Dissertation at the Faculty of Biology of the Ludwig-Maximilians-Universität München

For the award of the degree
Doctor rerum naturalium

tRNA modifications as molecular fossils of a protocellular translation system

Chun-Yin Chan
aus
Hongkong

München, 2024

Diese Dissertation wurde angefertigt unter Leitung von Prof. Dr. Thomas Carell an der Fakultät für Biologie der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. Thomas Carell

Zweitgutachter: Prof. Dr. Kirsten Jung

Tag der Abgabe: 25.09.2024

Tag der mündlichen Prüfung: 11.02.2025

Eigenständigkeitserklärung

Hiermit versichere ich an Eides statt, dass die vorliegende Dissertation mit dem Titel

_____

von mir selbstständig verfasst wurde und dass keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, wurden in jedem Fall unter Angabe der Quellen (einschließlich des World Wide Web und anderer elektronischer Text- und Datensammlungen) kenntlich gemacht. Weiterhin wurden alle Teile der Arbeit, die mit Hilfe von Werkzeugen der künstlichen Intelligenz de novo generiert wurden, durch Fußnote/Anmerkung an den entsprechenden Stellen kenntlich gemacht und die verwendeten Werkzeuge der künstlichen Intelligenz gelistet. Die genutzten Prompts befinden sich im Anhang. Diese Erklärung gilt für alle in der Arbeit enthaltenen Texte, Graphiken, Zeichnungen, Kartenskizzen und bildliche Darstellungen.

München/23.09.2024

(Ort / Datum)

CHUN YIN, CHAN

(Vor und Nachname in Druckbuchstaben)

CHUN YIN, CHAN

(Unterschrift)

Affidavit

Herewith I certify under oath that I wrote the accompanying Dissertation myself.

Title: tRNA modifications as molecular fossils of a protocellular translation system

In the thesis no other sources and aids have been used than those indicated. The passages of the thesis that are taken in wording or meaning from other sources have been marked with an indication of the sources (including the World Wide Web and other electronic text and data collections). Furthermore, all parts of the thesis that were de novo generated with the help of artificial intelligence tools were identified by footnotes/annotations at the appropriate places and the artificial intelligence tools used were listed. The prompts used were listed in the appendix. This statement applies to all text, graphics, drawings, sketch maps, and pictorial representations contained in the Work.

Munich/23.09.2024

(Location/date)

CHUN YIN, CHAN

(First and last name in block letters)

CHUN YIN, CHAN

(Signature)

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 23.09.2024 CHUN YIN, CHAN

(Unterschrift)

Erklärung

Hiermit erkläre ich, *

☒ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

☒ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

☐ dass ich mich mit Erfolg der Doktorprüfung im Hauptfach .....

und in den Nebenfächern .....

bei der Fakultät für ..... der .....

(Hochschule/Universität)

unterzogen habe.

☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den 23.09.2024 CHUN YIN, CHAN

(Unterschrift)

*) Nichtzutreffendes streichen

Contributions

Results 2.1 – I participated in the organic synthesis of the modified RNA nucleoside phosphoramidites, the synthesis of their respective RNA oligonucleotides and UV-melting experiments.

Results 2.2 – I participated in the organic synthesis of the modified RNA nucleoside phosphoramidites, the synthesis of their respective RNA and 2'OMe RNA oligonucleotides, and the investigations in RNA and 2'OMe RNA template-directed peptide coupling reactions and urea cleavage reactions.

Results 2.3 – I designed and contributed to all of the described experiments.

Results 2.4 – I designed and contributed to all of the described experiments.

Results 2.5 – I designed and contributed to all of the described experiments.

Acknowledgements

I am expressing my sincere gratitude towards a number of people who provided me with their continual support and guidance throughout the journey of my doctoral study. First and foremost, I would like to thank my doctoral supervisor, Prof. Dr. Thomas Carell, for his consistent guidance and motivation. His enthusiasm in multiple aspects of science ranging from chemistry to biochemistry has continually inspired me and immersed me in multidisciplinary knowledge throughout the years. His research group consists of scientists with diversifying backgrounds and expertise allowed new ideas to sparkle and be discussed freely. Under this environment, I learnt to address scientific problems by designing experiments systematically, which slowly developed my scientific mindset to become an independent researcher.

My second thank goes to my other evaluators and thesis advisory committee members, who provided valuable comments and insights for my doctoral work. This includes Prof. Dr. Kirsten Jung, Prof. Dr. Joachim T. Haug, Prof. Dr. Thomas Nägele, Prof. Dr. Jörg Nickelsen, Prof. Dr. Claude Becker, Dr. Sabine Schneider (TAC) and Prof. Dr. Anja Hoffmann-Röder (TAC).

Next, I thank the Marie Skłodowska-Curie Actions who funded my doctoral study for 3 years. This would not be possible without the continuous effort invested by Prof. Dr. Andrew Kellette from Dublin City University, coordinator Marco Cavallaro from Accelopment and all of the other participating principal investigators in the consortium. My Ph.D. was hugely benefited from the program not only by its training programs, but also the network that it built with other early-stage researchers.

I would like to particularly thank my senior colleagues back then, Dr. Milda Nainytė, Dr. Felix Müller and Dr. Luis Escobar, even though they have already left the group. They provided me with a lot of aids in the early days of my study. When I first joined the group with limited synthetic experience, they were patient teachers who equipped me with the essential skills to become a successful Ph.D. candidate. My thank then goes to my direct colleague, Johannes Singer, who chose to join my project and worked closely with me since then in my final 2 years. Without his diligence and insights, our study on thiouridine-modified RNA could not have developed into the current state, which accounts for a major part of this dissertation. He is always a reliable and competent working partner.

A special thank goes to Dr. Philipp Holliger from the MRC-LMB, Cambridge, who hosted me as a visiting Ph.D. student in his lab for three months. The secondment significantly widened my horizon in the fields of molecular biology and biotechnology, during which I managed to meet top-notch scientists and learnt plenty of cutting-edge ideas. I also extend my thank to his team and particularly to Dr. Edoardo Gianni, who is a passionate scientist and provided solid mentoring to me during my stay in the Holliger's group. Together, we engaged in numerous meaningful scientific discussions and designed the ribozyme experiments described in this thesis.

I would like to especially thank Dr. Luis Escobar, Dr. Edoardo Gianni and Dr. Felix Müller again, for their valuable time invested in going through the draft of my thesis. Their comments and suggested corrections improved the coherence and clarity of the thesis significantly.

A big thank goes to the permanent staffs in the Carell group, especially Dr. Markus Müller, Kerstin Kurz, Dr. Nädä Raddaoui and Frau Slava Gärtner, who handled most of the administrative and routine works in the secretary office and in the lab. In particular, I would like to thank the extra efforts that they put while we were moving the lab from Haus F to ICEM. Without them, the Carell group would not be running.

I would also like to thank Dr. Céline Douat from the Huc's research group for synthesizing and characterizing the D-Pro-Pro-Glu-NH_2 peptide, and Dr. Benjamin Tast from the LMU biomedical center core facility for his assistance in collecting and analyzing the flow cytometry data.

I would like to thank Prof. Dr. Christof Osman, Nadine Hamze and the student council in organizing and managing the graduate school of life science Munich (LSM). I enjoyed a lot being a Ph.D. student in the graduate school, with well-designed program structure, social events and annual retreats.

Next, I would also like to thank all amazing people that I met in the group and, more generally, during my stay in Germany. Particularly to my closest colleagues and friends that were not mentioned above – Dr. Katerina Pappa, Johann de Graaff, Dr. Giacomo Ganazzoli, Ewa Węgrzyn, Dr. Eva Schönegger, Elsa Peeve, Alexander Pichler, Andreas Reichl, Yasmin Gärtner, Stylianos Xefteris, Fabian Hernichel, Dr. Dilara Özdemir, Dr. Corinna Pleintinger, Dr. Anna-Lena Halbritter, Felix Xu, Annika Tölke, Dr. Ivana Mejdrova, Erik Boinowitz, Nikolai Diukarev, Kathrin Halter, Tobias Kernmayr, Szymon Drewniak, Marion Wierl, and many others fantastic people that I met during my doctoral study. I will always remember these few years in my life, with all of the conference trips that we travelled, all of the laughters from our stupid jokes and all of the meaningful discussions that we had. For a person who had never imagined moving out of Hong Kong before my Ph.D., you all made my stay in Germany colorful and unforgettable. It was a grace to meet all of you in my life.

Last but not least, I am beyond words to express my gratitude to my ever-supportive family, and friends that I grew up with in Hong Kong. My parents and sister are always supportive and reassuring in no matter what decision I made, which gave me the courage and motivation to face my daily challenges. A special thank goes to Ms. Mei Sin Li (aka Shirley Li), who is the best listener I ever know and spent countless hours with me on video calls during my darkest time. I also thank friends that I grew up with in the church, who provided their mental support throughout the years. A special thank to Hoi Kiu Wong (aka Katie Wong) and Yan Kin Lee (aka Victor Lee), who I already knew in Hong Kong and happened to study also in the LMU. They provided valuable advice and helps in my arrival in Munich. Finally, I have to thank my bachelor thesis supervisor, Prof. Dr. Chun Kit Kwok, who was the first to convince me to pursue a Ph.D. degree, introduced me to Thomas, my doctoral supervisor, and continued to give me career advice from time to time, even after I left his group.

「天下之至柔，馳騁天下之至堅。」
－《道德經》老子
"The softest thing in the world overcomes the hardest thing in the world."
- Laozi (teacher of Confucius)

List of publications

Publications (and preprint) resulting from this doctoral work:

1. Chan, C.-Y. (co-first), Singer, J. & Carell, T. Membrane bound geranylated RNAs establish a primitive peptide synthesis system. *bioRxiv*, 2024.2008.2002.606298 (2024). DOI: 10.1101/2024.08.02.606298.
2. Müller, F., Escobar, L., Xu, F., Węgrzyn, E., Nainytė, M., Amatov, T., Chan, C.-Y., Pichler, A. & Carell, T. A prebiotically plausible scenario of an RNA–peptide world. *Nature* **605**, 279-284 (2022).
3. Nainytė, M., Müller, F., Ganazzoli, G., Chan, C.-Y., Crisp, A., Globish, D. & Carell, T. Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. *Chem. Eur. J.* **26**, 14856-14860 (2020).

Other publications:

4. Chen, X., Xue, G., Zhao, J., Zhang, Y., Zhang, S., Wang, W., Li, Y., Yuan, J., He, L., Chan, C.-Y., Liu, Y., Chen, W., Zhao, Y., Hu, P., Sun, H., Kwok, C.K. & Wang, H. *Lockd* promotes myoblast proliferation and muscle regeneration via binding with DHX36 to facilitate 5′ UTR rG4 unwinding and *Anp32e* translation. *Cell Rep.* **39**, 2211-1247 (2022)
5. Umar, M.I., Chan, C.-Y. (co-first) & Kwok, C.K. Developemnt of RNA G-quadruplex (rG4)-targeting *L*-RNA aptamers by rG4-SELEX. *Nat. Protoc.* **17**, 1385-1414 (2022)
6. Zhang, Y., Wu, Y., Zheng, H., Xi, H., Ye, T., Chan, C.-Y. & Kwok, C.K. Proteomic and Transcriptome Profiling of G-quadruplex Aptamers Developed for Cell Internalization. *Anal. Chem.* **93**, 5744-5753 (2021)
7. Chan, C.-Y. (first) & Kwok, C.K. Specific Binding of a *D*-RNA G-Quadruplex Structure with an L-RNA Aptamer. *Angew. Chem. Int. Ed.* **59**, 5293-5297 (2020)
8. Chan, C.-Y. (co-first), Umar, M.I., & Kwok, C.K. Spectroscopic analysis reveals the effect of a single nucleotide bulge on G-quadruplex structures. *Chem. Commun.* **55**, 2616-2619 (2019)
9. Lyu, K., Chen, S.-B., Chan, C.-Y., Tan, J.-H. & Kwok, C.K. Structural analysis and cellular visualization of APP RNA G-quadruplex. *Chem. Sci.* **10**, 11095-11102 (2019)
10. Umar, M.I., Ji, D., Chan, C.-Y. (co-first) & Kwok, C.K. G-Quadruplex-Based Fluorescent Turn-On Ligands and Aptamers: From Development to Applications. *Molecules* **24**, 2416 (2019)
11. Chan, K.L., Peng, B., Umar, M.I., Chan, C.-Y., Sahakyan, A.B., Le, M. & Kwok, C.K. Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem. Commun.* **54**, 10878-10881 (2018)
12. Ma, L., Lin, X., Li, C., Xu, Z., Chan, C.-Y., Tse, M.-K., Shi, P. & Zhu, G. A Cancer Cell-Selective and Low-Toxic Bifunctional Heterdinuclear Pt(IV)-Ru(II) Anticancer Prodrug. *Inorg. Chem.* **57**, 2917-2924 (2018)

Patent filed:

13. Kwok, C.K., & Chan, C.-Y. (co-inventor) Method of Producing An Aptamer And Uses Thereof. Patent No.: US11,326,169., May 10, 2022

Fellowship

2020-2023 Early Stage Researcher of Marie Skłodowska-Curie Actions, grant agreement No. 861381 (Nature-ETN)

Secondment during this doctoral work

2022 Research group of Dr. Philipp Holliger, MRC-LMB, Cambridge, UK

Public presentations during this doctoral work

2024 Oral presentation. 8th Cambridge Symposium on Nucleic Acids Chemistry and Biology 2024. Cambridge, United Kingdom

2023 Poster presentation. Origin of Life Donostia Meeting (OLDM) 2023. Donostia, Spain

2022 Poster presentation. Bioinspired Complex Systems from Basic Science to Practical Applications. Neve Ilan, Israel

2022 Poster presentation. XVIIIth Symposium on Chemistry of Nucleic Acid Components (SCNAC), Český Krumlov, Czech Republic

Dedicated to my family and friends

List of content

Abstract

Centuries of debate on the topics of the origin of life on Earth resulted in the current viewpoint that small organic molecules formed by cosmic or atmospheric phenomena would slowly accumulate and react under prebiotic conditions on the early Earth to form essential precursors that are basic building blocks of primitive biopolymers. Short biopolymers resulting from the random chemical oligomerization of these building blocks could then be selected by their inhabiting environments, and those with the fittest physicochemical properties to accumulate and enrich would polymerize to acquire higher complexity in structures and functions, a process now regarded as 'chemical evolution'. Among the three component biopolymers of life, namely DNA, RNA and peptides, RNA possesses the capabilities to transfer genetic information by base pairing and perform a wide range of catalytic activities, when folded into ribozymes. Hence, the 'RNA world' theory indicates a key period of the chemical evolution in which RNA molecules served as a key hereditary molecule and evolved rapidly to catalyze rudimentary biochemical reactions that progressively shaped the modern biochemistry. However, the RNA world model faces several unanswered questions and one of those being the origin of translation, a process where RNAs are decoded to form peptides and evolved beyond the RNA world. Translation is characterized by two features - a template-directed peptide synthesis and(in) a defined genetic code dictionary. Attempting to address the mystery of the origin of translation, we looked into non-canonical nucleotides that are today found in the tRNAs and rRNAs of the translation machinery, which are highly conservative among all lifeforms on Earth and are considered as 'molecular fossils'. Since many of them were found to form parallelly with the canonical nucleic acid molecules in various prebiotic reaction pathways, they were likely to be incorporated into early RNA oligonucleotides and served important roles in the early functionalization and survival of RNA.

In this thesis, we designed the chemical synthesis of these modified RNAs and investigated their physicochemical properties. We built a model in which these modifications could perform template-directed peptide synthesis and acquire elementary chemoselectivity by liposome interactions.

1. Introduction

1.1 Early theories for the origin of life

The earliest unreligious theory of the origin of life can be traced back to the period of Aristotle (ca. 343 B.C.), when life was believed to emerge via the mechanism of *"spontaneous generation"*. This theory proposed that lower lifeforms, such as fleas or maggots, can appear from inanimate matters, like dust or dead flesh, by chance without the aid of any causal agent, like parents.[1-3] However, it immediately faced major challenges by the time of the Enlightenment, when microorganisms were first observed under the microscope by Leewenhoek, who proposed that life is originated from cellular interactions.[4] Further rebuttal experiments performed by Redi in 1668 showed that maggots can only grow in meat pieces contained only in open jars but not in sealed ones and later, in 1769, Spallanzani demonstrated the lack of microorganism growth in sealed broth sterilized by boiling in partial vacuum. Spontaneous generation was considered to be fully superseded when Pasteur, in the 19th century, sophisticated Lazzaro's experiment with a "swan-neck flask" to allow air but not air-borne microorganisms to reach a boiled broth and disproved the final possibility of air-dependent spontaneous generation (Fig. 1.1).[5,6]

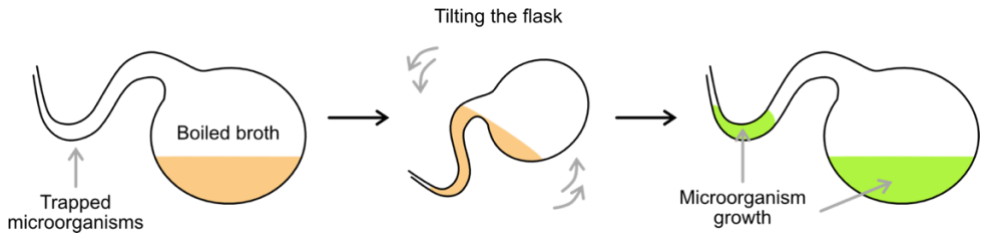
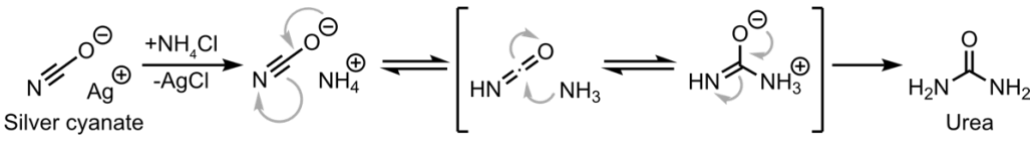


Fig. 1.1. Louis Pasteur's swan-neck flask experiment. *Microorganism growth was only observed in boiled broth stored in an open swan-neck flask after it made contact with the swan neck.*

The discovery of complex unicellular organisms, however, popularized ungrounded beliefs among chemists in the 19th century. For example, Berzelius's *Central Dogma of Chemistry* proposed that life-related and lifeless matters are strictly segregated and the generation of *"organic matters"* found naturally in vegetal and animal chemistry was physically impossible from *"inorganic matters"* that are present in minerals.[7] The divine veil of *"life comes only from life"*, as stated by Pasteur,[8] was slowly torn down in 1828 when Wöhler unexpectedly discovered the formation of urea from silver cyanate and ammonium chloride in an initial attempt to crystalize ammonium cyanate (Scheme 1.1).[9] Although its mechanistic details are not yet fully understood today,[10] the Wöhler synthesis marked the first abiotic formation of a biomolecule in human history from a lifeless, inorganic source.



Scheme 1.1. Wöhler synthesis of urea. *Proposed mechanism of silver cyanate (AgOCN) reacting with ammonium chloride (NH_4Cl) to form urea [$NH_2(C{=}O)NH_2$].[10]*

Three decades later in 1859, Darwin established the theory of biological evolution based on natural selection in "On the Origin of Species". It posited that species were differentiated based on the principle of mutation and propagation, which led to the survival of organisms possessing the fittest phenotypes in adaptation to their environment.[11] The work, however, was immediately criticized as incomplete by Bronn, his German translator and geologist, as it evaded the question of the origin of life.[12,13] Only after 12 years did Darwin, in his private letter to Hooker in 1871, consider the ideas from Spencer[14] and Thieslton-Dyer[15] and first mentioned that a *"little warm pond"* of inorganic matters, such as ammonia and phosphoric salts, could react in response to atmospheric phenomena, such as light, heat and electricity, and lead to the first living organism.[16] Darwin has never published any theory with regards to the origin of life in his lifetime.

Its extended idea, now regarded as *"chemical evolution"*, was coincidentally proposed by both Oparin in 1924[6] and later by Haldane in 1929[17], which suggested that *"commonly available volatiles"* could be chemically activated by either an early oxidizing atmosphere[6] or ultraviolet irradiation[17] to give precursors, like carbon dioxide and ammonia, of simple organic molecules, like amino acids. These precursors would then spontaneously condense and concentrate over time in the young oceans, resulting in a *"primordial soup"* that served as feedstocks for the prebiotic synthesis of early organic matters. These organic matters would then chemically evolve in mass and complexity and eventually gave rise to the first lifeform. The evidence of the Oparin-Haldane hypothesis, however, was only made available in 1952 by Miller and Urey in their groundbreaking experiment which named after themselves and published in 1953 (Fig. 1.2).[18] In a sealed container, a gaseous mixture representing an early atmosphere was circulated by refluxing a solution that mimics a primordial ocean. An electric discharge that represented atmospheric lightning was introduced by high voltage electrodes and applied to the gaseous phase, continuously. After a week, the solution became *"deep red and turbid"*. The initial analysis performed by Miller showed the formation of 5 different amino acids,[18] whereas recent re-analyses of the same mixture revealed the presence of almost all proteinogenic amino acids.[19] From Aristotle's "spontaneous generation of life" to the Miller-Urey's experiment showing "chemical generation of basic molecules of life", the historical debate of the origin of life had finally been concluded – life is a result of chemical evolution from organic molecules originated from inorganic matters that are chemically activated by natural forces.

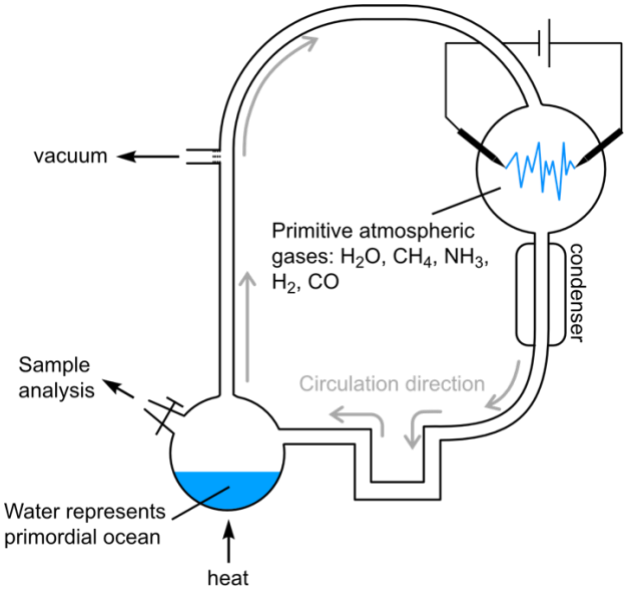


Fig. 1.2. Setup of the Miller-Urey experiment. *Steam was exposed to discharges in a presumably reducing early Earth atmosphere and condensed back to the solution in a closed reflux system.*[18]

1.2 Astronomical evidence of chemical evolution

The existence of a chemical evolution on Earth is evidenced by the analyses of extraterrestrial objects. In 1969, a carbonaceous chondrite (*i.e.* a class of meteorites composed of carbon-rich materials preserved since the primitive stage of the solar system)[20-22] was observed falling near Murchison in Victoria, Australia.[23] The follow-up gas chromatography-mass spectrometry (GC-MS) analyses of the collected samples from the core that experienced the least heat change during its atmospheric entry revealed at least 35 proteinogenic and non-proteinogenic amino acids[24] with a slight excess of the L-isomers.[25] Until now, more than 80 amino acids were found in similar meteorites,[26,27] alongside with a variety of nucleobases.[28-31] To rule out any possible terrestrial contamination, space missions equipped with a probe for on-site sample collection and subsequent mass spectrometry analysis on travelling celestial bodies, like ROSINA (Rosetta Orbiter Spectrometer for Ion and Neutral Analysis) from the European Space Agency,[32] also reported the formation of glycine, phosphorus and other precursors for life-related organic molecules, like hydrogen sulfide (H_2S) and hydrogen cyanide (HCN), in comet volatiles in space.[33,34] Telescopic observations confirmed the presence of various phosphorus[35-38] and cyanide[39-42] species also in the interstellar medium, suggesting a cosmochemical origin of the conserved organic molecules found in meteoritic samples. Collectively, these findings supported the interstellar formation of prebiotic chemical precursors.[43-46] One idea about how life could have started on Earth, therefore, suggests that the molecules were delivered to the Earth surface via the bombardments of celestial bodies.[47]

The investigation of early geochemical conditions on Earth is crucial to understand the prebiotic chemistry of the precursors on the early Earth and how they eventually led to the formation of more complex biomolecules, like nucleos(t)ides. By balancing multiple radiometric dating profiles from meteorites formed by akin solar materials, the formation of Earth is estimated to take place 4.54±0.05 Ga (billion years ago).[48] Its first period, known as

Hadean, is defined from its formation to that of its oldest rock (discovered by then), which is 4.5–3.8 Ga.[49] Most of the prebiotic chemistry is believed to have taken place in Hadean, as the earliest possible evidence of life was suggested by a $^{12}C/^{13}C$ profile of a graphite fractioned by potential biological activities, captured inside a zircon formed at 4.1 Ga.[50] During Hadean, the Earth probably experienced heavy bombardments from leftover solar materials after planetary formations.[51] While prebiotic precursors could accumulate from the frequent impacts, recent studies reconciled with cosmochemical tracers instead deduced a single colossal impact in late Hadean at ca. 4.48 Ga by a differentiated body named as Moneta that re-equilibrated Earth's atmosphere from being mildly reducing and chemically inert to highly reducing and chemically active.[52-54] This temporary active, H_2 containing atmosphere at ca. 4.48 - 4.30 Ga was essential for prebiotic chemistry, as it allowed the existence of C and N atoms in their reduced form (e.g. CO, NH_3, H_2 and CH_4) form, which gave rise to the oxidation states of the C and N atoms found in the chemical structures of RNA nucleosides (Fig. 1.3).[55] While the existence of a highly reducing Hadean atmosphere is still under debate, most of the proposed prebiotic chemistry pathways that describe the formation of amino acids, nucleosides and metabolites are based on this presumably highly reducing condition.

