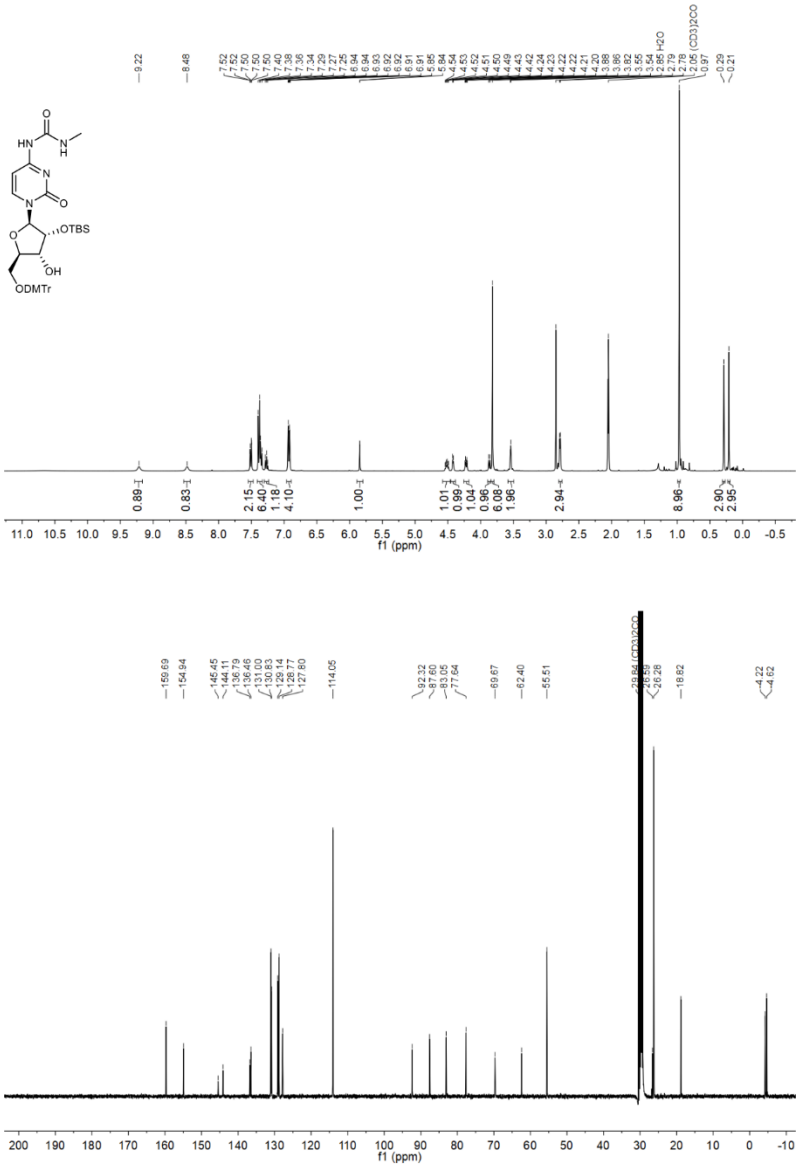


^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 30

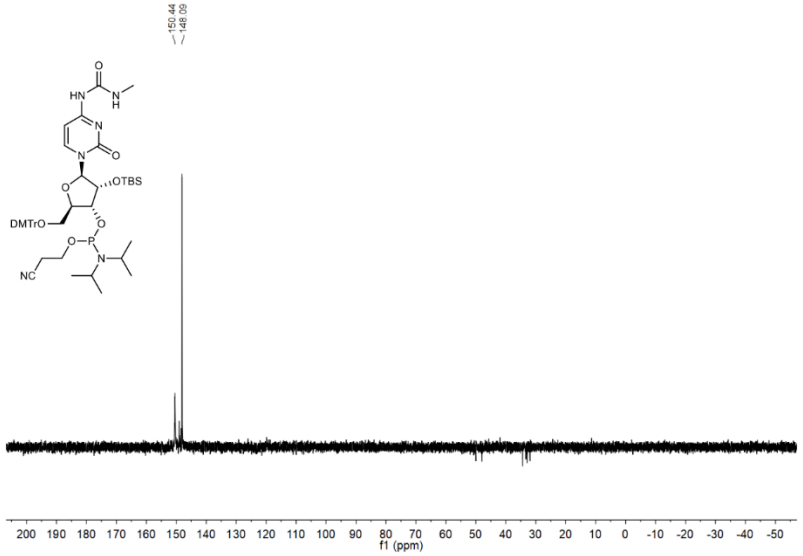




^{1}H and $^{13}C\{^{1}H\}$ NMR spectra of compound 31



$^{31}P\{^{1}H\}$ NMR spectra of compound 32



14. References

[1] G. R. Fulmer, A. J. M. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw, K. I. Goldberg, *Organometallics* **2010**, *29*, 2176-2179.

[2] G. W. Breton, M. Turlington, *Tetrahedron Lett.* **2014**, *55*, 4661-4663.

[3] F. Himmelsbach, B. S. Schulz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59-72.

[4] F. Ferreira, F. Morvan, *Nucleosides, Nucleotides, and Nucleic Acids* **2005**, *24*, 1009-1013.

[5] S. G. J. Mochrie, *Am. J. Phys.* **2011**, *79*, 1121-1126.

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

[7] F. Müller, L. Escobar, F. Xu, E. Węgrzyn, M. Nainytė, T. Amatov, C. Y. Chan, A. Pichler, T. Carell, *Nature* **2022**, *605*, 279-284.

[8] V. Serebryany, L. Beigelman, *Tetrahedron Lett.* **2002**, *43*, 1983-1985.