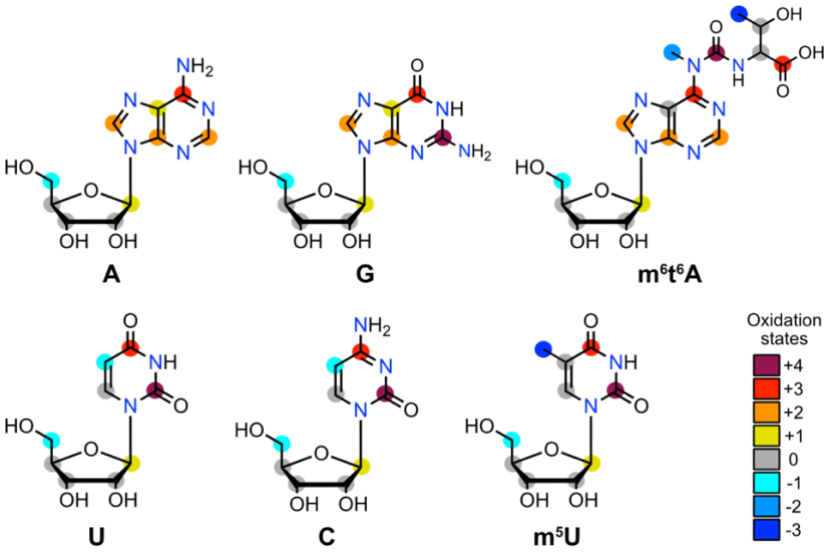
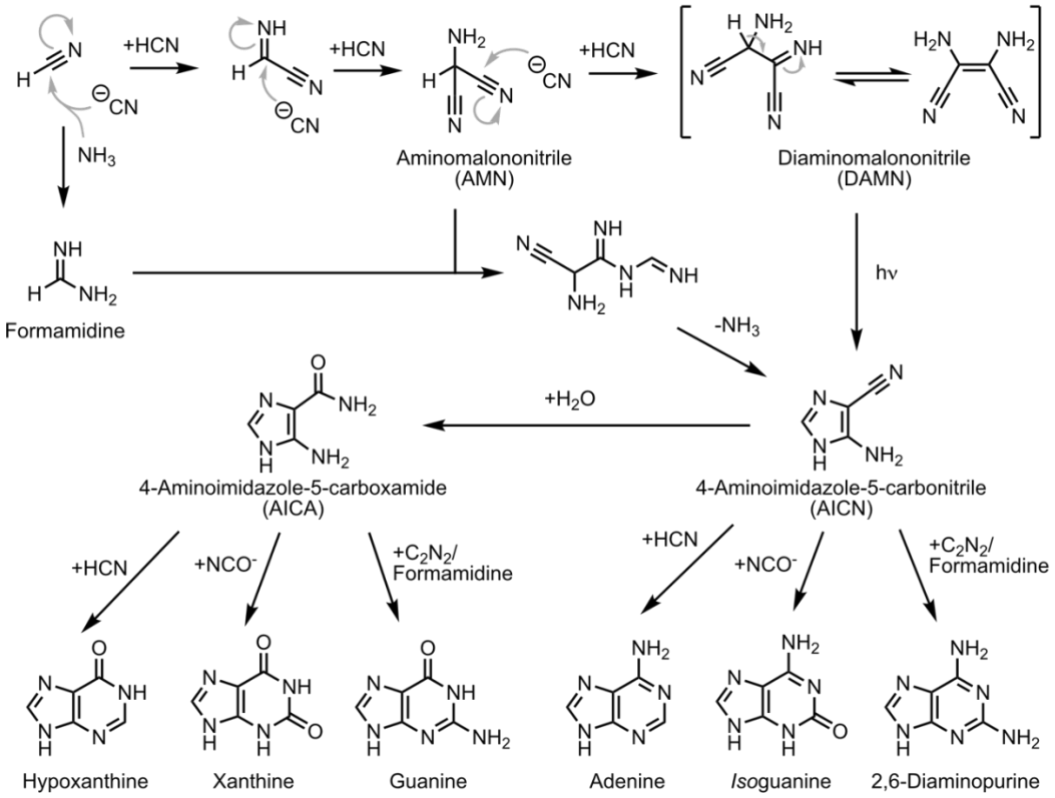


Fig. 1.3. Oxidation states of C and N atoms in RNA. *The C atoms in RNA has a range of oxidation states from -3 to +4 while all N atoms are -3, suggesting a redox-active environment for their formations.*[56,57]

1.3 Prebiotic formation of RNA nucleosides

Although it is impossible to confirm how life commenced on Earth, the exploration of potential chemical pathways leading to life building blocks under prebiotically plausible conditions is valuable to demonstrate *"how inanimate matters could slowly transition to living organic matters."*[58] Since it was deduced that RNA was a crucial biopolymer for the emergence of Darwinian evolution (see section 1.4), the prebiotic pathways leading to RNA building blocks became a pressing question in prebiotic chemistry. In the early 1960s, Oró proposed the first one-pot synthesis of adenine from concentrated HCN and ammonia at room

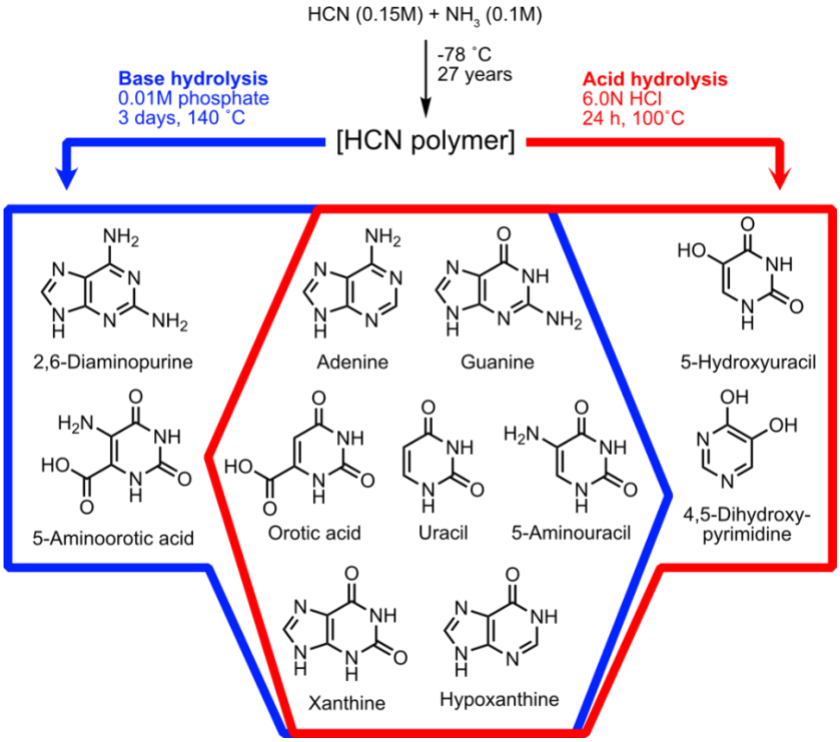
temperature.[59,60] Since the empirical formula of adenine is $C_5H_5N_5$, it is theoretically a cyclized pentamer of HCN. While the reaction suffers from a very low yield (ca. 0.5%), it was later improved (up to ca. 15%) by Takenishi and co-workers in 1966 by heating HCN in liquid ammonia.[61] The key intermediates, 4-aminoimidazole-5-carboxamide (AICA) and 4-aminoimidazole-5-carbonitrile (AICN),[62] were later suggested by Sanchez, Ferries and Orgel to have general applicability for the formation of various purine nucleobases (Scheme 1.2).[63,64] These results provided the first evidence of a potential parallel formation of canonical and non-canonical nucleic acid molecules.



Scheme 1.2. Synthesis of purine nucleobases from HCN. *Mechanism first proposed by Oró[60] and later modified and generalized by Orgel.[63,64]*

The requirement of a high HCN concentration (1.0-11.0 M) among these pathways, however, made their prebiotic plausibility questionable because HCN is unstable over a wide range of pH and temperature.[63,65] This observation prompted researchers to identify a promising prebiotic HCN source, since it is involved in several amino acid and nucleic acid formation pathways.[66] One of the focuses is a type of HCN polymer called "alzulmins". While ubiquitously detected in the solar system,[67,68] its exact structure, despite numerous proposed models, is still unconfirmed today (reviewed in ref.[69]). In 1963, Lowe and co-workers first reported the parallel formation of amino acids, fatty acids and purines (adenine and hypoxanthine) from the acid hydrolysis of HCN polymers.[70] The follow-up study by Ferris and colleagues showed the formation of diverse purine and pyrimidine products under both acidic and basic hydrolysis conditions.[71] This was the first report of prebiotic synthesis of pyrimidine nucleobases from oligomerized HCN. At the same time, Orgel suggested an alternative

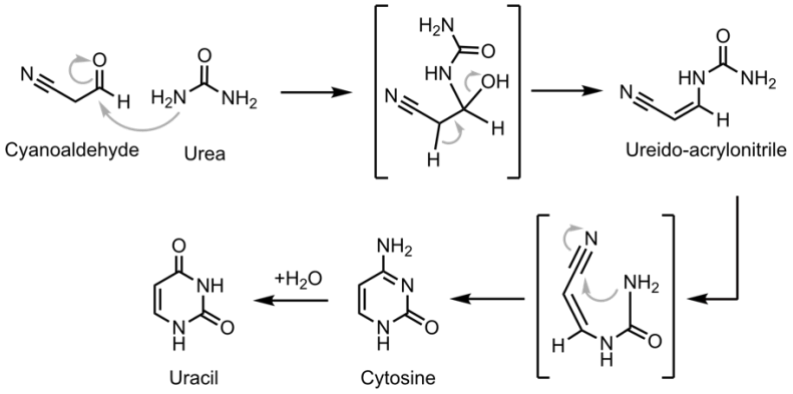
mechanism of concentrating HCN in eutectic mixtures by freezing a diluted solution at –23.4 ˚C.[72] This did not only increase the concentration of HCN to a sufficient level for Oró's pathway, but also catalyzed the formation of diaminomaleonitrile (DAMN), a key precursor of AICA and AICN (Scheme 1.2). These frozen conditions were suggested as plausible during Hadean due to the absence of greenhouse gases and a less luminous young sun.[73] This "cold origin of life" theory was further supported experimentally by Schwartz *et al*,[74] and Miller and his team.[73,75,76] Remarkably, Miller and coworkers reported a solution of ammonium cyanide (NH_4CN) formed by freezing gaseous HCN and ammonia at -78 ˚C for *27 years*, with a final HCN concentration of 0.15 M. The subsequent acidic and basic hydrolysis reactions of the supernatant formed a collection of at least 11 different nucleobases (Scheme 1.3).[65,75] The highly abundant non-canonical nucleobases in this result suggested a high possibility of their incorporation into the early nucleic acid biopolymers.



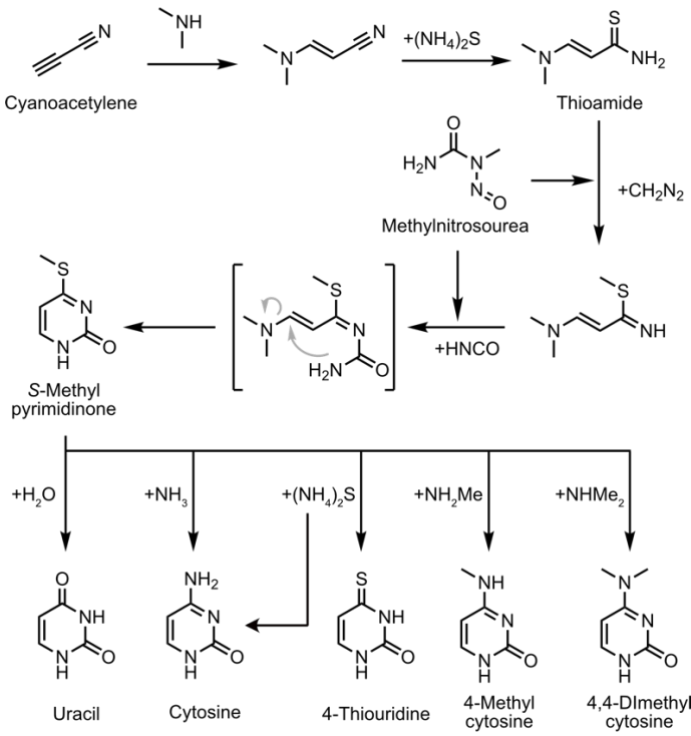
Scheme 1.3. Synthesis of nucleobases from the hydrolysis of HCN polymers. *Nucleobases formed by acid or base hydrolysis of HCN polymer reported by Miller.*[75]

Concurrently, various synthetic pathways of nucleobases based on other HCN-derived precursors were also explored. For example, cyanoacetylene (N≡CC≡CH) formed by electric discharge of methane and nitrogen, or heating acetylene with activated nitrogen, was shown by Orgel's team in 1966 that it can, in high concentration (> 0.1 M), react with urea and potassium cyanate to form cytosine,[77] which can further hydrolyze to uracil.[78] Since cyanate and cyanoacetylene hydrolyze readily in water to form the more stable cyanoaldehyde and ammonium carbonate, respectively, Ferris *et al* proposed the formation of 2,4-diaminopyrimidine by heating cyanoaldehyde with guanidine, which then hydrolyzed to cytosine and uracil in low yield.[79] This was later improved by Miller's team by heating

cyanoaldehyde with highly concentrated urea in various drying beach and lagoon models, to form cytosine and uracil in higher yields (Scheme 1.4),[80-82] without considering the potential decomposition and side reactions accompanied with the dry-down process.[83] Very recently, Carell and co-workers showed an alternative pathway of cyanoacetylene-based pyrimidine formation, which featured a more stable thioamide intermediate formed by reacting cyanoacetylene with dimethylamine. The thioamide could then react with methyl nitrosourea to form *S*-methylpyrimidinone, which reacts with different nucleophiles to form various canonical and non-canonical pyrimidines (Scheme 1.5).[84]



Scheme 1.4. Miller's synthesis of cytosine and uracil from cyanoaldehyde under dry-down conditions. *Mechanism of dry-down reaction of urea and cyanoaldehyde described by Miller.*[80]

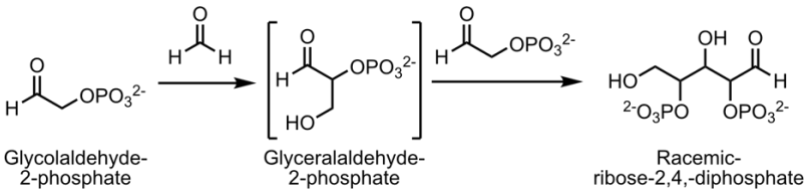


Scheme 1.5. Carell's synthesis of pyrimidines. *Cyanoacetylene-based pathway to canonical and non-canonical pyrimidine published by Carell's group.*[84]

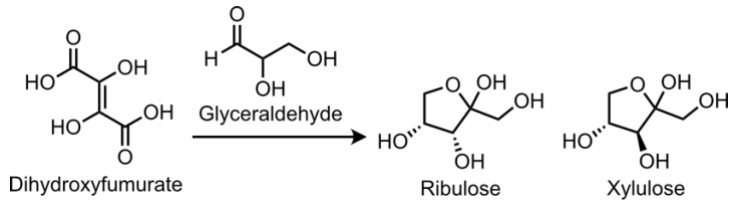
Another focus is the formamide-based pathways catalyzed by minerals. Although formamide can be readily formed by the hydrolysis of HCN ($HCN + H_2O \rightarrow HCONH_2$), its accumulation is prebiotically questionable. Formamide is highly hygroscopic and its further hydrolysis to ammonia and formic acid ($HCONH_2 + H_2O \rightarrow HCOOH + NH_3$) is kinetically 50-fold faster than its formation,[85] unless it is the protected by specific geochemical environment, like porous mineral surfaces.[86] Nevertheless, heating pure formamide at > 150˚C can lead to the formation of purine ring in high yield,[87] and a diversity of, again, canonical and noncanonical nucleobases if a catalytic amount of metallic oxides is present, like zeolite, silica and titanium oxide.[88,89] Clay minerals, first suggested by Bernal in 1951,[90] were also explored as potential inorganic catalysts. By heating formamide with various types of montmorillonite and aluminium-pillared clays, a plethora of nucleobases were obtained alongside with their key precursors such as AICA, as shown by Di Mauro, Saladino and colleagues.[91,92] A remarkable experiment from the same team explored potential catalytic properties of the Murchison meteorite material and resulted in a variety of heterocycles and carboxylic acids.[93] Similar investigations into various phosphate,[94] iron-sulfur,[95] zirconia,[96] borate,[97] and iron oxide[98] minerals all resulted in mixtures of canonical and non-canonical nucleobases, with varying selectivity of the reaction depending on the nature of the catalysts and, in some cases, the irradiation of UV light.[99,100] Alternatively, formamide could also be exploited as an HCN source via reductive dehydration and further reacted to form purine and adenine, as showcased by a recent work from Krishnamurthy and co-workers.[101]

Apart from pathways related to cyanoacetylene, cyanoaldehyde and formamide, alternatives from other non-HCN derived precursors were also suggested. For instance, uracil from malic acid and urea,[102] hypoxanthine and adenine from glycinamide and diformylurea,[103,104] and uracil and thymine from 5,6-dihydrouracil and 5,6-dihydrothymine, respectively.[105,106] The field of prebiotic synthesis has expanded drastically throughout the past 4-5 decades and it is impossible to cover all published pathways here. However, all of these results share one common conclusion: a potential co-emergence of canonical and non-canonical nucleic acid molecules.

The discovery of nucleobases in meteorites suggested that nucleobases and ribose were formed separately, and subsequently coupled. It was shown that ribose could be selectively stabilized and isolated from the complex formose reaction[107-109] with inorganic ions, such as lead,[110,111] borate[112,113] and silicate.[114] To avoid the formation of tar, Eschenmoser and colleagues proposed a pathway of aldolization of pre-phosphorylated, formaldehyde-based precursors, such as glycolaldehyde-2-phosphate,[115] to hinder potential keto-enol tautomerization and favored the formation of diphosphorylated ribose (Scheme 1.6).[116] Additionally, a glyceraldehyde pathway with dihydroxyfumarate recently suggested by Krishnamurthy's group demonstrated the formation of ribulose and xylulose in almost quantitative yield,[117] which can be then converted to ribose in a prebiotic manner (Scheme 1.7).[118,119] However, the required high concentration of formaldehyde (0.1 M) makes all formose-based pathways questionable in a prebiotic context, since it can readily degrade to methanol and formic acid in water via the Cannizzaro reaction.[120-122]

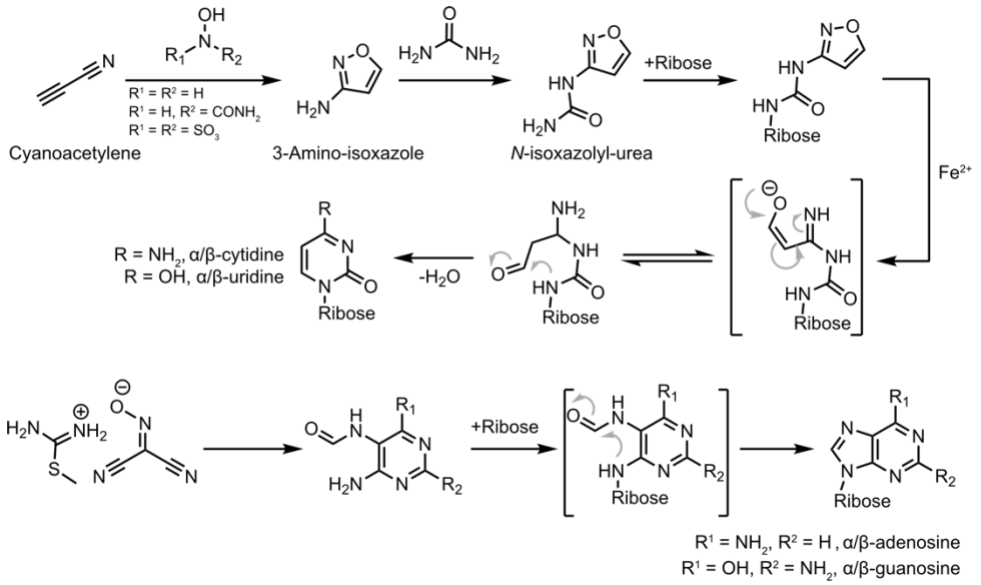


Scheme 1.6. Formation of ribose diphosphate. *Synthesis of ribose diphosphate from phosphorylated precursor to avoid tar formation proposed by Eschenmoser and colleagues.*[116]



Scheme 1.7. Formation of ribose precursors. *Krishnamurthy's pathway to ribulose and xylulose from dihydroxyfumurate.*[117]

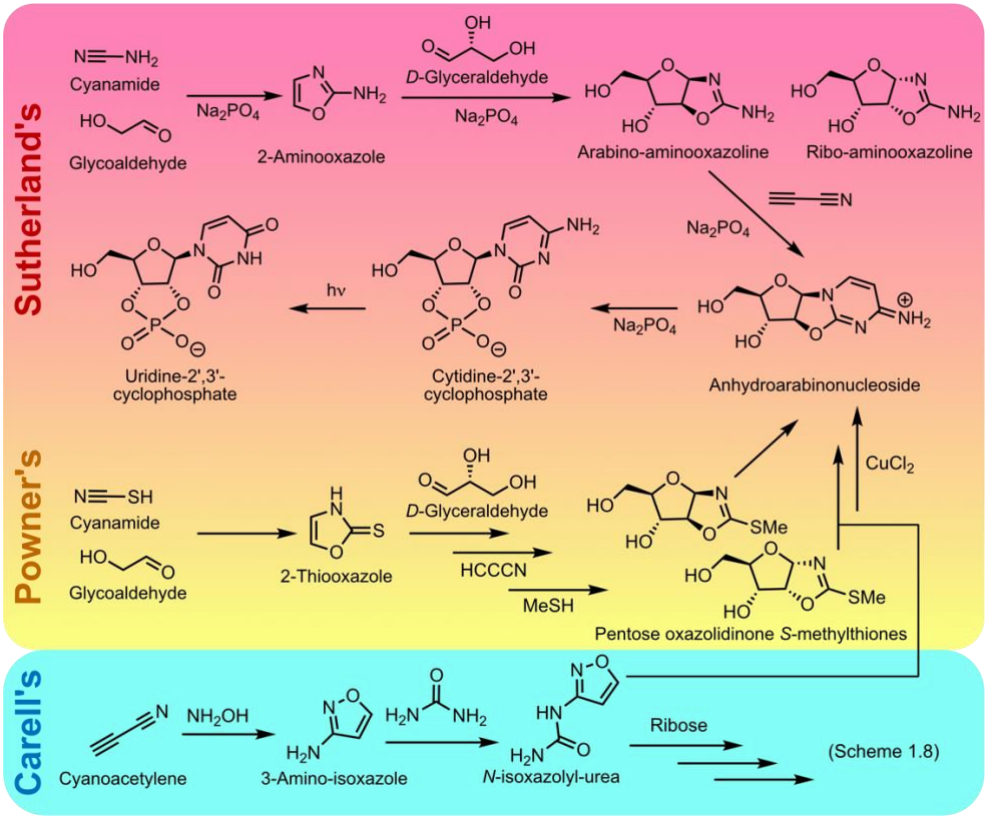
Nevertheless, all of the separate formation of nucleobases and ribose will eventually converge into the "nucleosidation problem" in nucleoside formation, since most of the attempts in direct coupling of nucleobases at the abasic site, especially pyrimidines, were (and still are) considered unsuccessful. In 1972, Orgel examined the direct coupling of purines onto *D*-ribose in the presence of Mg^{2+} ion and inorganic polyphosphates.[123,124] While the correct isomers of *β*-adenosine and *β*-guanosine were successfully detected in low yield (4 % and 9% respectively), incorrect isomers connecting the ribose at the exocyclic amines were found to be the major products.[123,124] The same studies also revealed that the glycosidic bond formation of pyrimidine nucleobases is much less favorable than the purines, which is due to the reduced nucleophilicity of the pyrimidine nitrogen at position 1.[125] Benner's team recently reported a successful regio- and stereoselective nucleosidation of ribose-1,2-cyclic phosphate with purine bases, generating their respective ribonucleoside 2'-phosphates.[126] Yet, similar nucleosidation was not observed on pyrimidines.[126] Based on the same cyclic phosphate precursor, a unified formation of all four canonical nucleosides were recently reported under dry heating conditions. However, only trace amount (< 1 %) of uridine and guanosine were detected.[127] A more innovative approach was recently explored by Carell and colleagues to nucleosidate late nucleobase precursors, followed by an on-ribose cyclisation, to form both canonical and non-canonical nucleosides (Scheme 1.8). While these pathways feature promising yields and regioselectivity, it is not chemoselective, leading to the formation of a mixture of ribofuranose and ribopyranose analogs.[128-130] The pathways are also ribose-dependent, whose prebiotic availability remains doubtful as mentioned above. Hence, the obvious shortcomings of direct nucleosidation made researchers start to consider alternative, indirect formation of nucleosides.



Scheme 1.8. Carell's prebiotic synthesis of purine and pyrimidine nucleosides. *A ribose-dependent pathway featuring on-ribose cyclisation of nucleobases.*[129,130] *Only the formation of the canonical nucleosides is shown.*

In 1970, Orgel reported the reaction of *D*-ribose with cyanamide, which produced the cyclic intermediate ribo-amino oxazoline and, eventually, the canonical *β*-ribocytidine.[131] Inspired by this work, Sutherland and Powner in 2009 reported the first ribose-free pathway leading to ribo-amino oxazoline and, subsequently, canonical cytidine and uridine nucleotides in high yields (Scheme 1.9).[132] Starting from smaller formaldehyde-derived precursors, like glycoaldehyde and glyceraldehyde that are considered to be more prebiotically plausible,[133] the pathway suggested the first co-formation of nucleobase and ribose in a step-wise fashion, albeit most of its steps are heavily dependent on the presence of a concentrated phosphate buffer (1.0 M). In the absence of phosphate, the undesired *β*-arabinoylcytidine would remain as the major product.[132] Similar pathways based on the same idea with improved yield and anomerization were subsequently reported.[134,135] Further efforts were made by Powner's and Szostak's groups to generate pyrimidine and purine nucleobases in a unified pathway with common precursors of pentose-oxazolidione thiones. While the formation of 8-oxo-purine nucleosides is possible, their conversion to the canonical ribonucleosides remained unsuccessful.[136] Moreover, all of these pathways assumed a homo-chiral *D*-glyceraldehyde origin, whereas both stereoisomers should have co-existed as a racemic mixture with similar abundance. While recent works proposed potential reasons for the opposing chiralities between nucleic acids with D-ribose and L-amino acids,[137,138] the question of why certain chirality is specifically assigned to certain biomolecules is never answered, and its wide discussion is beyond the scope of this thesis. More recently, Carell and colleagues explored the compatibility among the Carell's, Powner's and Sutherland's pathways, suggesting a parallel formation of nucleic acid molecules via multiple pathways in the prebiotic world (Scheme 1.9).[139] Although none of the pathways is perfect and our understanding of the early Earth environment is limited, the collective chemical effort has convinced molecular biologists about the prebiotic availability of a wide range of RNA building blocks, which led the origin of life question to a macromolecular level - how did

these building blocks evolve to polymerise, acquire structure, obtain catalytic activities and eventually lead to the first cell?



Scheme 1.9. Carell's, Powner's and Sutherland's synthesis of cytidine and uridine. *The ribose-independent pathway forms ribose indirectly with 2-aminooxazole and glyceraldehyde while heavily dependent on a concentrated phosphate buffer proposed by Sutherland*[132] *and the pentose-oxazolidione thiones pathway by Powner and Szostak*[136] *can be combined with the 3-amino-isoxazole pathway proposed by Carell.*[130,139]

1.4 The RNA world theory and its limitations

In 1953, Watson and Crick published a model of the molecular structure of double helical DNA[140] based on the result obtained by Franklin.[141] Shortly after, Crick proposed the "*Central Dogma of Molecular Biology"* in 1957 - a universal system to all lifeforms of genetic information being stored and replicated as DNA, delivered to RNA through transcription, and subsequently decoded to protein through translation in an irreversible manner.[142,143] This immediately led to the question of which biopolymer, RNA or protein, came first at the beginning of life, as the former stores "information" and the latter features "function". Functions are meaningless without being organized by information, whereas information is defined by its expression to functions.[144] This "chicken-and-egg conundrum" first received an acceptable answer when Cech in 1982[145] and Altman in 1983[146] discovered RNAs that catalyze self-splicing, referred as "ribozymes". The RNA world model, first suggested by Rich in 1962,[147] was thereafter coined by Gilbert in 1986, postulating a phase of evolution of life

centered around RNA that existed before the evolution of living systems into their modern DNA-protein world.[148] During this period, *"both the information needed for life and the enzymatic activity of living organisms were contained in the RNA molecules."*, as stated by Miller.[149] Another major support of the RNA world theory came from the studies of the ribosome and its mechanisms in catalyzing template-directed peptide synthesis of translation. Through decades of effort, the high resolution atomic structure of the ribosome was finally obtained from Ramakirshnan's,[150] Steitz's[151] and Yonath's[152,153] labs in the early 2000s. While proteins catalyze almost all cellular functions nowadays, the ribosome's crystal structures confirmed a long speculated idea – the active site known as the peptidyl transferase center is mainly formed by the ribosomal RNAs (rRNAs).[154] Furthermore, it was shown that rRNAs alone possess a reduced catalytic function of peptide synthesis in the complete absence of ribosomal proteins,[155] and confirmed that the ribosome is, indeed, a ribozyme assisted by proteins. These findings, accompanied by genomic studies revealing that rRNAs are highly conserved across all kingdoms of life,[156,157] suggested by Woese that evolution of the translation machinery can be dated back to the last universal common ancestor (LUCA) of all Earth's lifeforms.[156,158]

While an increasing number of natural ribozymes were reported after the discovery of Altman and Cech, many believed that most of the ancient ribozymes were lost throughout the evolution and therefore, it is difficult to explore the full capability of RNA catalysis. Until 1990, teams from Gold[159] and Szostak[160] separately described a novel method to evolve *de novo* functional RNAs *in vitro* from randomized sequences, namely systematic evolution of ligands by exponential enrichment (SELEX). At the very beginning, this method was designed to experimentally evolve "aptamers" - RNAs that fold to bind to a specific target, such as proteins, small molecules or metal ions. SELEX typically starts with a library with region(s) of randomized sequence flanked by defined primer sequences. Since different sequences possess different secondary structures and phenotypes, selection conditions, like exposing the library to a target molecule anchored on a solid support, is designed to selectively retain and isolate sequences that possess the desirable properties. The enriched library can then be amplified by a polymerase recognizing the primer region and re-introduced into a subsequent selection cycle, a process referred as 'mining'. After several rounds, the enriched library can be sequenced and *de novo* nucleic acids expressing designated phenotype can be identified (Fig. 1.4). Shortly after, methods were further developed to *in vitro* select *de novo* ribozymes.[161-163] These techniques, now referred as "directed evolution", allow molecular biologists to apply Darwinian selection pressure directly on biopolymers and understand the catalytic capabilities of ancient ribozymes by generating and studying their modern analogues.[164]

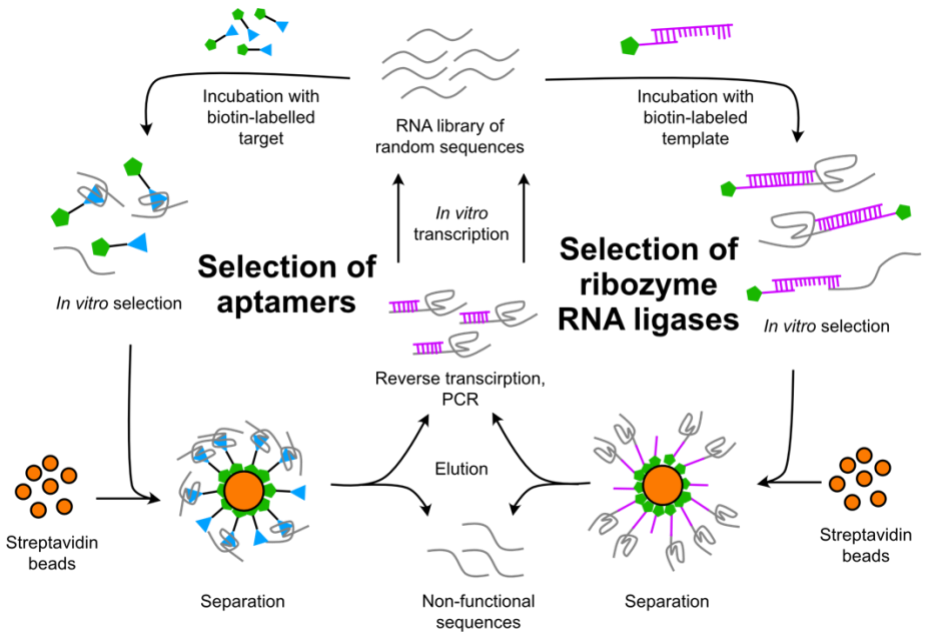


Fig. 1.4. *In vitro* selection cycles of aptamers (left) and RNA ligases (right). *Evolution of functional RNAs by applying selection conditions to a pool of randomized sequences.*

One of the most important catalytic activities in the RNA world is self-replication.[165] The discovery of natural ribozyme ligases in the late 1980s suggested a potential RNA-catalyzed RNA replications in the earlier stage of life.[166,167] In 1993, Bartel and Szostak isolated the first set of *de novo* RNA ligases using *in vitro* evolution from a pool of random sequences.[168] Among those, class I ligases were identified in 1995 to catalyse RNA elongation by forming a 5'-3' linkage, resembling the regioselectivity of protein polymerases.[169] Further evolution and engineering resulted in b1-207 (GenBank: U26413.1), a 119 nt. long ribozyme polymerase capable of extending the 3'-end of an RNA with nucleoside triphosphates (NTPs) in a template-directed manner, with $k_{cat} > 1\ s^{-1}$ and average fidelity >85%.[170,171] These ligases, however, suffer from a major limitation – they can only catalyze the polymerization with templates that were brought to their proximity by hybridizing themselves to a particular unpaired segment of the template, which resemble the activity of protein telomerases more than that of RNA or DNA polymerases.[168] Further selection and engineering from the Bartel's lab in 2001 introduced an additional random sequence domain to the catalytic core and removed this constrain, forming the 189 nt. long ribozyme R18 that allows general template-directed RNA primer-extension without any sequence requirement on the template strand.[172] It is noteworthy that the sequence of the catalytic core inherited from the class I ligase experienced little change during the mutation, indicating a high degree of evolutionary optimization.[172] While this initial ribozyme could only extend RNAs up to 15 nt. long, further evolutions in the past decades evolved ribozymes with better and better polymerization capabilities. Until now, one of the best ribozymes reported from the Holliger's group can polymerize RNAs up to 200 nt., longer than the ribozyme itself.[173] Very recently, Joyce's lab demonstrated the capabilities of ribozymes derived from the class I ligase to synthesize a functional class I ligase,[174] and to drive the synthesis of the hammerhead ribozyme in an artificial directed evolution

experiment.[175] All of these experiments displayed the wide self-replicating potential of RNAs in the early world.

Still, there are major questions that remained completely unanswered in the RNA word theory. The first is the disfavored accumulation of longer RNAs, which are not considered as stable biopolymers due to the spontaneous hydrolysis of its phosphodiester backbone. This hydrolysis is facilitated by the intramolecular nucleophilic attack of its 2' hydroxyl groups, forming a straight line with the phosphate oxygen, regarded as the 'in-line hydrolysis'. It can be enhanced by alkaline conditions and the presence of bi- or trivalent cations, like Mg^{2+}, Pb^{2+} and Eu^{3+}, which are common conditions used in prebiotic pathways.[176,177] This intrinsic instability makes RNA degradation much more favorable than its polymerization. As a result, ribozyme ribonucleases are much more abundant in nature, and easier to emerge from *in vitro* selections than ribozyme polymerases.[178] In addition, ribozyme ribonucleases are usually shorter (<100 nt.), which makes them prebiotically more plausible. For example, the shortest ribozyme of all types ever discovered is UUU, which promotes the cleavage of GpA in 5'-GAAA-3' in the presence of Mn^{2+},[179] whereas the shortest ribozyme polymerase reported is 150 nt. long.[180] It is hard to imagine how a net increase in RNA length could have happened in the early RNA population and, subsequently, allow the first RNA self-replicase to emerge from the primordial soup, without specific mechanisms to protect them from degrading.

Even if such a replicase could emerge and amplify under some very specific conditions, it would then face the second question - the "strand separation problem".[181] RNA secondary structures, like duplexes and G-quadruplexes, are very stable and rigid unlike their DNA counterparts, since the 2' hydroxyl groups provide extra hydrogen bond stabilization. An RNA GC repeats of ca. 20 nt. long in 1 M NaCl, can easily reach a melting temperature of >95 ˚C, exceeding the boiling point of water (predicted with ref.[182]). While this could potentially provide a certain degree of hydrolysis resistance for longer RNAs to accumulate, such structures, once formed, are almost impossible to de-hybridize and the RNA molecules are kinetically trapped. Therefore, in a hypothetical solution of concentrated RNA self-replicase, if given enough time, some of the molecules would start to unfold its kins and use them to synthesize its template strand (- strand). For a complete self-replication cycle, the self-replicase would then need to use this template strand to synthesize itself (+ strand). The template strand that got amplified alongside with the ribozyme is, however, a toxic non-competitive inhibitor to the system, as it possesses high complementarity and tends to unfold the ribozymes to dead-end RNA duplexes. Very recently, attempts from the Holliger's group to solve this problem showed that evolved ribozyme polymerases using trinucleotide triphosphates (NNNTPs), instead of nucleotide triphosphates (NTPs), can better invade the secondary structures of folded template strands.[183] Further experiments using similar polymerases with specific template constructs, like a circular template,[184] or by utilizing temperature and pH cycles,[185] allowed ribozyme polymerase to replicate from long RNA template duplexes. This system, indeed, is considered to be prebiotically plausible, as it is widely reported that nucleosides can be chemically activated and oligomerized to very short RNA fragments under prebiotic conditions. For instance, nucleotide monophosphates can be activated by imidazole to form phosphorimidazolides,[186-188] and oligomerize to form oligonucleotides.[189] While these nucleotide phosphorimidazolides alone can only form short linear or cyclic oligomers in aqueous solution, the presence of cations, such as Pb^{2+}, Zn^{2+} and Lu^{3+}, can catalyze formations

of longer products, like pentamers, tetramers and trimers.[190-192] These ion-assisted chemical oligomerizations were found to be promoted by eutectic conditions,[193,194] and the most effective metallic ion, uranyl (UO_2^{2-}), could form RNAs up to decamers to hexadecamers with predominantly 5'-2' linkages.[195,196] More remarkably, Ferris demonstrated oligomerization up to 50-mers with Montmorillonite.[197] Similar imidazole-based activations were also explored in non-enzymatic template extensions.[198,199] All of these suggested that longer RNAs, if possible to out-compete its degradation, were likely to form by ligation of non-enzymatically coupled RNA oligomers instead of by single addition of nucleotides like modern enzymes in a prebiotic context.

The third question comes to how an RNA world could have advanced into a protein world. This problem remains as the greatest mystery of the origin of life, as the RNA world does not describe the origin of translation. Translation is a universal process for almost all protein productions in all lifeforms. Its highly efficient and accurate operation today necessitates multi-facet cooperations of macromolecules and hence, a self-evidenced product of a complex evolutionary history. Its nature, however, can be broken down into two key components – a template-directed peptide synthesis and(in) a defined set of genetic codes. Inaccuracy in either one of these would lead to unfaithful translation with severe physiological costs. For the former, it is well described in the literature that low-fidelity codon-anticodon recognition, such as stresses produced by non-physiological Mg^{2+} concentrations, pH, temperature or the presence of ribosome-targeting antibiotics like streptomycin,[200-202] could drastically increase the production of misfolded proteins instead of their functional counterparts, and malfunction proteins fail to catalyze life-essential biochemical reactions is a death sentence to its organism. However, similar result can also appear if amino acids are not being loaded onto their cognate tRNAs accurately.[203,204] Mutations can still happen with an accurate ribosome when an error-ridden aminoacylation makes one codon to code for more than one amino acid, even though there is no change in the genetic information. In 1962, a remarkable experiment from Chapeville demonstrated that the ribosome can undergo translation with tRNA of a cysteine anticodon mischarged with an alanine just as efficient as the normal ones *in vitro*, revealing that peptide synthesis catalyzed by the ribosome and cognate aminoacylation catalyzed by the amino acid-tRNA synthetases (aaRS) are two separate systems that closely co-evolved.[205] Hence, in a prebiotic scenario where genetic codes were undefined, a template-directed peptide synthesis would be no different from a random oligomerization of amino acids, if the same codon can be assigned for all 22 amino acids unbiasedly. Indeed, it is even worse, as extra energy was consumed to load an unrecognized amino acid onto an unrecognized RNA, if we follow the modern biochemical logic. This presents an evolutionary paradox - the correspondence between the amino acids and the codons is provided solely by the aaRS, a group of accurately made proteins nowadays, but proteins cannot be made accurately without highly specific aaRS to provide the genetic codes.[206] This implies that the selection of certain amino acids must be linked to the selection of certain RNA molecules through other means in a protein-free prebiotic world, and they needed to be two dependent events, instead of two independent ones. Yet, this first genetic code dictionary must be very ambiguous, as it is very likely to be based on the innate physicochemical properties of both the RNA and amino acid molecules, like hydrophobicity. Despite numerous speculations,[207] no experimental attempt was reported to try to address this problem, until very recently from a collaboration between

the Sutherland's and Jäschke's groups.[137] Their work hypothesized that a 3-letter code upstream of the amino acid loading site could introduce chemoselectivity on the loading reaction. They showed that when a specific amino acid was reacted with a library of RNA duplexes with random codes, a favored loading was observed for certain sequences. However, when the favored sequence was later exposed to a mixture of amino acids, almost all of them failed to react preferentially with their corresponding amino acids, an example of treating the selection of amino acid and the selection of RNA as two independent events, whereas they should be co-dependent.[137]

In this thesis, we describe an experimental model that treats the selections of RNAs and oligopeptides as dependent events, and attempts to address the problem of the origin of translation and genetic code. This model combines the concept of protocell and the molecular fossil theory, which postulates that early RNAs were functionalized by non-canonical nucleobases that were formed parallelly with the canonical ones in prebiotic pathways. In short, we hypothesize that RNAs and oligopeptides with lipid-binding capability can interact with the surface of a liposome membrane and allowed an elementary co-recognition based on their physicochemical properties, which formed the basis of the first genetic code dictionary.

2. Results

2.1 Synthesis of modified oligonucleotides and their physicochemical properties

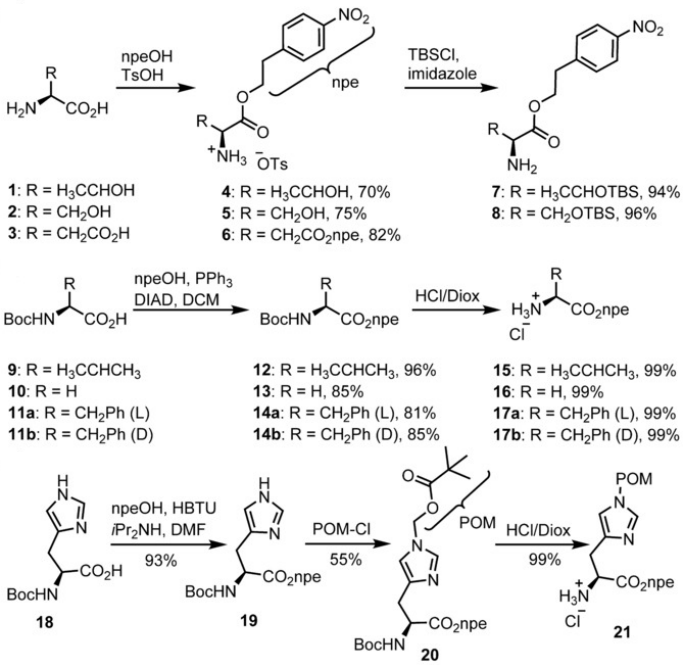
We initially hypothesize that aminocarbamoyladenosines (*aa*^{6}A) found in tRNAs can both base-pair and incorporate amino acids into RNA structures. We wished to observe different physicochemical properties of the RNA duplex when the incorporated *aa*^{6}A is charged with different amino acids and therefore, served as a starting point for a primitive genetic code dictionary. To this end, we began with the synthesis of various *aa*^{6}A-containing oligonucleotides. The results of this section are published in ref.[208,209] and redescribed here. (Information on all oligos described in 2.1 is listed on p. 25 and section 4.2)

2.1.1 Synthesis of *aa*^{6}A phosphoramidites and their respective oligonucleotides

We designed the synthesis of *aa*^{6}A nucleoside phosphoramidites (*aa* = Asp, Gly, His, Phe, Thr,[210] Ser & Val) and their incorporation into DNA and RNA via solid-phase synthesis. To begin with, we first prepared protected amino acid analogs (Scheme 2.1). As a general strategy, we protected the carboxylic group of all amino acids as an *p*-nitrophenylethyl (npe) ester. For amino acids that contain a hydroxyl or carboxylic acid side-chain, like Asp, Thr and Ser (**1-3**), the npe group was coupled via a reported pathway that does not have side reactions with the amine or hydroxyl groups, resulting in **4-6**.[210] Double equivalent of npeOH was used in protecting also the side-chain carboxylic acid in Asp (**6**). The hydroxyl groups of Thr and Ser were subsequently protected as TBS-ether to provide the final products, **7-8**. For amino acids with unreactive side chains, like Val, Gly and Phe, we employed the Mitsunobu reaction[211] for their npe couplings. Starting from commercially available Boc-protected amino acids **9-11**, we first protected the C-terminus with npe to form **12-14** and, subsequently, deprotected the *N*-Boc with 4 M HCl in dioxane to yield the ammonium salts **15-17** that can be readily precipitated from the reaction mixture by the addition of diethylether (remarks: it was discovered later after our publication that the C-protection and N-deprotection can be performed in one pot without column purification of **12-14**). For His, the Boc-protected **18** was coupled to form npe-ester **19** via hexafluorophosphate benzotriazole tetramethyl uronium (HBTU) activation to avoid side reaction(s) with the side-chain imidazole amine. This amine was subsequently protected with pivaloyloxymethyl (POM) chloride to form **20**, prior to the acidic Boc-deprotection which gives **21** as the final product.

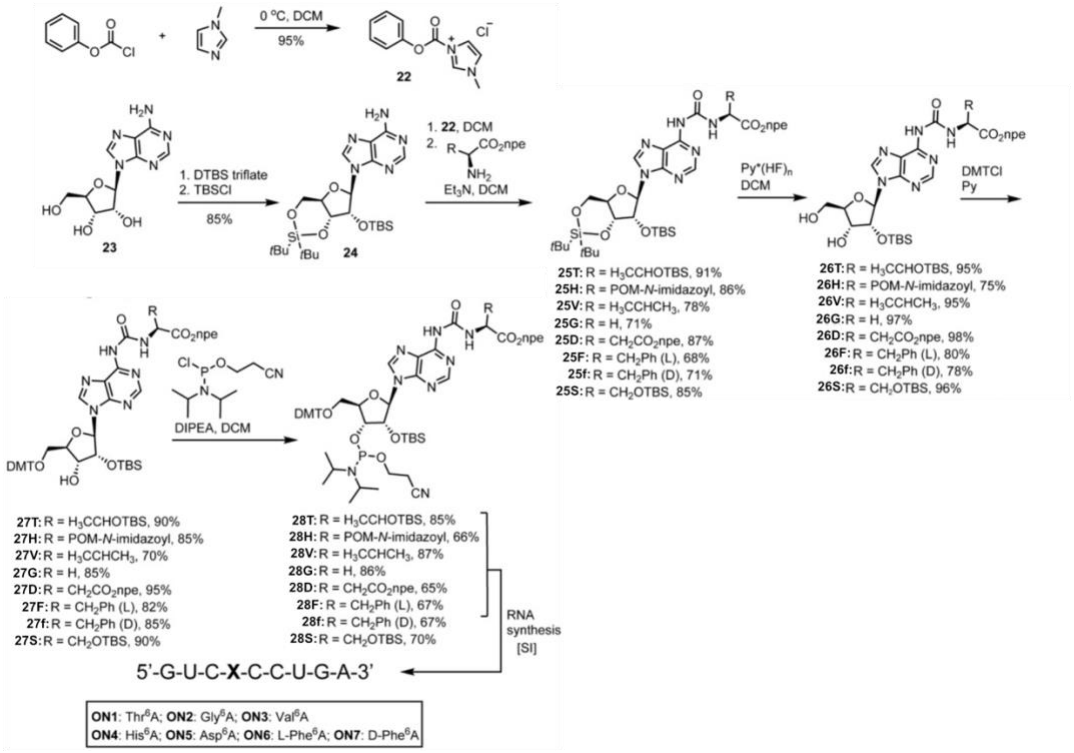
With the protected amino acids in hand, we then synthesized their respective phosphoramidites (Scheme 2.2). A urea moiety can be formed by employing 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**22**), which readily precipitated out by mixing phenyl chloroformate with *N*-methylimidazole.[212] In parallel, free adenosine (**23**) was 3',5' silyl-protected with a cyclic di-*tert*-butylsilyl group and 2' silyl-protected by TBS-ether in one step, forming **24**. The protected adenosine **24** was then reacted with the activated carbamate **22** and then the protected amino acids (**7-8**, **15-17** & **21**) to form amino acid-coupled adenosine nucleosides **25[*aa*] (*aa* = IUPAC amino acid code)** with good yields in all cases. The

nucleosides were then converted to their respective phosphoramidites following standard phosphoramidite synthesis pathway. First, selective deprotection of the 3' and 5' was achieved by reacting **25[*aa*]** with HF•pyridine in an ice bath (**26[*aa*]**),[213,214] the primary 5' OH group was then protected by reacting with 4,4'-dimethoxytritylchloride (DMTCl) (**27[*aa*]**),[215] and the 3' OH group phosphitylated to form the final products (**28[*aa*]**). The phosphoramidites were then employed for the solid-phase RNA synthesis as described in Fig. 1.6 without any adjustment to the synthesis cycle. To deprotect the npe group, an additional step was introduced to incubate the controlled pore glass (CPG) beads in 10% 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU) in THF for 2 h at r.t. before ammonia cleavage (before step 4 in Fig. 1.6). RNA oligos **ON1-7** were successfully synthesized with their respective modifications incorporated.

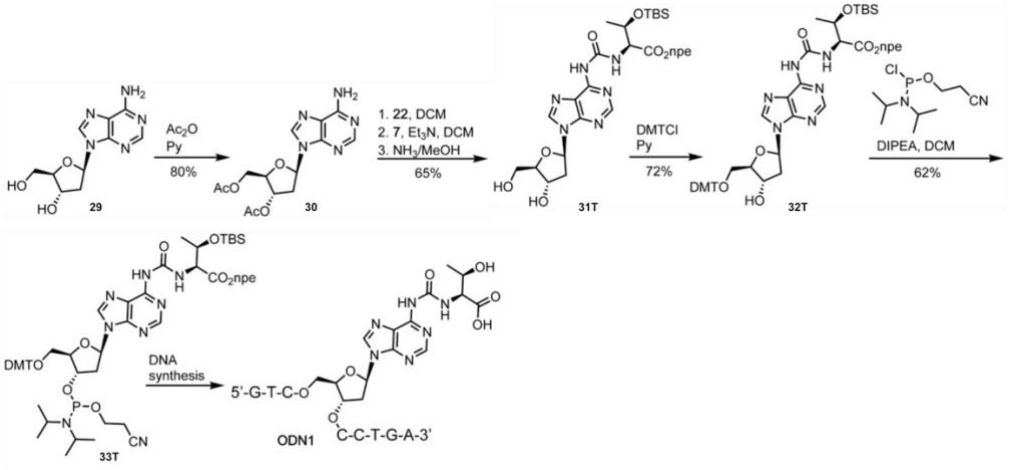


Scheme 2.1. Protection of amino acid building blocks. *General strategy for protecting carboxylic groups of amino acids by npe and other side-chain functional groups by their corresponding protecting groups. Reprinted from ref.*[208]

To also investigate how these modifications could affect DNAs, although unfound in nature, a representative DNA phosphoramidite of t^6dA was also synthesized (Scheme 2.3). With minor adjustments, the 5' and 3' OH groups of deoxyadenosine **29** were first protected by acetylation to form **30**. A similar amino acid coupling step with **7** and **22**, followed by basic deacetylation with ammonia in one step to form **31T**. Standard phosphoramidite synthesis that protects 5' OH with DMTCl (**32T**), followed by 3' OH phosphitylation to form the final product (**33T**) was performed. DNA oligo **ODN1** was synthesized again with solid-phase synthesis with the added npe deprotection step.



Scheme 2.2. Synthesis of *aa*^{6}A phosphoramidites. *General strategy for synthesis of aa^{6}A phosphoramidites. Modified from ref.*[208]



Scheme. 2.3. Synthesis of t^{6}dA phosphoramidite. *The synthesis is adjusted from that of its RNA analogs described in Scheme 2.2. Modified from ref.*[208]

The DNA and RNA oligonucleotides were purified by high performance liquid chromatography (HPLC). Fig. 2.1 shows the crude and isolated HPL-chromatograms of **ON1** and **ODN1** and their respective MALDI-TOF mass spectra, indicating high purity of the obtained RNA and DNA molecules.

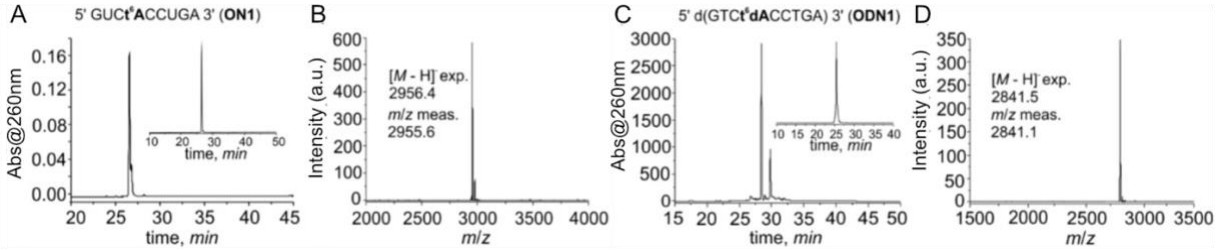


Fig. 2.1. Characterization data of ON1 and ODN1. *(A) HPL-chromatograms of crude (big) and purified (small) ON1. (B) MALDI-TOF mass spectrum of purified ON1. (C) and (D) are respective HPL-chromatograms and MALDI-TOF mass spectrum for ODN1. Modified from ref.*[208]

2.1.2 Physicochemical properties of *aa*^{6}A-modified oligonucleotides

Theoretically, the *aa*^{6}(d)A can exist in 2 conformations, s-*trans* and s-*cis*, based on which sides the intramolecular H-bond is formed between the urea moiety and the purine amine (Fig. 2.2).[216] In particular, the s-*trans* conformation expose the exocyclic amine to the Watson-Crick edge, allowing potential formation of a canonical base-pair with uracil. The s-*cis* conformation, however, has the amino acid branching towards the Watson-Crick edge and therefore, prohibit base-pairing. To investigate the favored conformation of *aa*^{6}A in an RNA duplex, a series of UV-melting experiments were carried out with duplexes formed between **ON1-7** with their universal counterstrand, 5'-CAGUGGACU-3', and compared their results with a corresponding canonical duplex (**D2**). Their results were summarized in Fig. 2.2. In all cases, a single clear melting point was observed, suggesting that only one conformer of *aa*^{6}A existed in the RNA duplexes, regardless of the amino acid entity. The *aa*^{6}A incorporation destabilized the RNA duplex (**D1)**, reducing its melting temperature by 10-15 ˚C, when compared to the A:U canonical duplex (**D2**). When two *aa*^{6}A were incorporated, no duplex formation was observed. For the single t^{6}dA incorporated DNA analog (**D3**), even stronger decrease in melting temperature was observed, reducing the value by more than 20 ˚C (**D3** & **D4**). These data suggested that, in agreement with the tRNA crystal structure,[217] *aa*^{6}A is unlikely to adopt an s-*trans* conformation and therefore, it is incapable of base-pairing and strongly disrupt the duplex stability. We were also interested to find out if *D*- and *L*-stereoisomers of a bulky amino acid, like Phe, would affect the duplex stability differently. However, we observed the same melting temperature for both *L*-Phe and *D*-Phe coupled *aa*^{6}A.

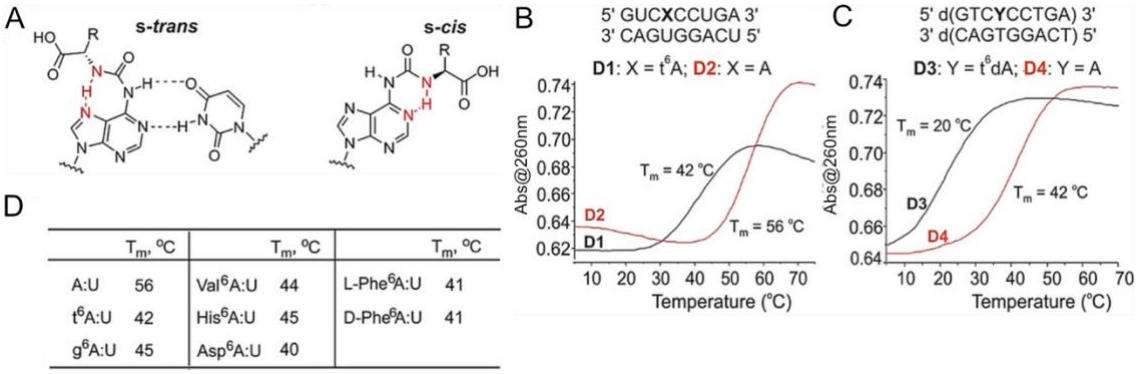


Fig. 2.2. UV-melting of *aa*^{6}(d)A duplexes. *(A) Possible conformations of aa^6(d)A. (B) RNA and (C) DNA UV-melting curves of aa^6(d)A containing duplexes in comparison with their canonical analogs. (D) Summary of melting temperatures determined by UV-melting experiments of RNA duplexes with different aa^{6}A nucleotides incorporated. Adapted from ref.*[208]

Next, we investigated the possibility to incorporate multiple *aa*^{6}A nucleotides as bulges into RNA duplexes (Fig. 2.3). This allows us to build an RNA duplex decorated with various amino acid side chains, a potential construct that allowed the RNA world to transition to an RNA-peptide world. Therefore, we prepared an 18 bp RNA duplex (**D5**) with 3 t^{6}A incorporated in one of the oligos as bulges. To our surprise, the melting temperatures were indistinguishable when compared with its canonical counterpart (**D6**), where the bulges were replaced by unmodified adenosines. We then constructed the duplex **D7**, featuring a bulge of three *aa*^{6}A nucleotides bearing the "catalytic triad" amino acids, namely Ser, Asp and His, directly next to each other. This duplex is reasonably stable with a melting temperature of 45 ˚C. Although we did not observe any catalytic activity for **D7**, we believe that a potential RNA structure putting these modifications in proper positions could make primitive catalytic properties possible on these amino acid-decorated RNAs.

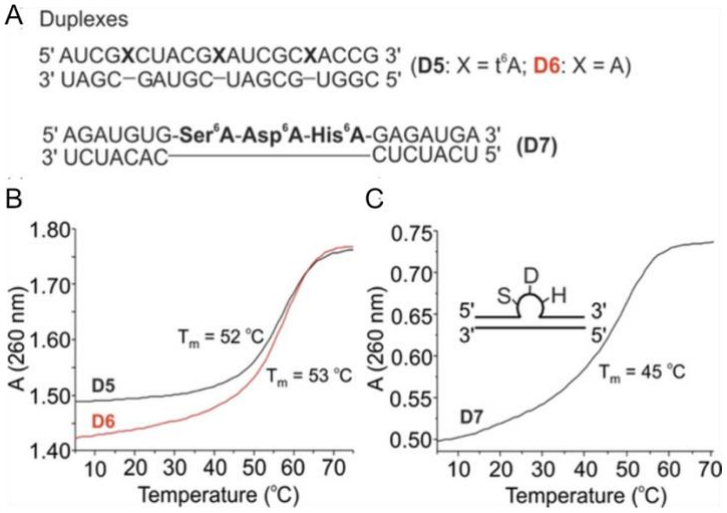


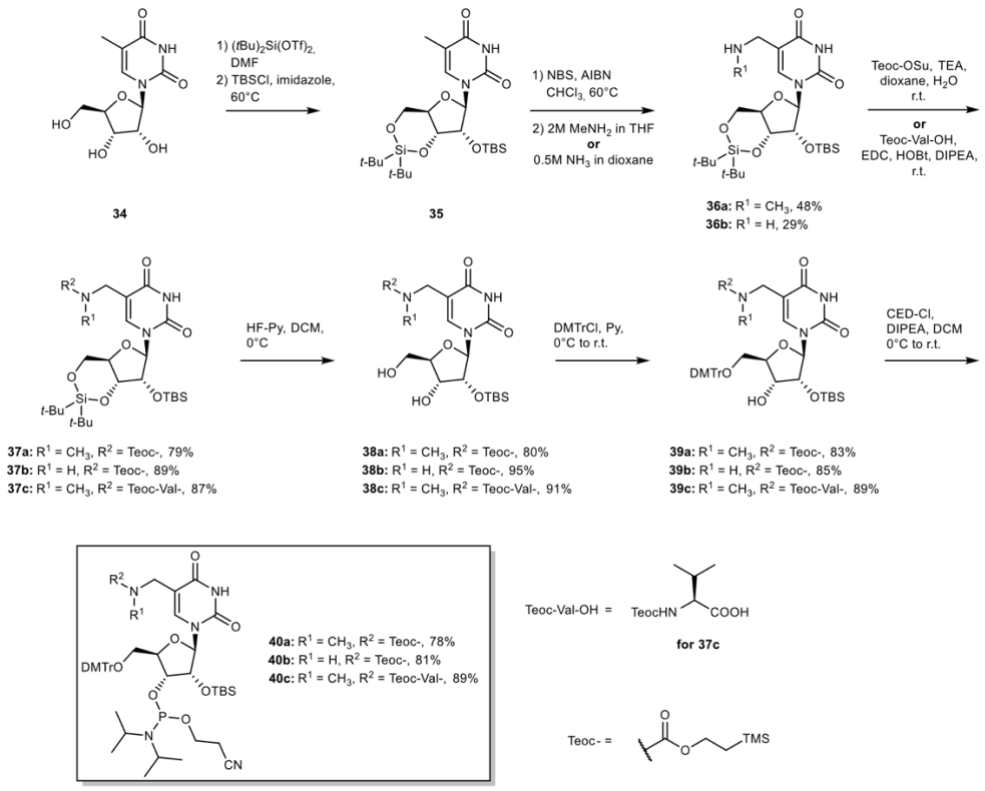
Fig. 2.3. UV-melting of RNA duplexes with bulged *aa*^{6}A. *(A) Sequences of RNA duplexes. (B) UV-melting curves showing similar stability of triply single-nucleotide bulged duplexes of*

*modified **D5** and canonical **D6**. (C) UV-melting curve of **D7** containing a three- nucleotide bulge with aa^{6}A nucleotides charged with different amino acids. Adapted from ref.[208]*

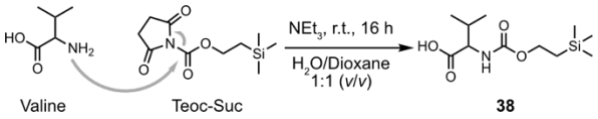
In summary, our data showed that *aa*^{6}A nucleotides are likely to adopt an s-*trans* conformation and hence, unable to form stable base-pairs with uridine when placed in a duplex RNA, hindering them to confer sequence information. However, they allow the incorporation of amino acid into the bulges of RNA duplexes regardless of its sequence context without a significant drop in stability, showcasing a potential way for early RNAs to acquire protein-like functionality by structurally positioning these modified nucleotides, which are formed simultaneously with the canonical ones,[218] to form a more sophisticated active site beyond the canonical building blocks.

2.1.3 Synthesis of (m)nm^{5}U phosphoramidites and respective oligonucleotides

The synthesis of (m)nm^{5}U phosphoramidites (Scheme 2.4) was adapted from a reported pathway for 5-hydroxymethylcytidine.[219] Starting from 5-methyluridine (**34**), the 5' OH, 3' OH and 2' OH were protected similarly by the cyclic di-*tert*-butylsilyl and the TBS groups, respectively (**35**). The exocyclic 5-methyl group was then radically brominated via a Wohl-Ziegler reaction induced by azobisisobutyronitrile (AIBN) and *N*-bromosuccinimide (NBS),[220] followed by an immediate substitution by concentrated methylamine in THF or ammonia in dioxane to form silyl-protected mnm^{5}U (**36a**) or nm^{5}U (**36b**) in one step, respectively. Since methylamine is more nucleophilic than ammonia, the yield of the mnm^{5}U analog is slightly better. To protect the amine or methylamine from the solid-phases synthesis cycle, we coupled it with the 2-(trimethylsilyl)ethoxycarbonyl (teoc) protecting group according to a literature procedure,[221] forming **37a-b**. They were then subjected to standard phosphoramidite synthesis steps as previously described, forming the corresponding mnm^{5}U (**38a-40a**) and nm^{5}U (**38b-40b**) phosphoramidites. We were also interested to see if the peptide coupling reactions between *aa*^{6}A and (m)nm^{5}U would be different after the first amino acid was transferred (see section 2.2). Hence, we also synthesized phosphoramidites for a Val-coupled mnm^{5}U (vmnm^{5}U) by slightly adjusting the synthetic pathway. Teoc-protected valine can be first prepared by reacting free valine with teoc-succinimide under basic condition to form **38**, which can be activated by a conventional amino acid activator, like 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), to couple with the methylaminomethyl group of **36b** to form the valine-coupled analog **37c**. The corresponding phosphoramidite can then be formed following the same procedure without any modification (**38c-40c**). The incorporation of the phosphoramidites proceeded smoothly without any adjustment needed for the solid-phase synthesis cycle. To remove the teoc group, the solid support (CPG beads or High Load Glen UnySupport™ polystyrene beads) was suspended in saturated $ZnBr_2$ solution of *i*PrOH/$MeNO_2$ (1/1, *v*/*v*) at r.t. for 16 h. It was then washed with DCM and dried before continuing to the ammonia cleavage, deprotection and purification steps.



Scheme 2.4. Synthesis of (m)nm^{5}U phosphoramidites. *General strategy for synthesis of (m)nm^{5}U phosphoramidites. Adapted from ref.*[209]



Scheme 2.5. Synthesis of teoc-Val-OH. *Valine can be protected with Teoc-Suc under basic conditions.*[221] *This step was also applied to* ***36a-b*** *to form* ***37a-b*** *in Scheme 2.4.*

Name	Sequence information	Polymer	HPLC	MALDI-TOF	
	Sequence (5'-3')		0-30% B t_R (min)	*m/z* calcd. for $[M-H]^-$	found
ON1	GUC(t^6A)ACCUGA	RNA	26.5	2956.4	2955.6
ODN1	GTC(t^6dA)ACCTGA	DNA	25.1	2841.5	2841.1
ON2	GUC(g^6A)ACCUGA	RNA	25.2	2913.4	2912.8
ON3	GUC(v^6A)ACCUGA	RNA	22.5	2955.5	2954.9
ON4	GUC(h^6A)ACCUGA	RNA	23.7	2993.4	2992.9
ON5	GUC(d^6A)ACCUGA	RNA	21.0	2971.4	2970.9
ON6	GUC(f^6A)ACCUGA	RNA	29.8	3003.5	3002.6
ON7	GUC((*D*-Phe)^{6}A)ACCUGA	RNA	31.2	3003.5	3002.4
ON8	AUCG(t^6A)CUACG(t6A)AUCGC(t6A)ACCG	RNA	31.5	7109.1	7107.8
ON9	AGAUGUG(s^6A)(d^6A)(h^6A)GAGAUGA	RNA	25.3	6042.9	6041.7

Table 2.1. Oligonucleotides described in section 2.1. *Left to right: oligo's nomenclature, sequence information, polymer type and its characterization data including HPLC retention time (gradient 0-30 % buffer B) and MADI-TOF mass spectrometry data. Modifications m^6aa^6A are named after the single letter IUPAC code for the corresponding amino acid. Information for all oligos in this thesis is summarized in the appendix.*

2.2 A prebiotically plausible scenario for peptide synthesis with tRNA modifications

After knowing the basic physicochemical properties of *aa*^{6}A-modified RNAs, we became interested in exploring the roles of amino acid modified-nucleotides on a primitive translation system. We wondered if these modifications could form an alternative template-directed peptide synthesis cycle when put in close proximity with other modifications consist of a free amino group, like nm^{5}U and mnm^{5}U. Such peptide synthesis would feature a more stable urea linkage for peptides and RNAs to co-accumulate in comparison to the ester or acyl phosphate mixed anhydride linkages that we observe now in nature. Our results of this section are published in ref.[209] and redescribed here. (Information on all oligos described in 2.2 is listed in p.41 and section 4.2)

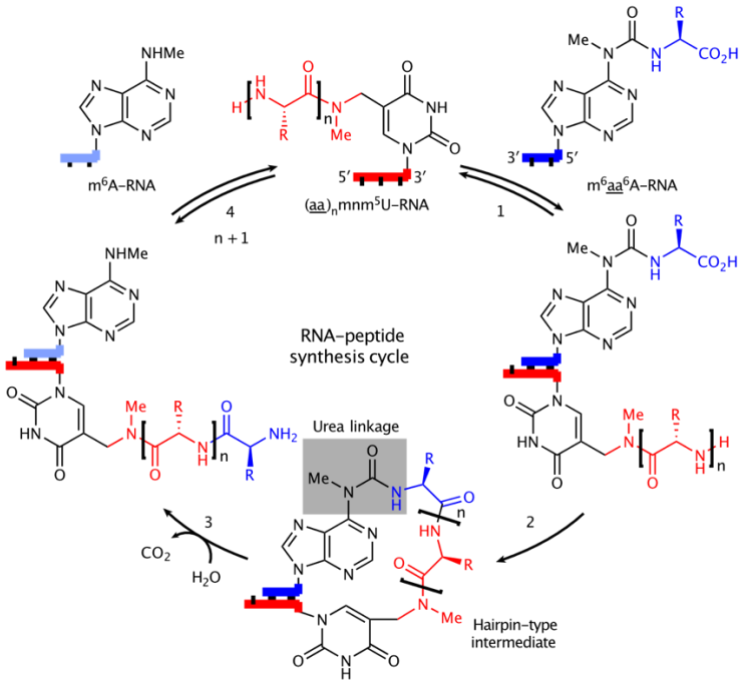


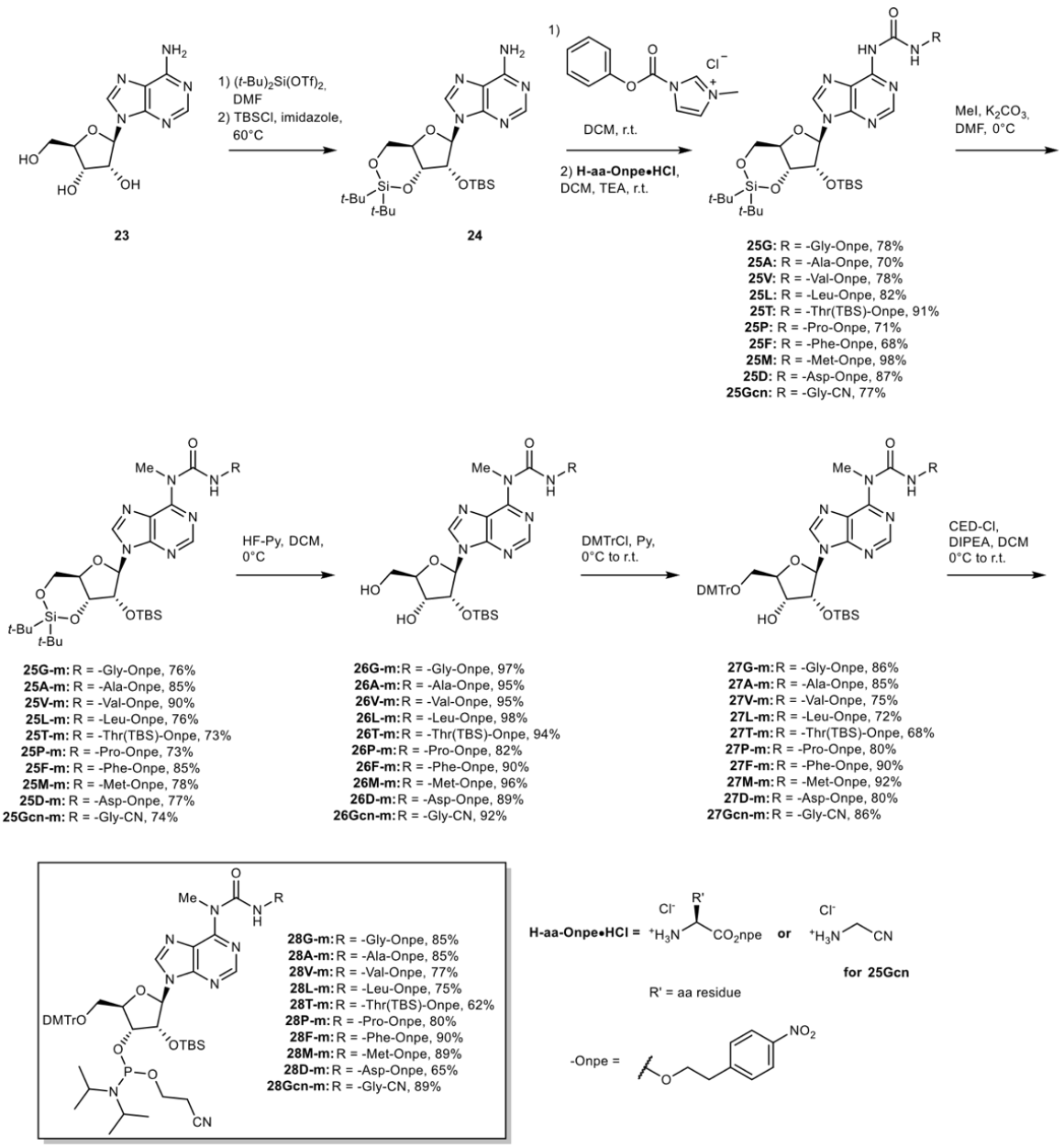
Fig. 2.4. Novel template-directed peptide synthesis cycle. *Peptide synthesis cycle described in this section based on (m)^{6}aa^{6}A and (m)nm^{5}U: 1) hybridization of complementary oligos, 2) peptide coupling assisted by a carboxylic acid activator, 3) cleavage of the (m)^{6}aa^{6}A urea linkage, and 4) dissociation of the reacted m^{6}A strand. Reprinted from ref.[209]*

2.2.1 Peptide coupling on m^6*aa*^{6}A and (m)nm^{5}U modified RNAs

To better explore if the peptide coupling can be affected by the identity of the amino acid charged onto *aa*^{6}A, we decided to expand our library of *aa*^{6}A phosphoramidites (Scheme 2.6). To better facilitate a later urea cleavage step (see section 2.2.2), we decided to add one additional methylation step to our synthesis to produce m^6*aa*^{6}A phosphoramidites (**25[*aa*]** to **25-28[*aa*]-m**).

Next, we synthesized m^6*aa*^{6}A-containing RNA and 2'OMe RNA oligos (**ON10-[*aa*]m** and **ON11-16**) and (m)nm^{5}U-containing oligos (**ON17-[*aa*](m)** and **ON18-21**) to investigate potential template-directed peptide couplings. To quantify the coupling efficiency, we created calibration curves by injecting canonical oligos **CON1-6** with sequences resembling the

modified oligos with incremental volumes into the HPLC (Fig. 2.5). To screen for the best amino acid activator for this purpose, RNA duplexes formed by m^6*aa*^{6}A-modified **ON10-Gm** and (m)nm^{5}U-modified **ON17**, **ON17-m** or **ON17-Vm** (50 µM each) were reacted with selected activators (50 mM) from the literature, EDC/Sulfo-NHS,[222] DMTMM•Cl[222] or MeNC,[223] in 100 mM MES buffer (pH 6) and 100 mM NaCl at r.t. overnight (Fig. 2.6). The results showed that DMTMM•Cl is the best coupling reagent, and primary amines in **ON17** and **ON17-Vm** in general react better than the secondary amine in **ON17-m**. We then furthered our study with DMTMM•Cl and coupled **ON17-m** and **ON17-Vm** with different m^6*aa*^{6}A oligos **ON10-[*aa*]m**. While we did not observe any major yield difference among the amino acids except that m^{6}a^{6}A has an unusual high yield (51%) with mnm^{5}U RNA. A follow-up kinetic analysis revealed that the nature of the amino acid affects the coupling rate, ranging from 0.12±0.02 h^{-1} for m^{6}g^{6}A to >1 h^{-1} for m^{6}f^{6}A. This, however, is likely attributed to the innate bias of the activator DMTMM•Cl. To ensure that the observed mass is indeed the expected RNA-peptide, control experiments were carried out for duplexes of an m^{6}g^{6}A strand and an unmodified strand (**ON10-Gm**:**CON2**), and an m^{6}g^{6}A strand and a teoc-protected vmnm^{5}U strand (**ON10-Gm**:**ON17-teoc**), we observed no reaction in both cases (Fig. 2.7).



Scheme 2.6. Synthesis of m^6*aa*^{6}A phosphoramidites. *Modified from ref.*[209]

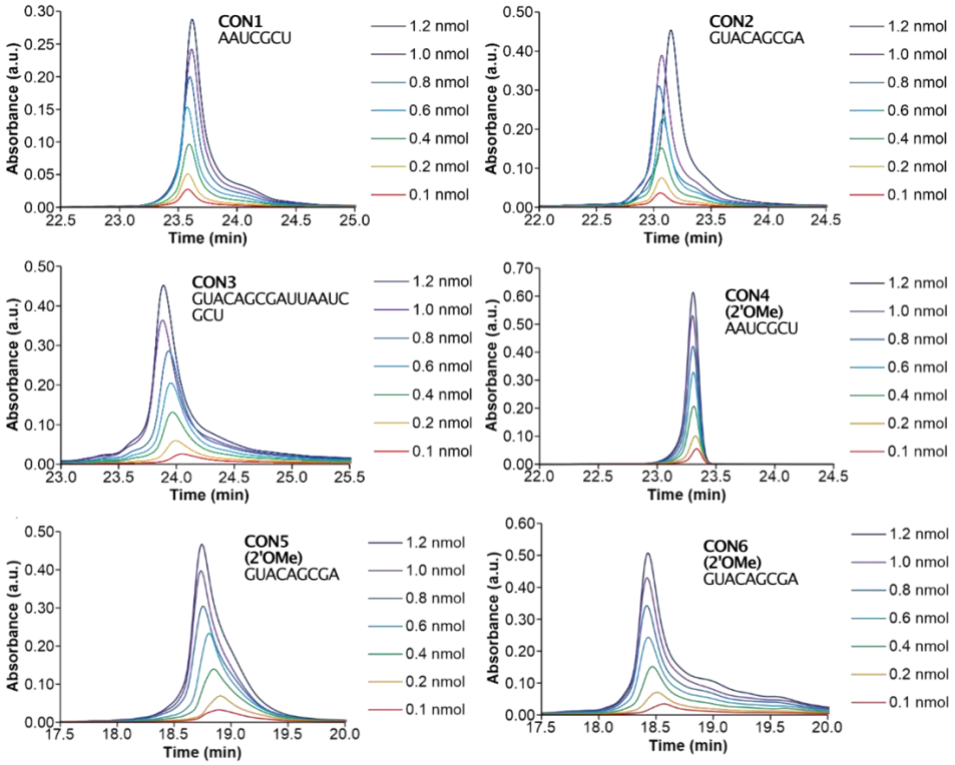


Fig. 2.5. Calibration curves of CON1-6. *Calibration curves of canonical oligos resembling sequences of m^{6}aa^{6}A and (m)nm^{5}U oligos and their products are used to quantify the reactions. Modified from ref.*[209]

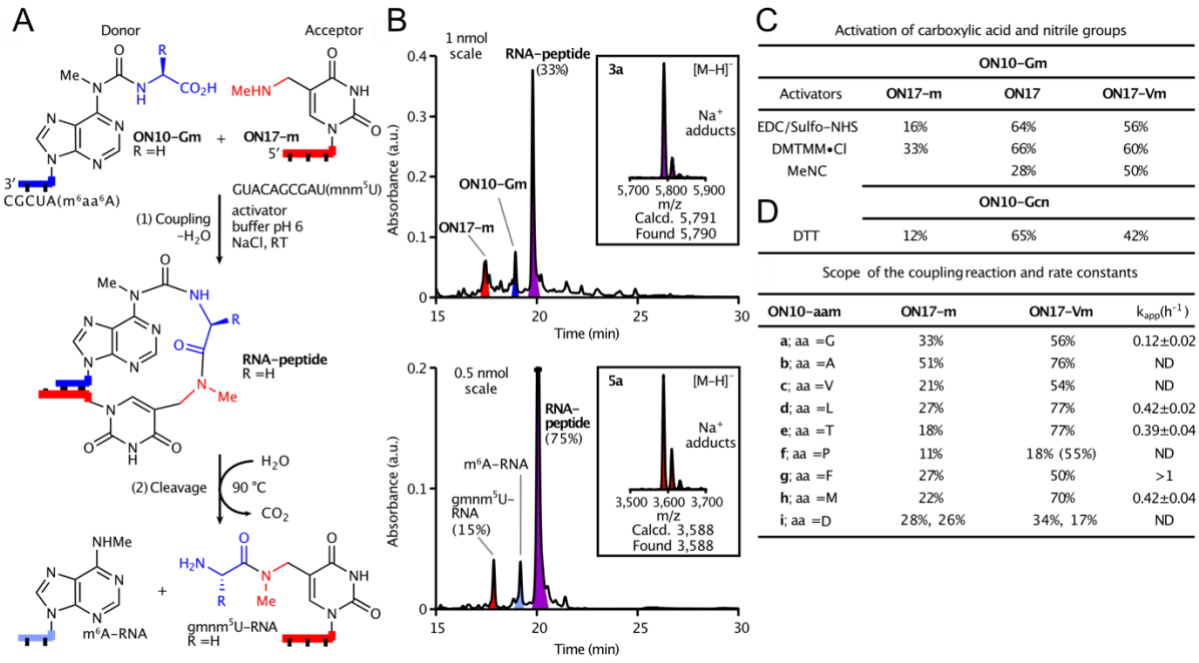


Fig. 2.6. Peptide couplings on RNA duplexes with terminal m^6*aa*^{6}A and (m)nm^{5}U. *(A) Scheme of coupling reaction and subsequent urea cleavage between an **ON10-[aa]m** and **ON17-m**. (B) HPL-chromatograms of the coupling and cleavage steps. (C) Screening of selected amino acid activators from the literature. (D) Coupling yields and kinetics of DMTMM•Cl assisted couplings between **ON10-[aa]m** and **ON17-m,** and EDC/Sulfo-NHS assisted couplings between **ON10-[aa]m** and **ON17-Vm**. Modified from ref.*[209]

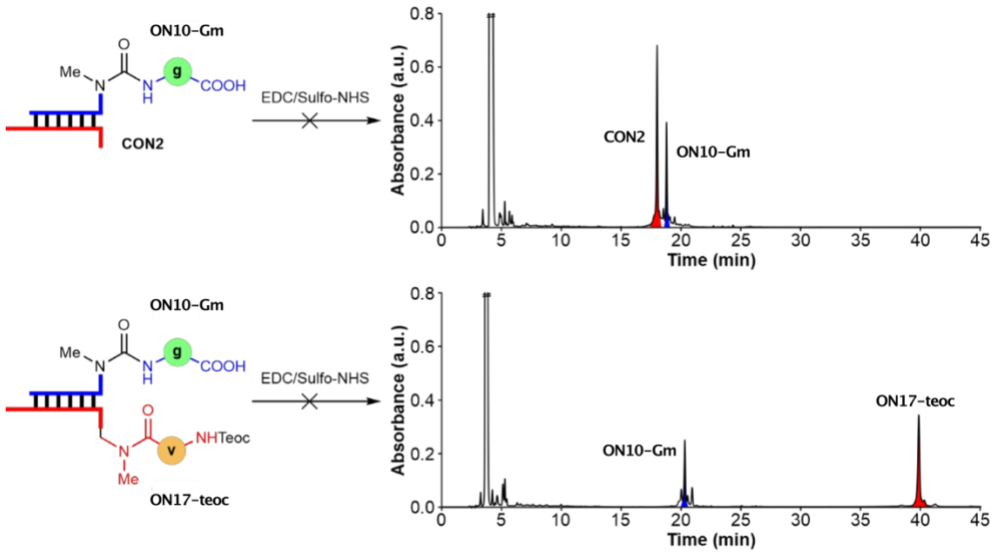


Fig. 2.7. Control experiments for peptide coupling. *Absence of RNA-peptide peaks when reacting m^6g^6A-containing* ***ON10-Gm*** *with (top) unmodified* ***CON2*** *and (bottom) teoc-protected* ***ON17-teoc****. Modified from ref.*[209]

We next investigated how the duplex length would affect the coupling efficiency. To enhance duplex stability, we employed 2'OMe RNAs. We reacted mnm^5U containing **ON20** with m^6aa^6A containing **ON15** and **ON16**, forming duplexes of 4 and 2 nucleotides, respectively (Fig. 2.8). While the yield in general decreased, we still observed detectable product for duplex **ON20:ON16** when we reacted at 0 ˚C and 1 M NaCl.

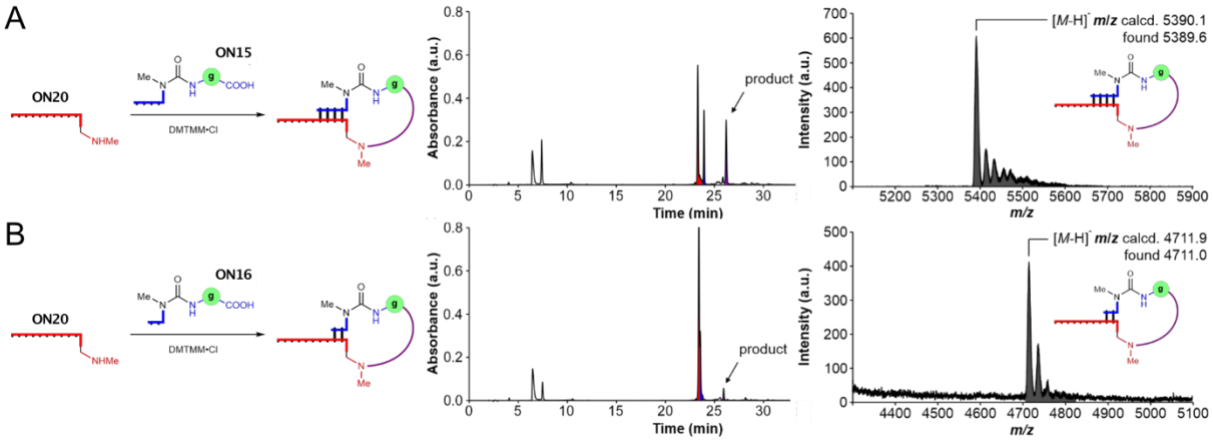
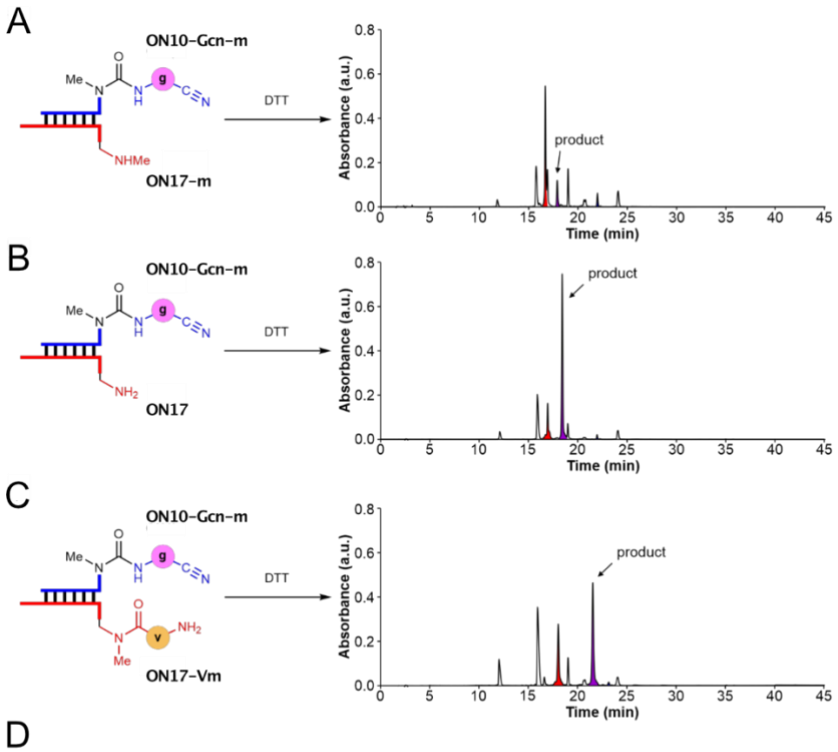


Fig. 2.8. Peptide couplings with shortening duplex. *Coupling reaction of 2'OMe mnm^5U-modified RNA* ***ON20*** *with 2'OMe m^6g^6A-modified RNA (A)* ***ON15*** *at r.t. and 100 mM NaCl, and (B)* ***ON16*** *at 0 ˚C and 1 M NaCl. Modified from ref.*[209]

To explore the prebiotic plausibility of such coupling, we investigated the coupling of an amino nitrile derivative, a recently reported amino acid precursor for peptide formation,[224] of the modified base, $m^6g_{CN}{}^6A$. By synthesizing the phosphoramidite **28Gcn-m** (Scheme 2.6), we prepared **ON10-Gcnm** and reacted it with (m)nm^5U oligos **ON17, ON17-m** and **ON17-**

Vm under described thiol activation conditions: 50 mM dithiolthreitol (DTT) and 100 mM borate buffer (pH 8) at r.t. We observed the formation of desired RNA-peptide hairpins in all cases within a few hours, in which the reaction with **ON17** had the highest yield of 66% using EDC/Sulfo-NHS as activator (Fig. 2.9).



D

$m^6g_{CN}{}^6A$ strand	(m)nm^5U strand	Average Yield ± Error (%)
ON10-Gcn-m; $m^6g_{CN}{}^6A$	**ON17-m**; mnm^5U	12±1
	ON17; nm^5U	65±2
	ON17-Vm; $vmnm^5U$	47±2

Fig. 2.9. Coupling reaction between $m^6g_{CN}{}^6A$ and (m)nm^5U oligos. *Coupling reactions of* ***ON10-Gcn-m*** *with (A)* ***ON17-m****, (B)* ***ON17****, and (C)* ***ON17-Vm****. (D) Summary of yields the described reactions. Modified from ref.*[209]

We next measured the stability of the RNA-peptide products, as the terminal peptide linkage should greatly increase its melting temperature. We compared the melting temperatures of RNA duplex **ON10-Gm:ON17-Vm** and its covalent RNA-peptide hairpin product (Fig. 2.10). The melting temperature of the former was determined to be 30.5 ˚C. However, when we measured the latter, we observed a two-state melting curve, with melting temperatures of 28.4 ˚C and 80.1 ˚C. We, therefore, proposed that the hairpin is more likely to be unfolded in lower temperature and formed an intermolecular RNA-peptide-RNA duplex. To support this idea, we carried out concentration-dependent UV-melting experiments with 3, 5, and 8 µM RNA-peptide, and the result supported this idea by showing that only the first melting temperature is concentration-dependent, whereas the second is not. Combining these results, the second melting temperature of the RNA-peptide indicates the unfolding of the hairpin structure and suggested that the formation of a terminal peptide clamp increased the melting

temperature of this transition phase from 30.5 ˚C of the **ON10-Gm:ON17-Vm** duplex to 80.1 ˚C of a linear, unfolded RNA-peptide-RNA.

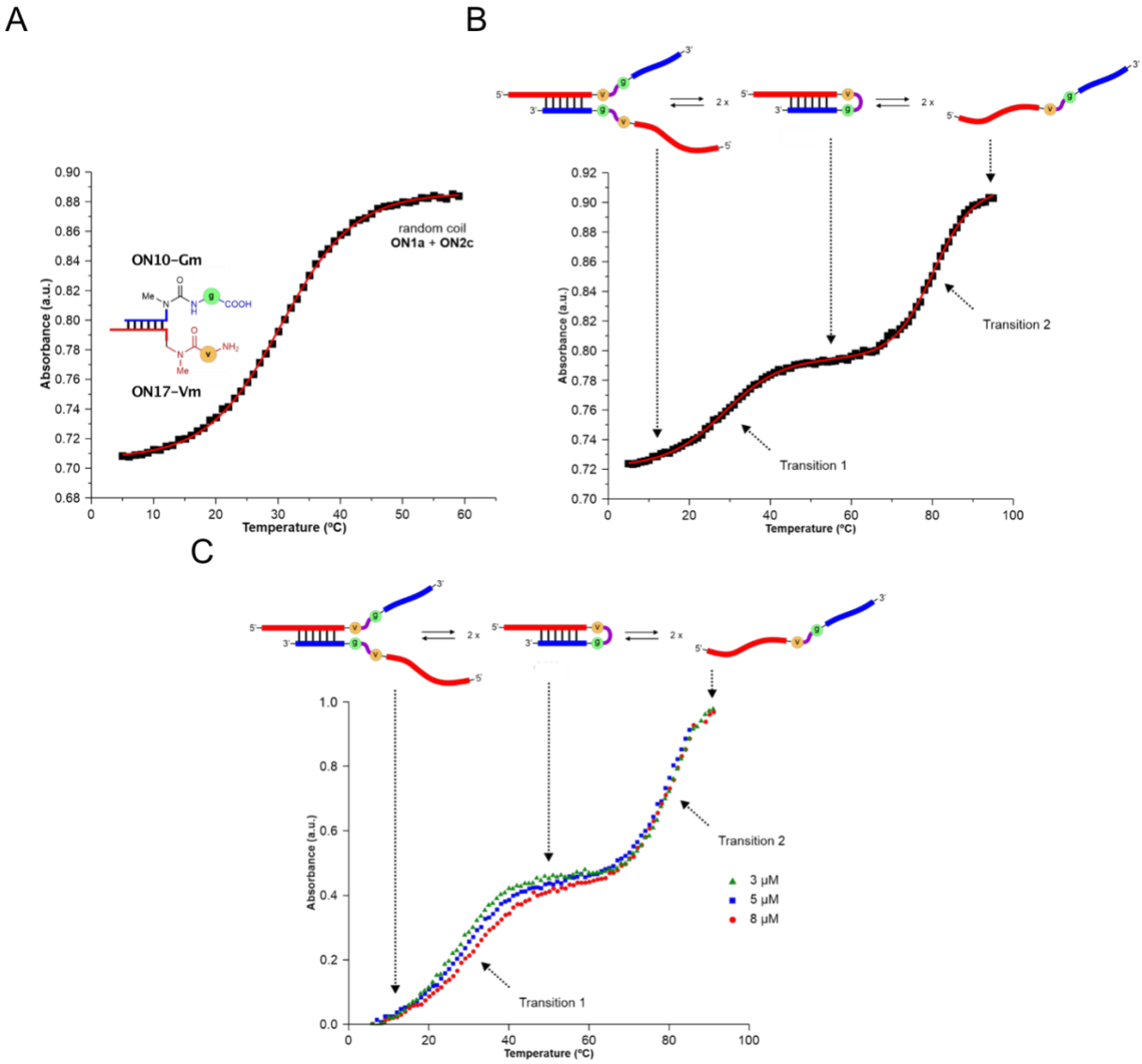
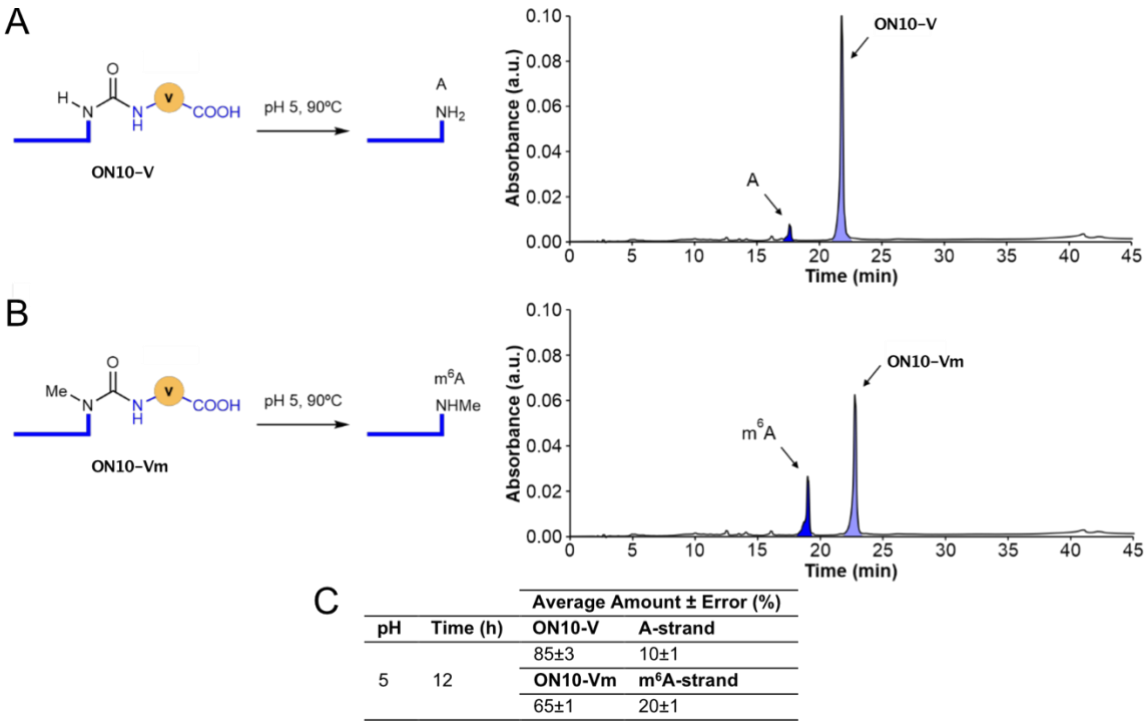


Fig. 2.10. UV-melting curves of duplex ON10-Gm:ON17-Vm and its RNA-peptide hairpin product. *UV-melting experiments of (A) duplex **ON10-Gm:ON17-Vm** (T_m = 30.5 ˚C) and (B) its RNA-peptide hairpin product (T_{m1} = 28.4 ˚C, T_{m2} = 80.1 ˚C). (C) Concentration-dependent melting curves suggested that the structure of first RNA-peptide transition phase is intermolecular. Modified from ref.*[209]

2.2.2 Urea cleavage of RNA-peptide hairpins

At the same time, we also studied the urea cleavage of the m^6*aa*^{6}A moiety to release the reacted oligos as m^{6}A and *aa*-(m)nm^{5}U modified strands. To begin with, we first heated the v^{6}A- and m^{6}v^{6}A-modified oligos, **ON10-V** and **ON10-Vm**, alone in 100 mM acetate buffer (pH 5) and 90 ˚C for 12 h in two separate experiments. Their follow-up HPLC analyses showed that the cleavage of **ON10-Vm** is faster than that of **ON10-V**, which made us decide to use m^6*aa*^{6}A instead of *aa*^{6}A for this study (Fig. 2.11).

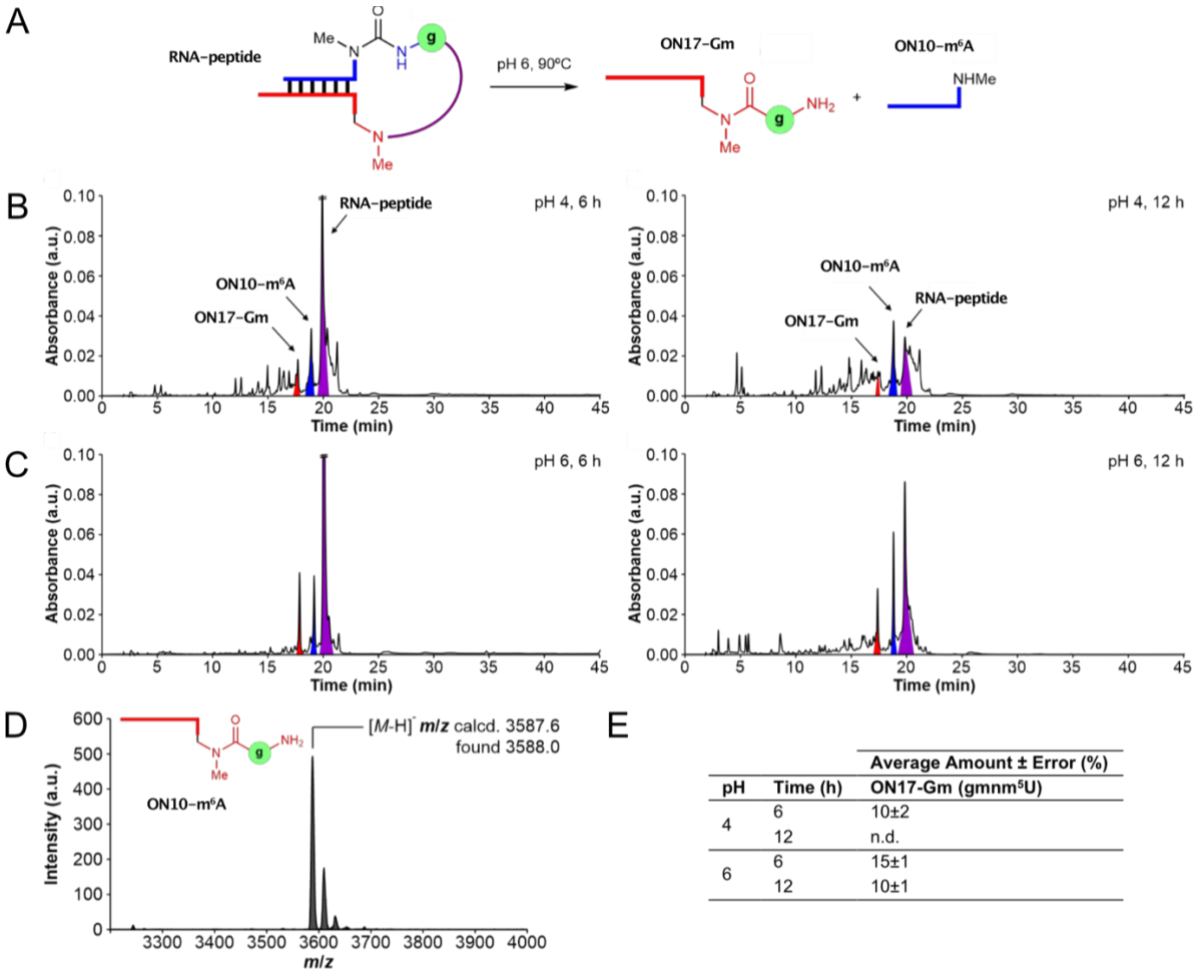


C

		Average Amount ± Error (%)	
pH	Time (h)	ON10-V	A-strand
		85±3	10±1
5	12	ON10-Vm	m⁶A-strand
		65±1	20±1

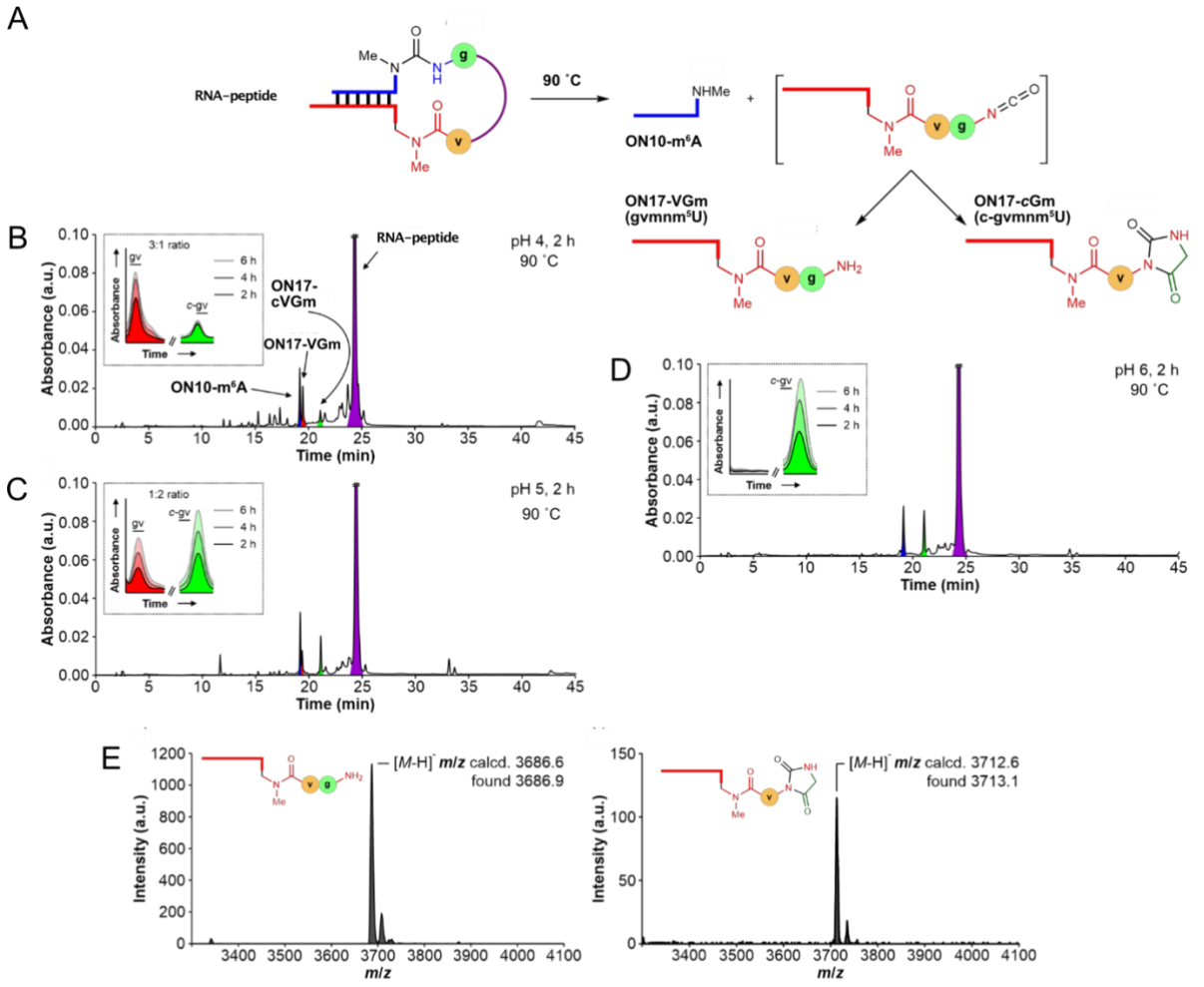
Fig. 2.11. Urea cleavage of (m)^{6}v^{6}A-modified RNAs. *Urea cleavage carried out on single stranded (A)* ***ON10-V*** *and (B)* ***ON10-Vm****. (C) Summary of the yields after the reaction, showing that* ***ON10-Vm*** *has a faster cleavage rate. Modified from ref.*[209]

We then studied the urea cleavage of the RNA-peptide hairpins. To begin with, we screened for the optimal pH and reaction time for the cleavage of the RNA-peptide hairpin formed by **ON10-Gm**:**ON17-m** and discovered that incubation in 100 mM MES buffer (pH 6) for 6 h at 90 ˚C produced the highest yield (15±1%) of the amino acid-transferred **ON17-Gm** (Fig. 2.12). We then investigated the cleavage of a slightly longer RNA-peptide produced from reacting **ON10-Gm**:**ON17-Vm**. Surprisingly, we observed parallel formation of the desired RNA-peptide product (**ON17-GVm)** and its hydantoin analog formed by the cyclisation of the peptide amine and the terminal carboxylic acid of the transferred Gly (**ON17-*c*GVm**) (Fig. 2.13). The hydantoin **ON17-*c*GVm** is considered as an undesired product, as it leads to a dead end of the peptide elongation. We shortly realized that the formation of hydantoin is pH sensitive. When we reacted the RNA-peptide at pH 6, we observed an exclusive formation of **ON17-*c*GVm** with a 25±1% yield. However, when reacted at reduced temperature and acidic conditions, like 60 ˚C and pH 4, we observed the -GVm:-*c*GVm ratio to be optimal at 7:1 after 5 d (Fig. 2.14), indicating that the formation of the hydantoin side products can be controlled by pH, reaction time, and temperature. The low yield of the urea cleavage can be attributed to the degradation of RNA, as a longer smear was observed in the HPL-chromatograms as the reaction time increased.



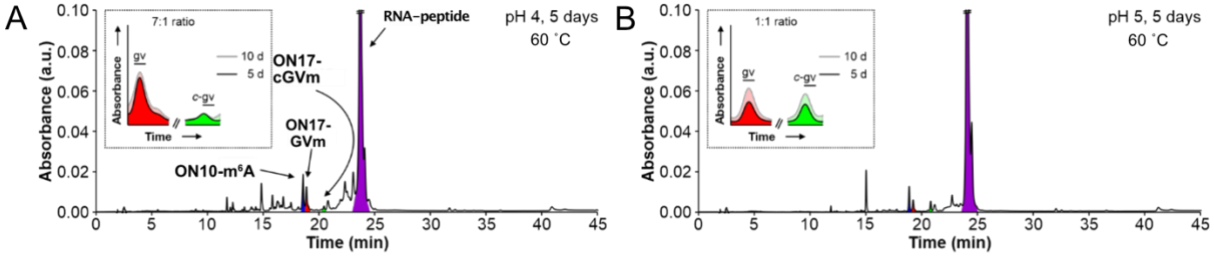
pH	Time (h)	Average Amount ± Error (%) ON17-Gm (gmnm⁵U)
4	6	10±2
	12	n.d.
6	6	15±1
	12	10±1

Fig. 2.12. Urea cleavage of ON10-Gm:ON17-m hairpin. *(A) Reactions at pH4 & 6 (B & C) at 90 ˚C to form* ***ON10-m^{6}A*** *and (D)* ***ON17-Gm****. (E) Summary of* ***ON17-Gm*** *yields. Modified from ref.*[209]



		Average Amount ± Error (%)				
pH	Time (h)	RNA-peptide	ON10-m⁶A	ON17-GVm (gvmnm⁵U)	ON17-cGVm (c-gvmnm⁵U)	Ratio (-GVm/-cGVm)
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	-

Fig. 2.13. Urea cleavage of ON10-Gm:ON17-Vm hairpin at 90˚C. *(A) Reactions at pH4-6 (B-D) to form* ***ON10-m⁶A****, (E)* ***ON17-VGm*** *and (E)* ***ON17-cVGm****. (F) Acidic condition favoring the formation of* ***ON17-VGm****. Modified from ref.*[209]



		Average Amount ± Error (%)				
pH	Time (days)	RNA-peptide	ON10-m⁶A	ON17-GVm (gvmnm⁵U)	ON17-cGVm (c-gvmnm⁵U)	Ratio (-GVm/-cGVm)
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

Fig. 2.14. Urea cleavage of ON10-Gm:ON17-Vm hairpin at 60˚C. *Reactions at pH4 & 5 (A & B). (C) Conditions of 60 ˚C, pH 4, for 10 d favors the formation of* ***ON17-VGm*** *over the hydantoin* ***-cVGm****. Modified from ref.*[209]

2.2.3 Formation of longer peptides on RNAs

We next investigated how the length of the generated peptide affects the coupling reactions. To better produce the RNA-peptides, we designed an on-bead coupling protocol to synthesize various 5'-m^6*peptide*^{6}A-RNA-3' and 3'-*peptide*-mnm^{5}U-RNA-5' oligos efficiently. After the deprotection of the npe or teoc groups from the solid support of the solid-phase synthesis, the beads were suspended in a DMF solution of DMTMM•BF_4 and *N*,*N*-diisopropyl-*N*-ethylamine (DIPEA) together with the desired npe-protected (for m^6*aa*^{6}A) or Boc-protected (for mnm^{5}U) amino acids or peptides at r.t. for at least 1 h. The beads were then washed and dried and submitted to a subsequent npe or Boc deprotection step (1:1 TFA/DCM for 5 min., r.t.) and coupled to the next amino acid/peptide until the desired peptide sequence was formed. The RNA-peptide can then continue to the cleavage, deprotection and purification steps in accordance with the standard RNA preparation procedure as described above.

Having the longer RNA-peptides in hand, we hybridized various m^6*aa*^{6}A oligos with *peptide*-mnm^{5}U oligos and reacted them with the mentioned peptide coupling protocol (Fig. 2.15). While we observed decent coupling yields from 40-60%, we did not observe any selectivity of the coupling reaction based on the peptide length (up to 5+1). In all cases, subsequent urea cleavage (pH 4, 90˚C) afforded dipeptide- to hexapeptide-decorated RNAs in 10-15% yield.

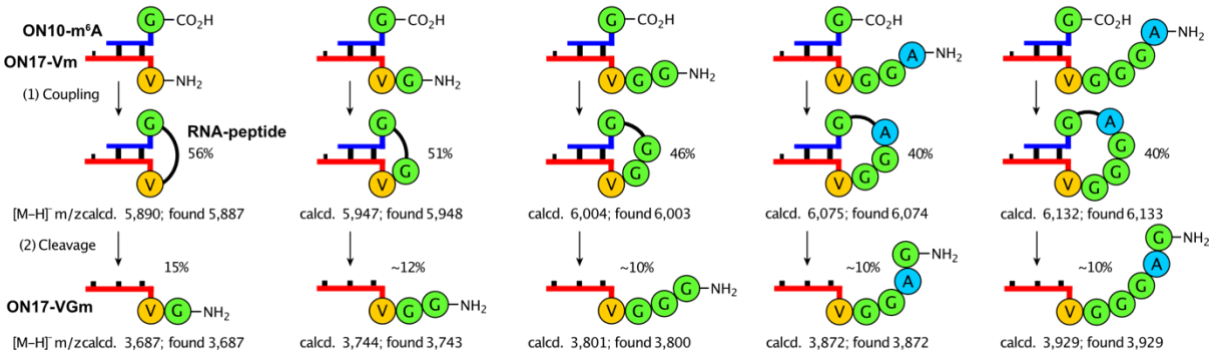


Fig. 2.15. Growth of longer peptide structures on RNA. *Separated peptide coupling and urea cleavage reactions indicated minor effects of peptide length on the synthesis cycle. Modified from ref.*[209]

We then investigated the possibility of forming these longer peptide-decorated RNAs by transferring multiple amino acids at once from an m^6*peptide*^{6}A-modified RNA. Such m^6*peptide*^{6}A moieties could be prebiotically available when peptides are reacted with a nitrosated *N*^{6}-methylurea adenosine (Fig. 2.16). For example, when *N*^{6}-methylurea adenosine was treated with $NaNO_2$ (5% H_3PO_4) and subsequently with triglycine (pH 9.5), the peptide-coupled adenosine nucleoside adenosine *3g*^{6}A can be detected in around 65% yield. While the initial study was carried out on a nucleoside, a systematic study on peptide loading on RNA oligos modified by nitrosated *N*^{6}-methylurea adenosine was later published by our group.[225] In two separate experiments, we coupled and urea-cleaved duplexes formed by an m^6*peptide*^{6}A-modifed and a *peptide*-mnm^{5}U-modifed oligos and obtained the final *peptide*-mnm^{5}U-modifed with similar yields (Fig. 2.16).

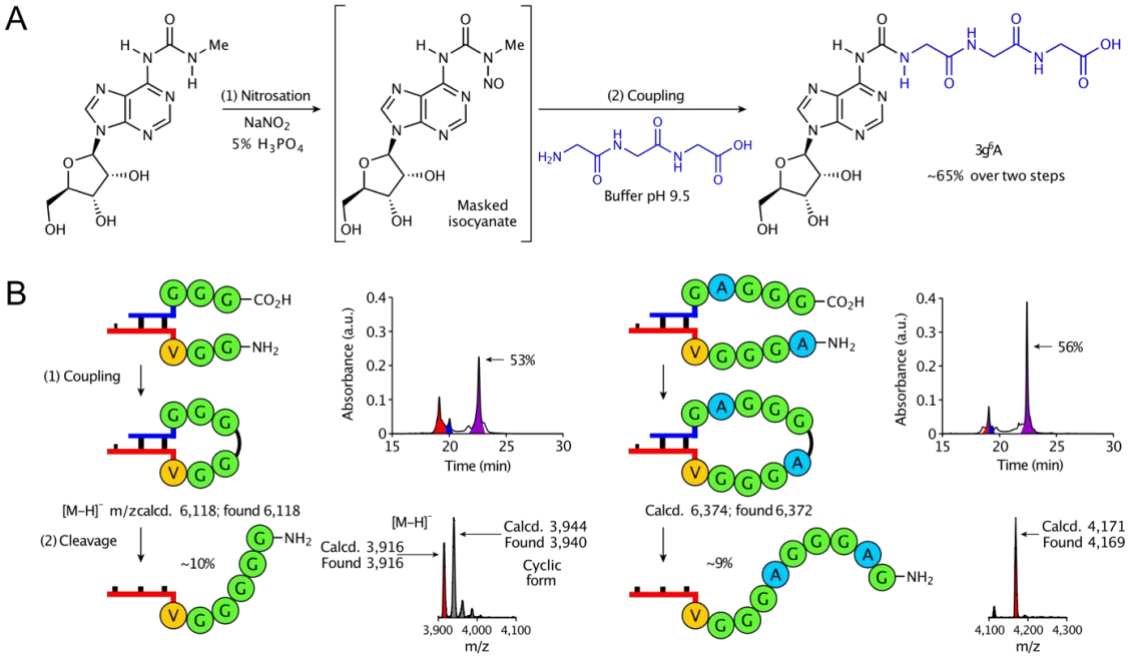


Fig. 2.16. Formation of RNA with long peptide structures by fragment coupling. *(A) Proposed mechanism of peptides captured by nitrosated N^6-methylurea adenosine. (B) Coupling and cleavage reactions of m^6peptide^{6}A and peptide-mnm^{5}U modified duplexes. Modified from ref.[209]*

2.2.4 Multiple peptide growth on a single RNA structure

We next investigated if multiple peptide growth can happen at different positions on a single RNA template simultaneously (Fig. 2.17). We hybridized a single 21mer RNA oligo modified by a middle gmnm^{5}U and a terminal nm^{5}U (**ON18**) with two short RNA oligos modified by a terminal m^{6}v^{6}A (7mer **ON10-Vm** and 10mer **ON11**). After reaction with DMTMM•Cl, we observed peptide bond formation happened successfully on both reaction sites with a yield of 35%. In a separate setup, we hybridized a 22mer RNA oligos, whose 5' and 3' ends are modified by an m^{6}v^{6}A and an m^{6}g^{6}A, respectively (**ON12**), with two short RNAs with one incorporated a middle vmnm^{5}U (**ON19**) and the other a terminal mnm^{5}U (**ON20**). With a similar yield, we observed the successful formation of the three-strand product linked by two peptide bridges.

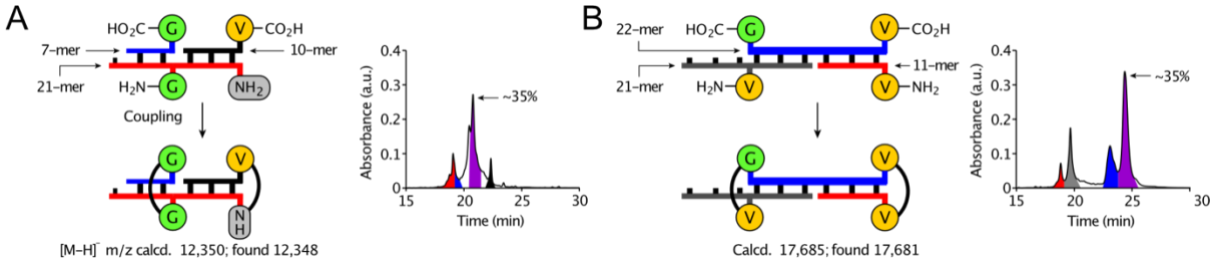


Fig. 2.17. Parallel peptide syntheses on one template. *Two examples (A & B) showcasing simultaneous peptide couplings on a single template. Modified from ref.[209]*

2.2.5 Effects of base pairing on peptide coupling reactions

To confirm that the peptide coupling is sensitive to the RNA's structure, we designed two competition experiments (Fig. 2.18). In the first one, we incubated a terminal vmnm^{5}U modified oligo (**ON17-Vm**) with two counterstrands of different length, with the shorter one being m^{6}g^{6}A modified (**ON10-Gm**), and the longer one being m^{6}v^{6}A modified (**ON13**). We detected, after the coupling reaction, an exclusive formation of the RNA-peptide product from the longer duplex **ON17-Vm**:**ON13** (49% yield) but not the shorter **ON17-Vm**:**ON10-Gm**, suggesting that the more stable duplex has a higher coupling preference. In the second experiment, we incubated the same **ON17-Vm** with an m^{6}l^{6}A modified complementary strand (**ON10-Lm**) and an m^{6}g^{6}A modified strand with a mismatch sequence (**ON21**). After the coupling, we detected only the RNA-peptide from the complementary duplex **ON17-Vm:ON10-Lm** (65% yield) but not from the mismatch **ON17-Vm:ON21**. This indicates that the reaction is catalyzed by the proximity effect when the terminal m^6*aa*^{6}A and *aa*mnm^{5}U modifications are brought together by the formation of a stable RNA duplex.

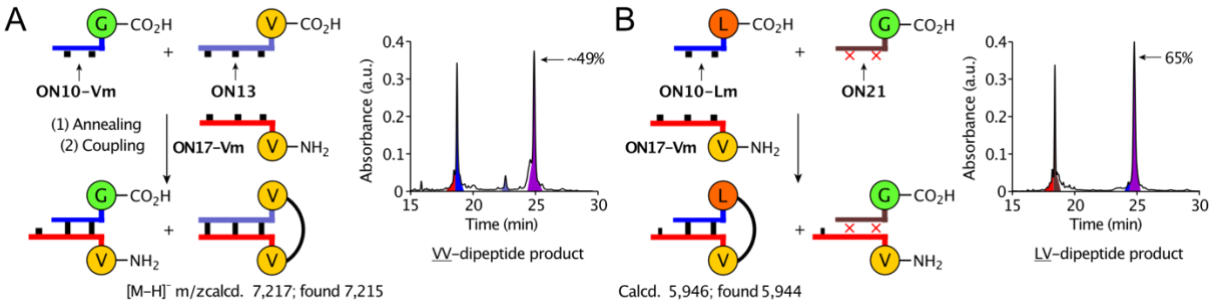


Fig. 2.18. Competition experiments on effects of base pairing. *Competitions (A) between longer and shorter m^{6}aa^{6}A counterstrands and (B) between complementary and mismatch sequence showcased the coupling reaction's sensitivity towards duplex stability and sequence specificity, respectively. Modified from ref.*[209]

2.2.6 Continuous peptide synthesis on RNA

We finally explored if multiple cycles of coupling and urea-cleavage peptide synthesis can undergo continuously with minimal separation of reaction materials. To improve the stability against RNA in-line hydrolysis during the urea cleavage step, we replaced the canonical A, U, C, G nucleotides in **ON17-Vm** by their ribose-modified 2'OMe analogs, forming the 2'OMe RNA **ON20**. In between coupling and urea-cleavage cycles, we introduced a size-exclusion membrane filtration step to remove excessive DMTMM•Cl and its reacted residues. In an initial experiment, we reacted **ON20** with **ON10-Gm** in the first and second cycles, and then with **ON10-Fm** in the third cycle. After the second and third cycle, a small sample of the reaction mixture (~0.5 nmol scale) was separated and analyzed by HPLC (Fig. 2.19). We observed an overall yield of around 18% for the formation of the desired *aa*-mnm^{5}U 2'OMe RNA **ON20-GGm** after the second cycle, and the **ON20-FGGm** in about 10% overall yield after the final cycle.

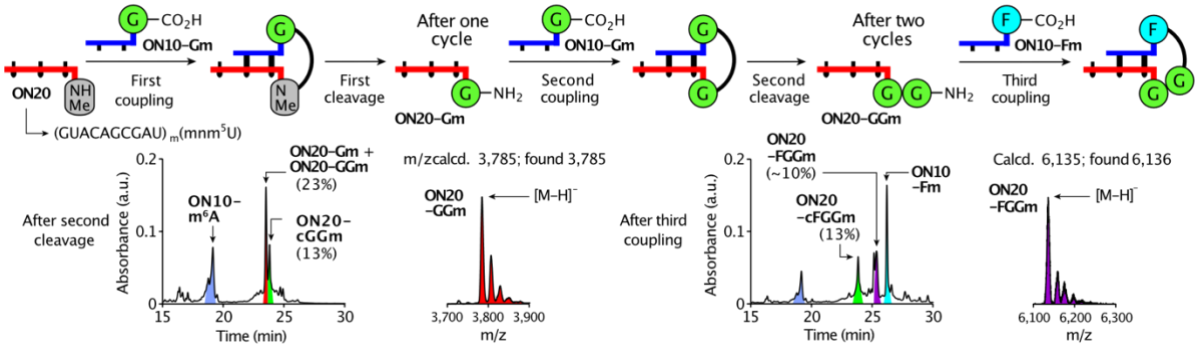


Fig. 2.19. Continuous peptide synthesis on RNA. *The mnm^{5}U modified 2'OMe RNA oligo (**ON20**) was reacted with **ON10-Gm** twice and with **ON10-Fm** once in a continuous manner. The final desired product, **ON20-FGGm**, was detected with an overall yield of about 10%. Modified from ref.*[209]

Taking all into account, our results in this section introduced a novel template-directed peptide synthesis model in an RNA-peptide world. By incorporating tRNA modified nucleotides m^6*aa*^{6}A and (m)nm^{5}U into RNA oligos, we showcased a peptide synthesis cycle by transferring the loaded amino acid from an m^6*aa*^{6}A nucleotides to an (m)nm^{5}U nucleotides via peptide coupling and urea cleavage. We showed that this peptide coupling is sensitive towards the RNA structure, and preferentially reacts with the more stable duplex with a matching sequence. We also showed the system's capability to continuously synthesize peptide with minimal purification. However, it is worth mentioning that this system shows almost no selectivity towards different amino acids or peptides being coupled. In our kinetic study of the coupling reactions, where a sample was isolated and injected to the HPLC every 2 h from the reactions, we observed no difference in the final yield of the formed RNA-peptide hairpins, regardless of the amino acid(s) loaded on the *aa*^{6}A (Fig. 2.20). Only a rate difference was observed for specific amino acids in the initial hours of the reactions. To create the first genetic code dictionary for translation, a selectivity would be required that can select peptides based on their innate physicochemical properties.

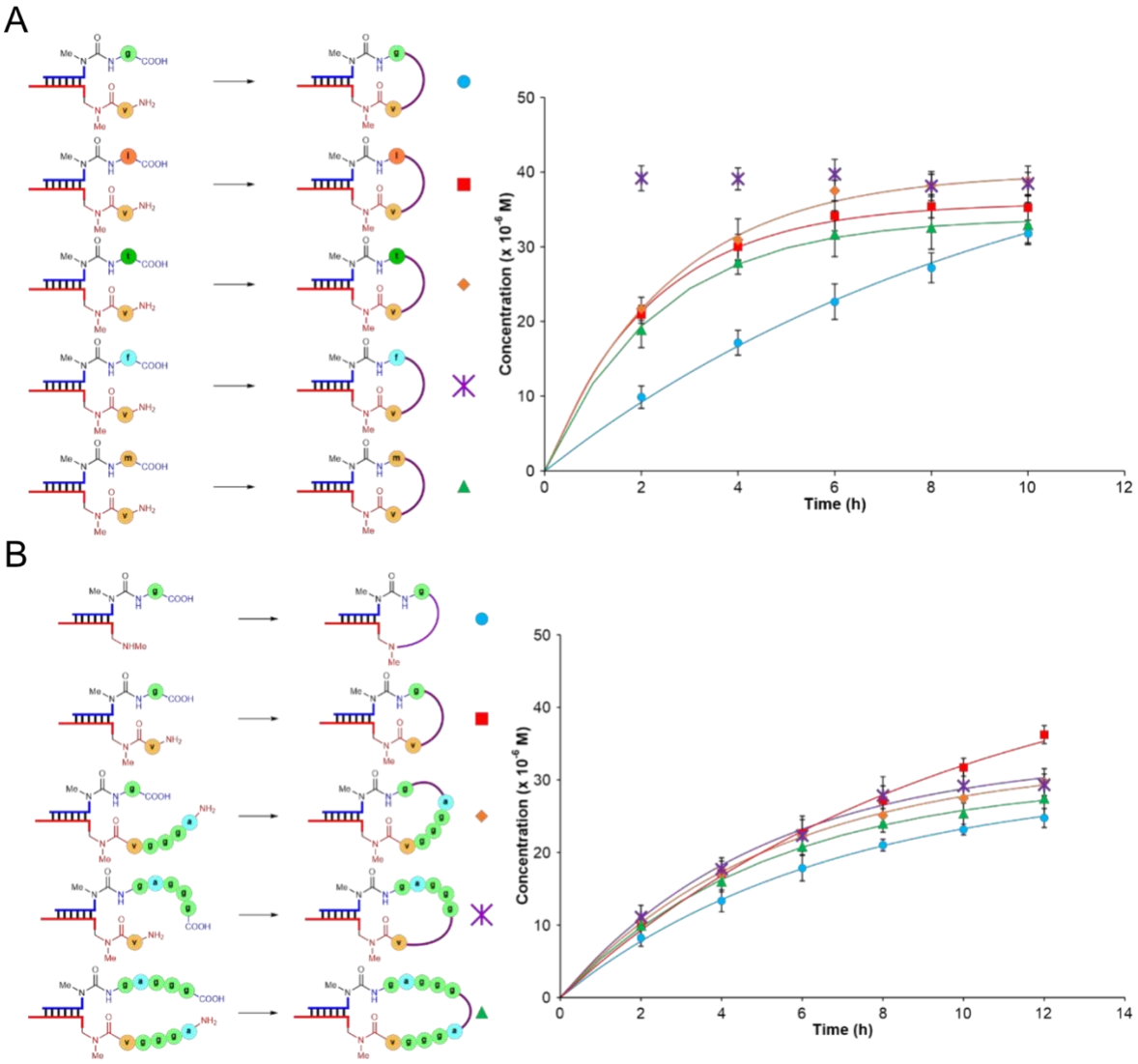


Fig. 2.20. Reaction rates of various peptide coupling reactions. *(A) While significant reaction rate difference can be observed for shorter peptides, their final yields are very similar. (B) Reaction rate difference was not observed in the coupling reaction of longer peptides. Modified from ref.*[209]

Name	Sequence information	Polymer	HPLC	MALDI-TOF	
	Sequence (5'-3')		0-40% t_R (min)	*m/z* calcd. for $[M-H]^-$	found
ON10-Gm	(m^6g^6A)AUCGCU	RNA	18.8	2277.4	2278.4
ON10-Am	(m^6a^6A)AUCGCU	RNA	20.2	2291.4	2290.0
ON10-Vm	(m^6v^6A)AUCGCU	RNA	22.2	2319.4	2317.8
ON10-Lm	(m^6l^6A)AUCGCU	RNA	24.3	2333.4	2331.6
ON10-Tm	(m^6t^6A)AUCGCU	RNA	18.9	2321.4	2320.0
ON10-Pm	(m^6p^6A)AUCGCU	RNA	18.0	2317.4	2316.8
ON10-Fm	(m^6f^6A)AUCGCU	RNA	24.5	2368.4	2365.4
ON10-Mm	(m^6m^6A)AUCGCU	RNA	23.2	2351.4	2350.4
ON10-Dm	(m^6d^6A)AUCGCU	RNA	17.2	2335.4	2334.3
ON10-Gcnm	(m^6gcn^6A)AUCGCU	RNA	21.2	2258.4	2258.5
ON10-V	(v^6A)AUCGCU	RNA	20.6	2305.4	2302.2
ON11	(m^6v^6A)CUAUUGAGU	RNA	22.3	3300.5	3301.1
ON12	(m^6v^6A)AUCGCUGUACCCUAUUGAGU(m6g6A)	RNA	23.1	7231.0	7233.7
ON13	(m^6v^6A)AUCGCUGUAC	RNA	23.2	3604.6	3603.4
ON14	(m^6g^6Am)AUCGCU	2'OMe RNA	23.8	2375.5	2374.4
ON15	(m^6g^6Am)AUCG	2'OMe RNA	23.6	1736.4	1735.1
ON16	(m^6g^6Am)AU	2'OMe RNA	23.1	1058.2	1058.2
ON17-m	GUACAGCGAU(mnm^{5}U)	RNA	17.4	3530.5	3529.7
ON17	GUACAGCGAU(nm^{5}U)	RNA	17.8	3516.5	3515.9
ON17-Vm	GUACAGCGAU(vmnm^{5}U)	RNA	18.6	3629.6	3627.2
ON17-teoc	GUACAGCGAU(Teoc-vmnm^{5}U)	RNA	37.7	3773.7	3776.9
ON18	GUACAGCGAU(gmnm^{5}U)ACUCAAUAG(nm^{5}U)	RNA	18.7	6806.0	6806.4
ON19	GUACAGCGAU(vmnm^{5}U)ACUCAAUAGG	RNA	19.9	6858.0	6857.7

ON20	GUACAGCGAU(mnm^{5}U)	2'OMe RNA	23.0	3670.5	3670.4
ON21	(m^{6}g^{6}A)AGCCCU	RNA	19.5	2276.4	2275.7
CON1	AAUCGCU	RNA	23.6	2162.3	2162.0
CON2	GUACAGCGA	RNA	23.1	3487.5	3486.9
CON3	GUACAGCGAUUAAUCGCU	RNA	23.9	5712.8	5711.7
CON4	AAUCGCU	2'OMe RNA	23.3	2261.6	2260.1
CON5	GUACAGCGA	2'OMe RNA	18.8	4772.7	4772.8
CON6	GUACAGCGAUUAAUCGCU	2'OMe RNA	18.6	6998.0	6995.1

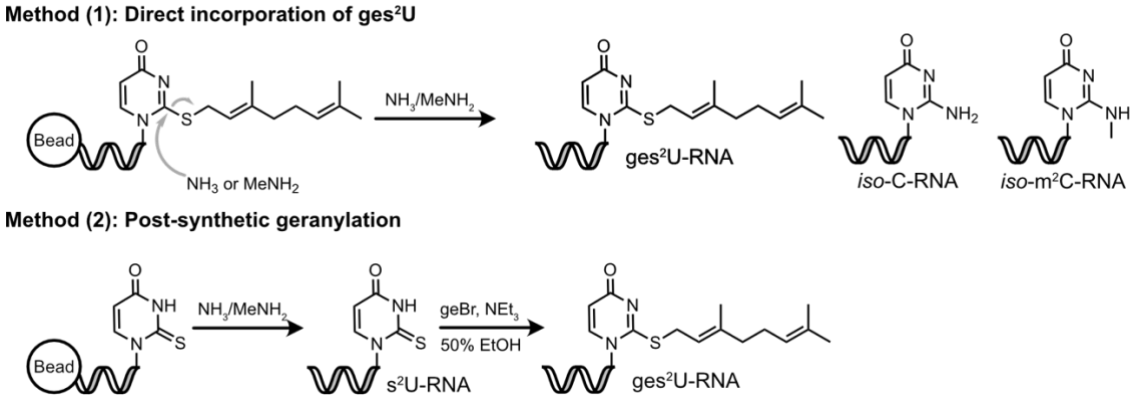
Table 2.2. Oligonucleotides described in section 2.2. *Left to right: oligo's nomenclature, sequence information, polymer type and its characterization data including HPLC retention time (gradient 0-40 % buffer B) and MADI-TOF mass spectrometry data. Modifications m^{6}aa^{6}A are named after the single letter IUPAC code for the corresponding amino acid. Information for all oligos in this thesis is summarized in the appendix.*

2.3 A protocellular model for a primordial translation system

A primitive correspondence between specific RNAs and amino acids was very likely to be founded on their intrinsic physicochemical properties. We wondered if lipophilic RNAs and peptides can be co-recognized by phase separation. RNA modifications, like *S*-geranyl-2-thiouridine (ges^{2}U) and its analogs, are characterized by a long terpene moiety and serve as promising molecular fossils for RNA lipidation. Nowadays, these modifications serve as identity determinants for aaRS to distinguish structurally similar tRNAs.[226] Thus, it is plausible that they were also relevant to a primitive co-selection of amino acids and RNAs. Here, we explored the idea that ges^{2}U-decorated oligos can be anchored onto liposome membranes. By doing so, we also hypothesized that the aforementioned template-directed peptide synthesis can become chemoselective when the microenvironment of the coupling reaction changes. The results of this chapter are uploaded to a preprint server.[227] (Information on all oligos described in 2.3 is listed in p.58 and section 4.2)

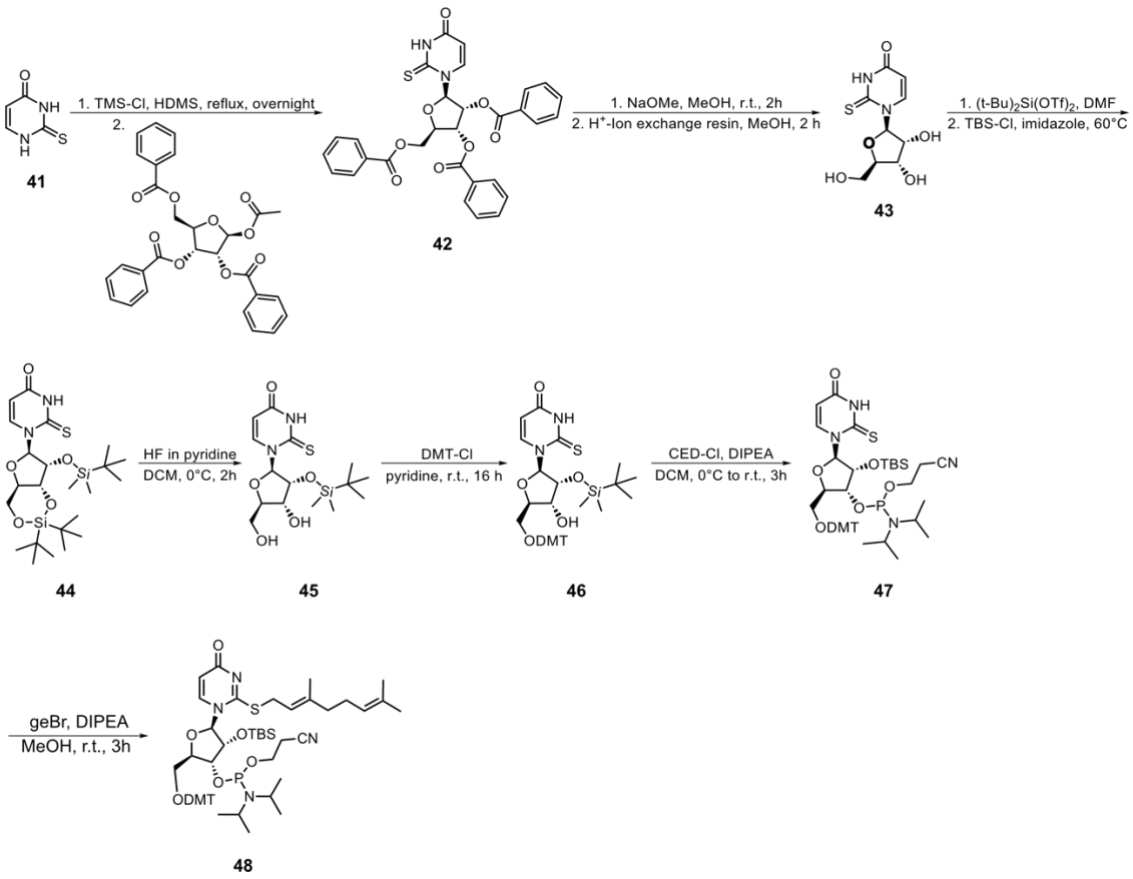
2.3.1 Synthesis of RNAs containing multiple ges^{2}U units

The synthesis of ges^{2}U-modified oligonucleotides is widely discussed in the literature.[228-230] However, all of the published results only described the synthesis of RNA oligos containing one ges^{2}U unit. The ges^{2}U incorporation can be achieved by 2 methods: (1) direct incorporation of a ges^{2}U phosphoramidite into the solid-phase RNA synthesis, or (2) incorporation of 2-thiouridine (s^{2}U) followed by a post-synthetic geranylation with geranyl bromide (geBr) in 50% EtOH and NEt_3 to ges^{2}U (Scheme 2.7). The major problem of (1) is that during the deprotection and cleavage from the solid support, a nucleophilic substitution was detected by ammonia or methylamine, replacing the geranylthiol moiety and forming *iso*-C or *iso*-m^{2}C as side products.[228] On the other hand, method (2) avoids potential side reactions, yet requires an additional purification step.



Scheme 2.7. Methods for synthesis of ges^{2}U-containing RNAs. *(Top) Direct incorporation producing the desired ges^{2}U strand, together with side products. (Bottom) Post-synthetic geranylation avoids side products, but requires additional purification.*

To find out the best method for the incorporation of multiple ges^{2}U nucleotides into RNAs, s^{2}U phosphoramidite was synthesized in multigram scale and a portion was converted into ges^{2}U phosphoramidite according to pathways reported in the literature (Scheme 2.8).[231,232] Briefly, 2-thiouracil (**41**) was first nucleosidated onto protected ribose by classic Vorbrüggen chemistry (**35**). After deprotection, the s^{2}U nucleoside (**42**) was silylated and proceeded through normal phosphoramidite synthetic pathway to form the s^{2}U phosphoramidite (**43-47**), which can be converted into the ges^{2}U phosphoramidite by geranylation (**48**). Both the s^{2}U and ges^{2}U phosphoramidites were incorporated into oligos successfully with standard synthesis cycle without any adjustment.



Scheme 2.8. Synthesis of (ge)s^{2}U phosphoramidites. *Synthetic pathway to produce s^{2}U and ges^{2}U phosphoramidites according to literature.*[231,232]

We first synthesized a 5'-U$_5$(ges^{2}U)-3' oligo (**ON22-geS**) with method (1). As expected, we observed partial conversion of the incorporated ges^{2}U into *iso*-C and *iso*-m^{2}C by HPLC, leaving only ~10% of **ON22-geS** (Fig. 2.21). However, when we synthesized the double ges^{2}U-modified oligo, 5'-U$_4$(ges^{2}U)$_2$-3' (**ON23-geS**), from its s^{2}U precursor (**ON23-S**) using method (2), we observed a clean HPLC with most of the **ON23-S** fully converted to **ON23-geS** (Fig. 2.22). Due to its easy purification and higher yield, we concluded that method (2) is better for synthesizing RNAs with multiple ges^{2}U nucleotides and continued our study with this protocol.

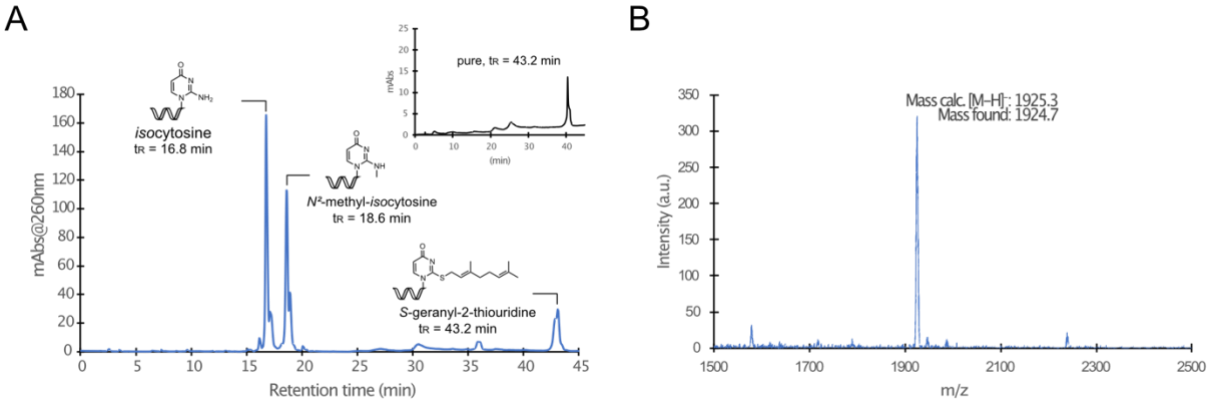


Fig. 2.21. Synthesis of ges^{2}U-RNA with direct incorporation. *(A) HPL-chromatogram (0-40% buffer B) of method (1) showed conversion to iso-C and iso-m^{2}C side products, producing ges^{2}U-RNA in low yield. (B) Mass spectrum of the purified* ***ON22-geS****. Adapted from ref.*[227]

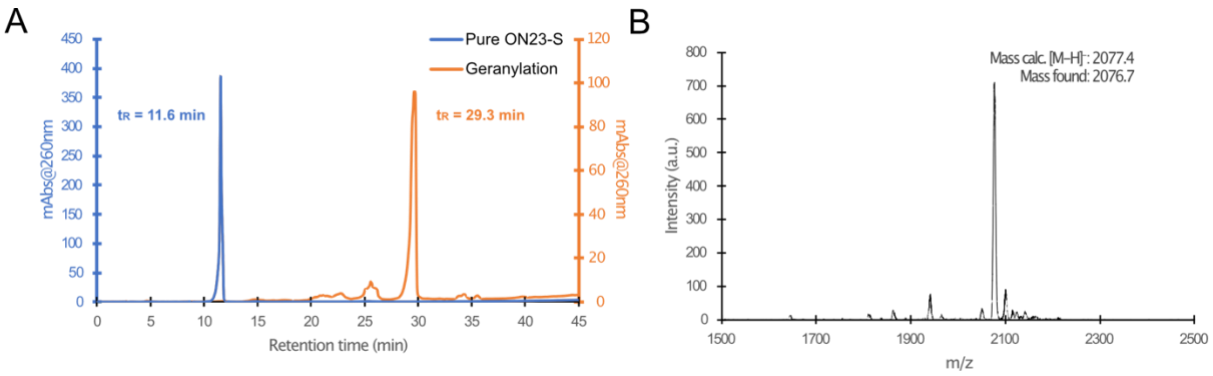


Fig. 2.22. Synthesis of ges^{2}U-RNA with post-synthetic geranylation. *(A) Overlay of HPL-chromatograms (0-80% buffer B) of pure* ***ON23-S*** *(blue) and crude* ***ON23-geS*** *after geranylation (orange), showing almost no side product. (B) Mass spectrum of the purified* ***ON23-geS****. Modified from ref.*[227]

2.3.2 Lipidation of ges^{2}U-modified RNAs

We next investigated the lipidation capability of ges^{2}U-modified RNAs. For this purpose, we synthesized a series of 5'FAM-labeled variants of homo-U oligos, such as 5'-FAM-U$_5$(ges^{2}U)-3' (**ON22-5FAM-geS**) and 5'-FAM-U$_4$(ges^{2}U)$_2$-3' (**ON23-5FAM-geS**). We then prepared zwitterionic Egg PC liposome by lipid rehydration[233] and stained them by Nile Red™ (0.1% mol.). RNA oligos (10 µM) were added to the suspension (10 mM Egg PC, 100 mM MES buffer pH 6, 100 mM NaCl) and immediately analyzed by confocal microscope (Fig. 2.23). As expected, the non-geranylated reference oligo 5'-FAM-U$_4$(s^{2}U)$_2$-3' (**ON23-5FAM-S**) did not show any localization on liposome surface due to the absence of the fluorescence signal. However, only a single substitution of U to ges^{2}U already allowed **ON22-5FAM-geS** to weakly associate with the liposome surface (Fig. 2.24). This observation was enhanced when double ges^{2}U modified **ON23-5FAM-geS** was added to liposome, where we could observe a strong and exclusive FAM emission localized on the liposome surface. We next separated the two ges^{2}U units in the sequence, and even observed weakened lipid association in 5'-FAM-U$_2$(ges^{2}U)U$_2$(ges^{2}U)-3' (**ON24-5FAM-geS**) (Fig. 2.24). These results suggested that ges^{2}U

nucleotides can serve as lipid-anchoring agents for RNAs to phase separate from an aqueous solution.

Liposomes can enrich molecules over time in natural reservoirs by sedimentation. We wondered if lipid-binding RNAs can be selectively enriched by this mechanism. To ensure reproducibility, liposomes were extruded to ~100 nm diameter after rehydration and characterized by dynamic light scattering (DLS) prior to the addition of oligos (Fig. 2.25). We, therefore, spun down the RNA-liposome suspensions and resuspended the particles with fresh buffer. Even after three washing cycles, flow cytometry data indicated a strong retention of geranylated **ON23-5FAM-geS** but not non-geranylated **ON23-5FAM-S**, or ges^{2}U-modifed oligos with weak lipid association (**ON22-5FAM-geS & ON24-5FAM-geS**), suggesting that liposome association can allow selective enrichment of oligos with specific ges^{2}U arrangement (Fig. 2.23 & 2.24).

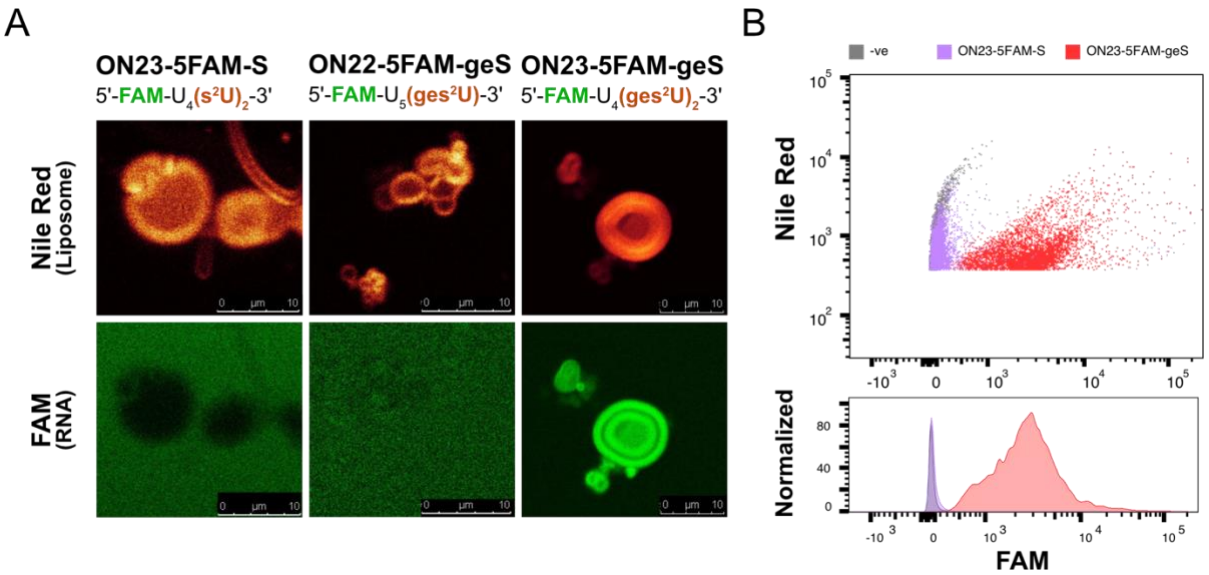


Fig. 2.23. Lipidation of RNAs by ges^{2}U. *(A) Confocal images of (ge)s^{2}U modified oligos showing the lipidating effect of ges^{2}U but not s^{2}U. (B) Flow cytometry data showing FAM signal from **ON23-5FAM-geS** but not **ON23-5FAM-S** after three wash cycles with liposomes. Negative control (-ve) contains only Nile Red™-stained liposomes and buffer. Adapted from ref.*[227]

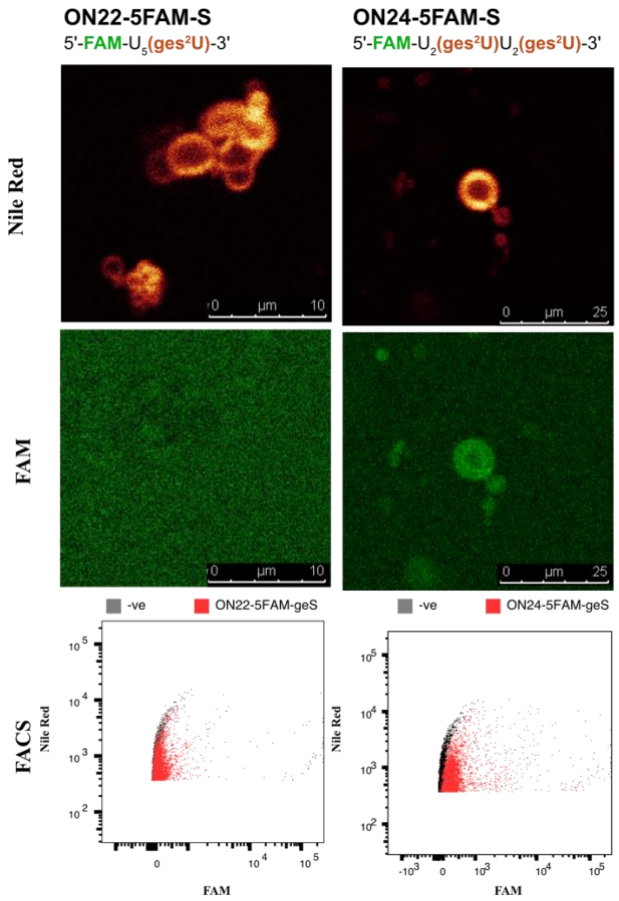


Fig. 2.24. Confocal images & FACS analysis of ges^{2}U oligos with weak lipidation. *(Top & middle) Confocal images showed weak association between Nile Red™-stained liposomes and FAM-labeled ges^{2}U-RNAs. (Bottom) Flow cytometry data showed that ges^{2}U-RNAs with weak lipid association cannot survive resuspension cycles. Adapted from ref.*[227]

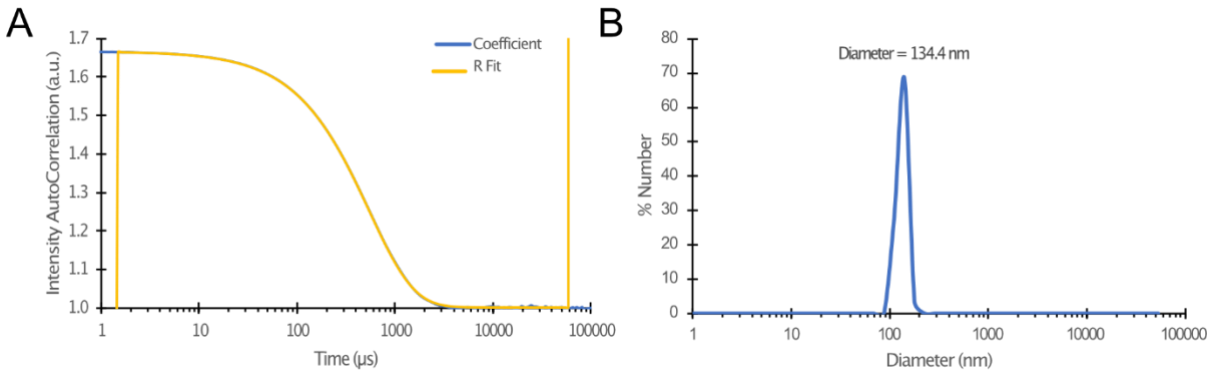


Fig. 2.25. Dynamic light scattering analysis of liposome suspension. *(A) Autocorrection of* coefficient *of scattering intensity over time (blue) and the selected regularisation fitting region (yellow). (B) The calculated size distribution (number %) of the liposome, with the peak at 134.4 nm diameter. Adapted from ref.*[227]

We next investigated if RNAs with longer canonical sequences can achieve this ges^{2}U-assisted lipidation. To this end, we synthesized oligos with two terminal ges^{2}U modifications with increasing length of canonical nucleotides and investigated their lipidation capabilities by the same confocal microscopy and flow cytometry protocols (Fig. 2.26). We first replaced the

four uridines in **ON23-5FAM-geS** to other canonical nucleotides (**ON25-5FAM-geS**) and observed that the liposome-binding is not U-specific. Extending the canonical:ges^{2}U nucleotide ratio, **ON26-5FAM-geS** (6:2) and **ON27-5FAM-geS** (9:2) also showed efficient association with lipids. However, when we furthered the extension to 12:2 in **ON28-5FAM-geS**, the localization effect weakened in its confocal image, suggesting a limitation in the number of hydrophilic canonical nucleotides per ges^{2}U nucleotide can lipidate.

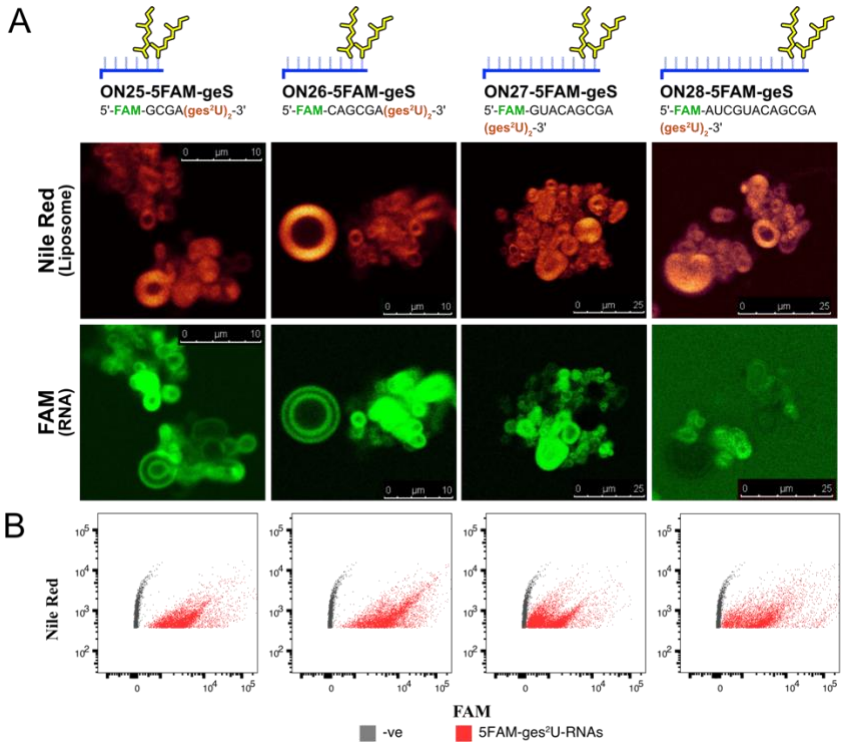


Fig. 2.26. Confocal images & FACS analysis of longer ges^{2}U oligos. *(A) Confocal images showed strong association between the described FAM-labeled ges^{2}U-RNAs except **ON28-5FAM-geS**. (B) Flow cytometry data of the respective FAM-labeled ges^{2}U-RNAs (red) and negative control (-ve) that contains only the stained liposomes and buffer. Adapted from ref.[227]*

Next, we asked if ges^{2}U modified oligos can capture unmodified counterstrands on the liposome surface by base pairing. Hence, we synthesized unlabeled ges^{2}U-modified oligos and hybridized them with their 3'-FAM labeled counterstrands and added the duplex to the liposomes (Fig. 2.27). Inspired by the above results, we started with unlabeled **ON26-geS** and base paired it with **ON29-3FAM**. We initially observed an exclusion of FAM signal from the liposomes in both confocal images and flow cytometry analysis, similar to the negative control with only **ON29-3FAM**. However, when we extended the ges^{2}U tail with one additional ges^{2}U (**ON30-geS**), we saw distinct accumulation of unmodified **ON29-3FAM** captured on liposomes. To confirm that the capturing is a result of duplex formation, the experiment was repeated with a 3'FAM labeled strand with mutated sequence (**ON31-3FAM**), in which we observed no lipid-binding (Fig. 2.28). Our results here showed that ges^{2}U nucleotides do not only equip RNAs with the ability to phase separate, but also decorate the liposome surface to achieve primitive molecular recognition based on base pairing.

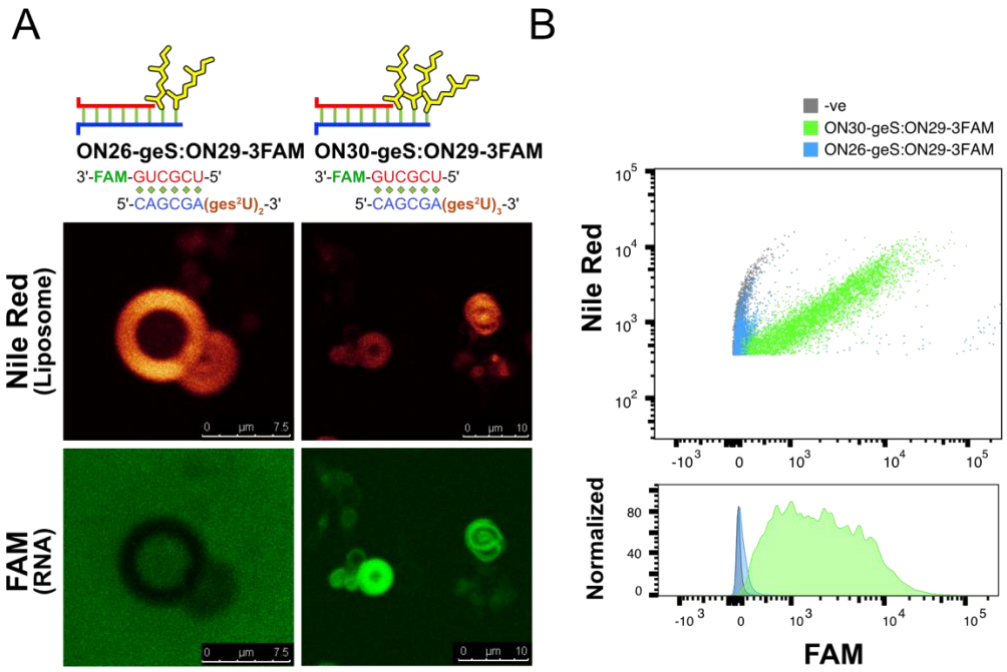


Fig. 2.27. Confocal images & FACS analysis of ges^{2}U RNA duplexes. *(A) Confocal images and (B) flow cytometry data showing the capability of ges^{2}U oligos to capture an unmodified counterstrand on liposome surface. Adapted from ref.*[227]

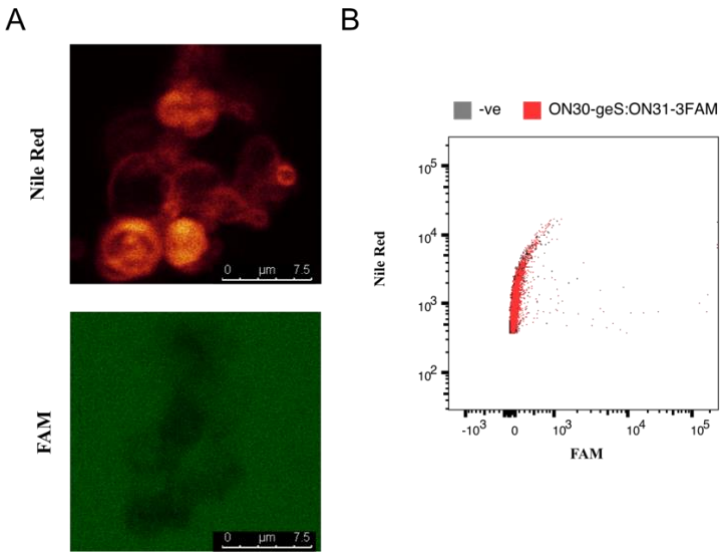


Fig. 2.28. Confocal images & FACS analysis of ges^{2}U RNA mismatch. *(A) Confocal images and (B) flow cytometry data showing the inability of a ges^{2}U oligo to capture an unmodified mismatch strand on liposome surface. Adapted from ref.*[227]

2.3.3 Liposome-catalyzed RNA geranylation

The above results inspired us to see if the lipid membrane can be utilized to concentrate hydrophobic substances in water and act as a reaction medium. Using geBr as a model compound of activated geraniols, we performed geranylation of single stranded RNA (ssRNA) with two s^{2}U units (**ON32-S**) in a solution of 20 mM geBr, 100 mM NaCl and 100 mM borate

buffer (pH 10) and compared the result in the absence (-Lip) and presence (+Lip) of liposome by HPLC and MALDI-TOF after 2 h. We observed that single and double geranylation were moderately and significantly enhanced, increasing the formation of **ON32-geS** from 3.1±0.4% to 10.5±1.8% (Fig. 2.29A). As geBr is immiscible with water, they are likely to be concentrated on the surface of the suspended liposomes with the ges^{2}U-modified RNA and therefore facilitated the geranylation.

We then investigated if lipid-anchored ges^{2}U-modified RNAs can recruit their s^{2}U counterstrand and specifically catalyze their geranylation in the presence of geBr and liposomes (I-III in Fig. 2.29). To the same geBr-liposome suspension, we added **ON32-S** and its complementary ges^{2}U RNA **ON30-geS**, forming an RNA duplex (dsRNA). Comparing to the geranylation of ssRNA with geBr-liposome, we observed a further increase in the fully geranylated **ON32-geS** formation by about 5-fold, from 10.5±1.8% to 48.2±2.7%. We repeated the experiment with a mutated ges^{2}U strand **ON33-geS**, and observed the yield dropped to 14.2±0.6%, similar to that of ssRNA **ON32-S** geranylation in liposome, confirming that the catalysis is based on duplex formation on the liposome surface. Our results suggested that pre-existed geranylated RNAs on the liposome surface can catalyze the geranylation by selectively recognizing their fully complementary strand.

We then ask if the process can be selectively bestowed on unpaired s^{2}U but not the base paired ones (IV-VII in Fig. 2.29). From the duplex **ON32-S**:**ON30-geS**, we removed the original s^{2}U overhang and replaced the terminal U:A base pair into an s^{2}U:A base pair, forming duplex **ON34-S**:**ON30-geS**. We reacted this duplex in the geBr-liposome suspension and observed a reduced geranylation from 30.0±1.2% to 16.0±1.0 to form **ON34-geS**, when compared with reacting the ssRNA **ON34-S** alone with geBr-liposome. To see if the system can promote site-specific geranylation when two s^{2}U sites are present on one molecule, we added after the terminal s^{2}U:A base pair an s^{2}U overhang, forming **ON35-S:ON30geS** and reacted it with the geBr-liposome. We observed an almost doubled formation of one of the single geranylated isomer with a longer retention time on the HPLC from 19.6±1.6% to 36.4±2.9%, and the suppression of the other isomer with a shorter retention time, when compared to the **ON35-S** ssRNA reaction with geBr-liposome. Collectively, and since terminal ges^{2}U has more conformational freedom and can better interact with the reverse phase of HPLC, we believe that the terminal s^{2}U was preferentially geranylated. Our system demonstrated how site-selective modifications could be facilitated in the presence of liposomes.

Finally, we studied if the proxy lipid phase can provide hydrolysis resistance to ges^{2}U on the anchored RNA (Fig. 2.30). We studied the stability of **ON12-geS** at pH 7 and pH 10, and compared it in +Lip (20 mM Egg PC) and -Lip and monitored its integrity over 4 days. While the liposome did not reduce the hydrolysis at pH 7, it significantly protected the ges^{2}U from hydrolysis at pH 10, increasing the integrity from 33.7±2.3% to 47.7±3.9% after 4 days. Our result here demonstrated that liposomes do not only mediate the formation of ges^{2}U but also provide protection from de-geranylation.

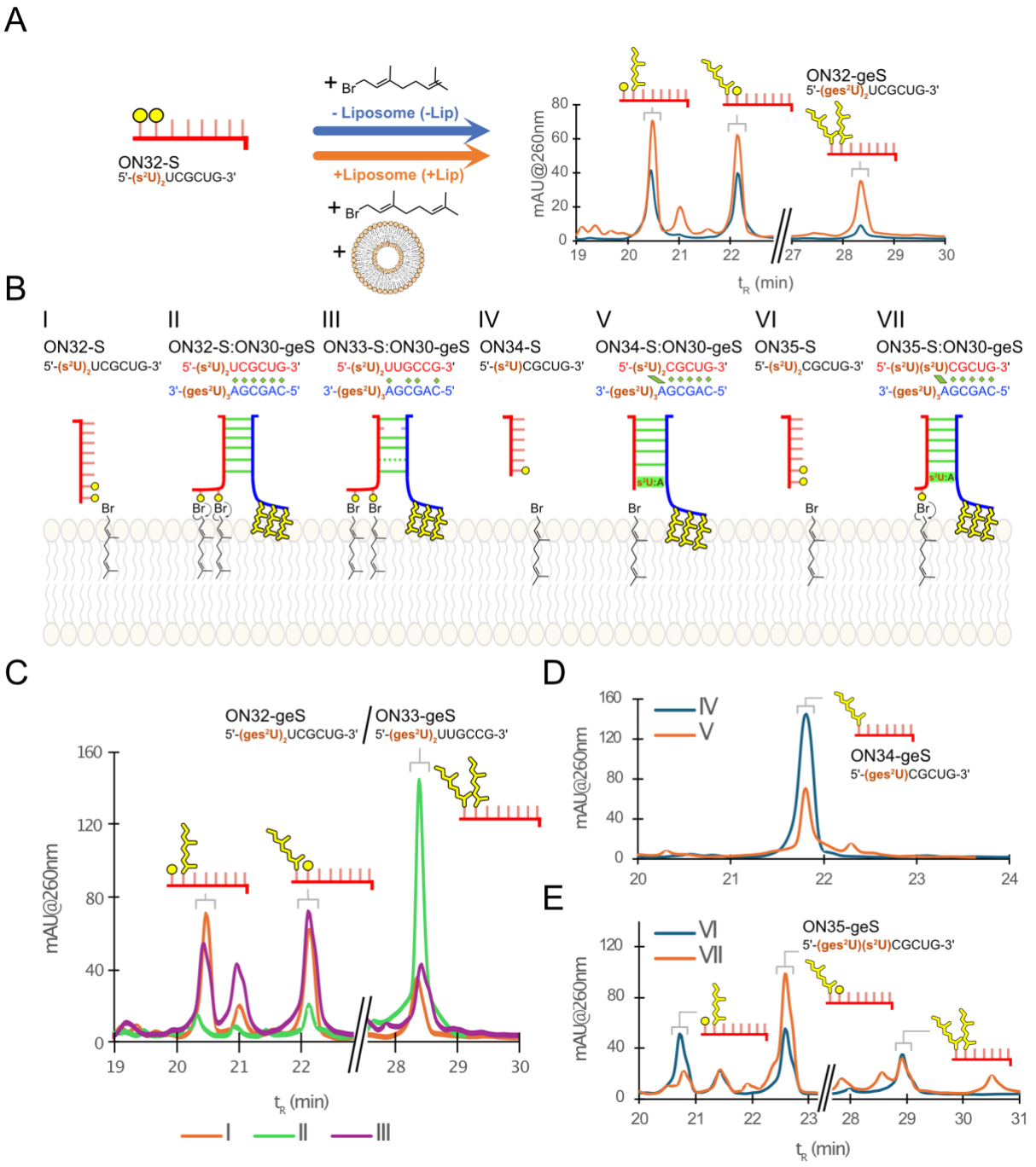


Fig. 2.29. Selective geranylation of s^{2}U on liposomal membrane. *(A) Liposome-promoted geranylation of **ON32-S**, showing an improved yield of **ON32-geS** formation. (B) Depiction of a series of experiments of sequence-specific (I-III) and site-specific (IV-VII) geranylation. (C) Overlaid HPL-chromatograms showing geranylation with liposomes takes place exclusively in a perfect duplex. (D) HPL-chromatograms showing the suppression of geranylation when the s^{2}U is protected by base pairing. (E) HPL-chromatograms showing the site-specific geranylation of unpaired s^{2}U but not base paired s^{2}U. Adapted from ref.[227]*

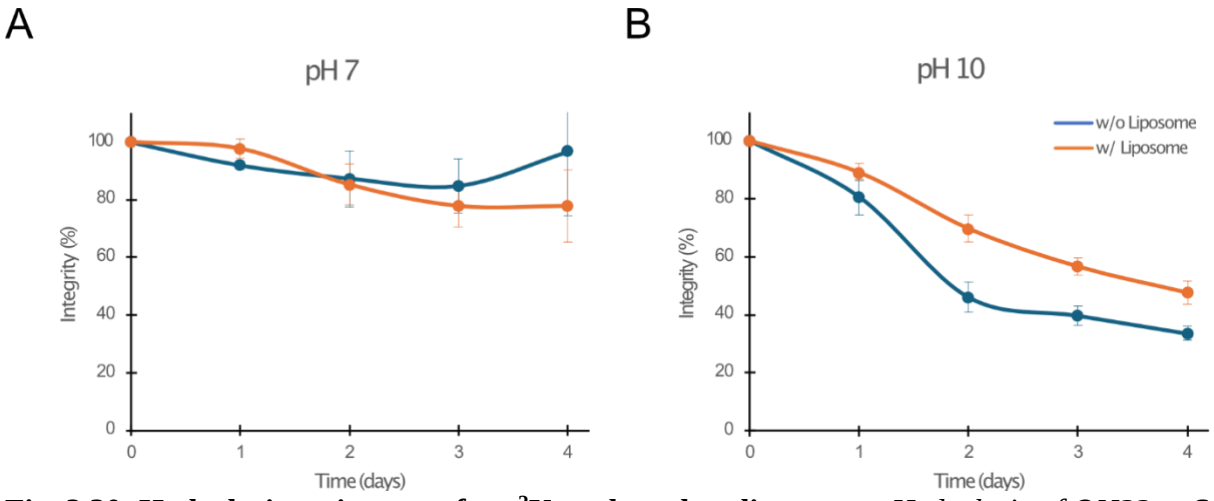
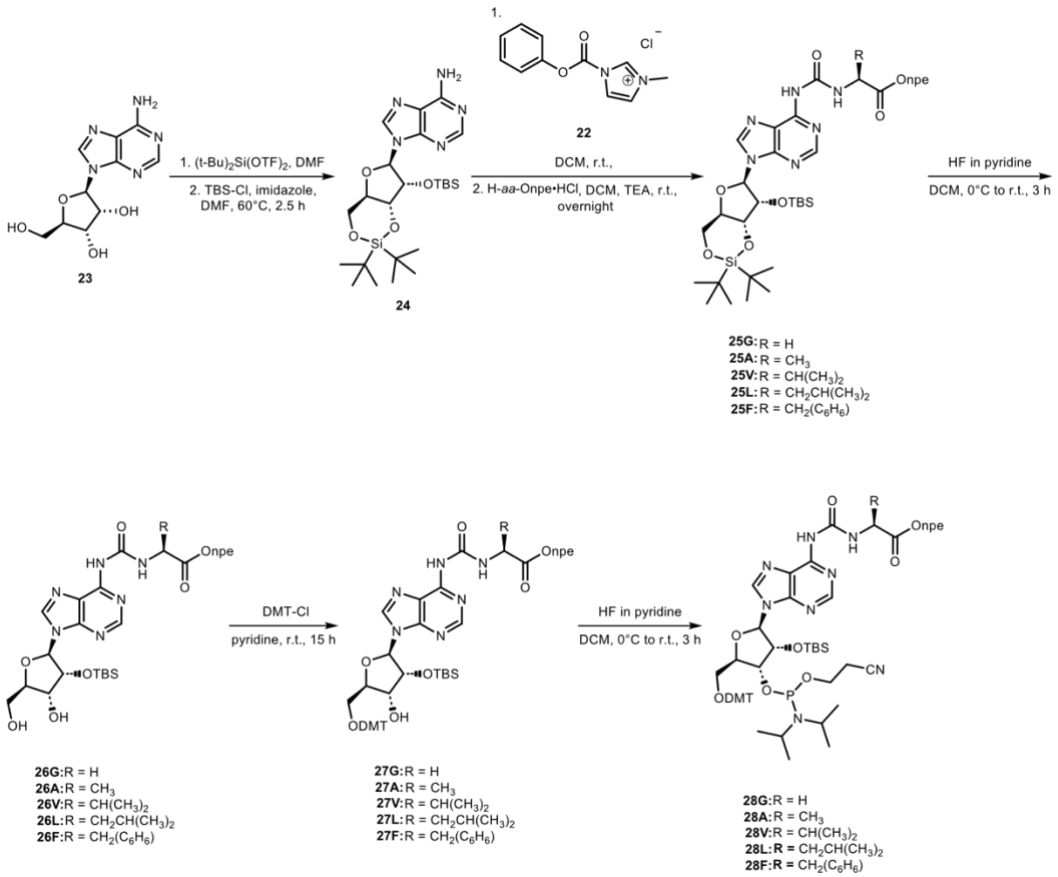


Fig. 2.30. Hydrolysis resistance of ges^{2}U anchored on liposomes. *Hydrolysis of **ON33-geS** over time in (A) phosphate buffer pH 7 and (B) glycine buffer pH 10 over 4 days in -Lip (blue) and +Lip (orange) setups. Adapted from ref.*[227]

2.3.4 Liposome effects on template-directed peptide couplings

We next studied how lipid-anchoring would affect the template-directed peptide couplings between the *aa*^{6}A and nm^{5}U nucleotides. We, therefore, synthesized a new library of *aa*^{6}A phosphoramidites for this purpose according to the described pathway (Scheme 2.9).



Scheme 2.9. Synthesis of *aa*^{6}A phosphoramidites. *General strategy for synthesis of aa^{6}A phosphoramidites. Adapted from ref.*[227]

Next, we synthesized RNA oligos with both ges^{2}U and either *aa*^{6}A or nm^{5}U incorporations. We first explored if the *aa*^{6}A and nm^{5}U are compatible with the post-synthetic geranylation of s^{2}U nucleotides (Fig. 2.31). We found out that the *aa*^{6}A did not react with the geBr during the geranylation step, whereas the primary amine of nm^{5}U needs to remain protected during the geranylation. After screening of reaction conditions, we discovered a mild teoc deprotection condition after the geranylation with minimal degradation of ges^{2}U and the oligo, by TBAF at 15 ˚C for 1.5 h. Using a 5-bp duplex system, we initially synthesized **ON36-g^{6}A-geS**, **ON37-nm^{5}U**, **ON38-g^{6}A**, and **ON39-nm^{5}U-geS** to study the effects.

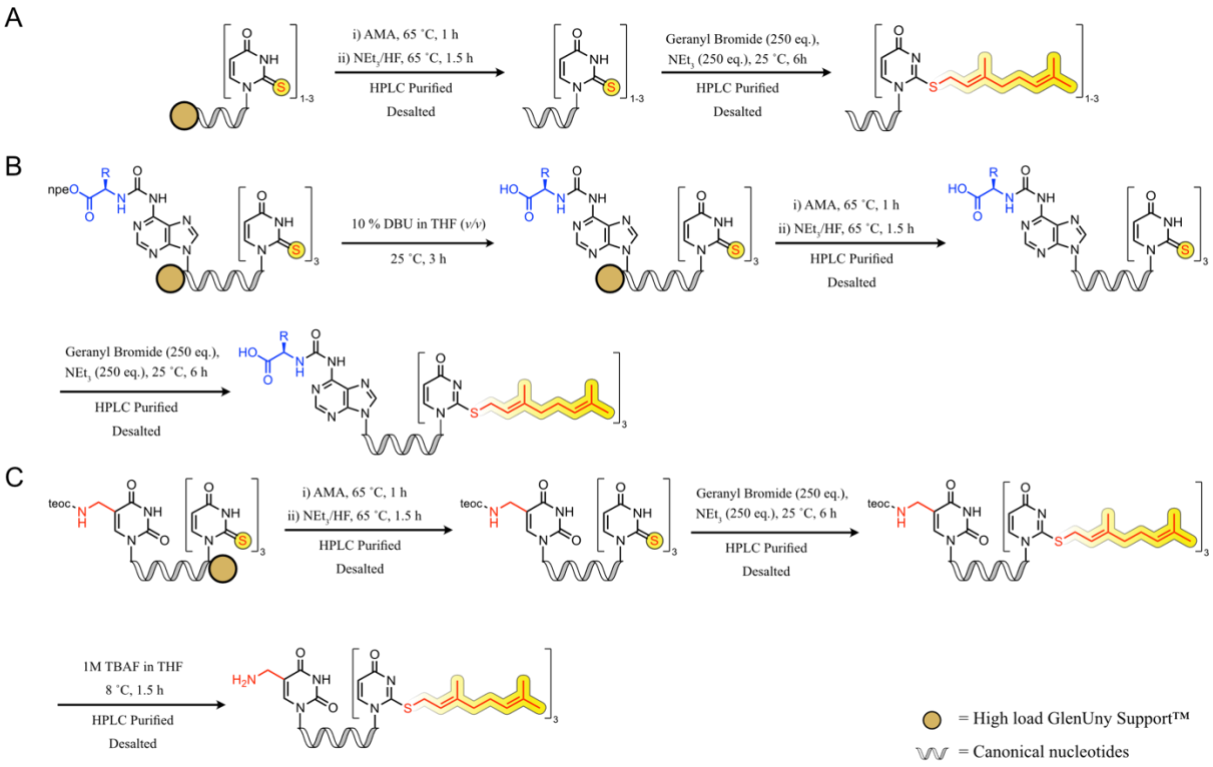


Fig. 2.31. Synthesis of RNA oligos with multiple modifications. *Synthesis of: (A) ges^{2}U-modified RNAs with post-synthetic geranylation; (B) aa^{6}A- and ges^{2}U-modified RNAs with npe-deprotection prior to post-synthetic geranylation. (C) nm^{5}U and ges^{2}U-modified RNAs under mild teoc-deprotection conditions after post-synthetic geranylation. Adapted from ref.*[227]

To ensure duplex formation at r.t., a UV-melting study was performed on **ON36-g^{6}A-S:ON37-nm^{5}U** with the same buffer and RNA concentrations as the reaction conditions, obtaining a melting temperature of 34 ˚C (Fig. 2.32). We then determined the yields of the RNA-peptide hairpin formation based on HPLC calibration curve sobtained from canonical oligonucleotides that substitute the modifications with their closest canonical analogs (**CON7-10**, Fig. 2.33). To obtain an overview of these similar reactions performed in the presence (+Lip) and absence (-Lip) of liposomes, relative yields were calculated by dividing the peak area of the formed RNA-peptide products in -Lip by that of it in +Lip (Fig. 2.34B and Table 2.4 for all coupling yields).

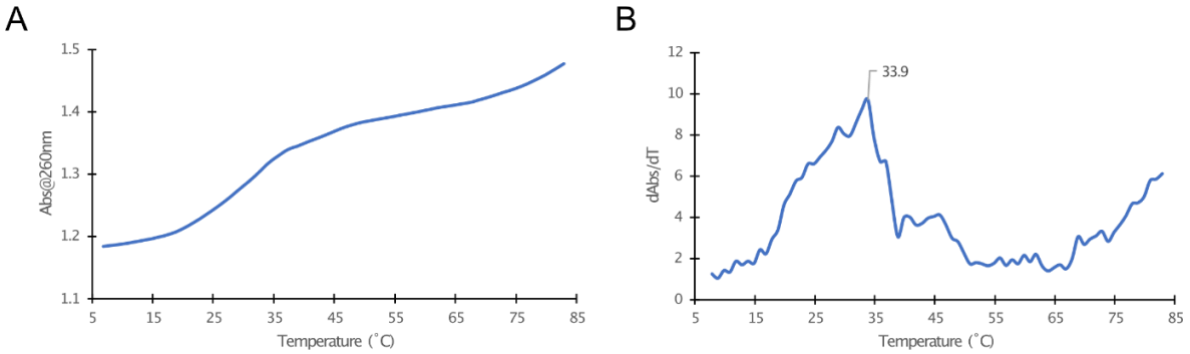


Fig. 2.32. UV-melting of ON36-g^{6}A-geS:ON37-nm^{5}U. *(A) Raw data and (B) derivative plots of UV-melting experiments of the 5 bp modified duplex. Adapted from ref.*[227]

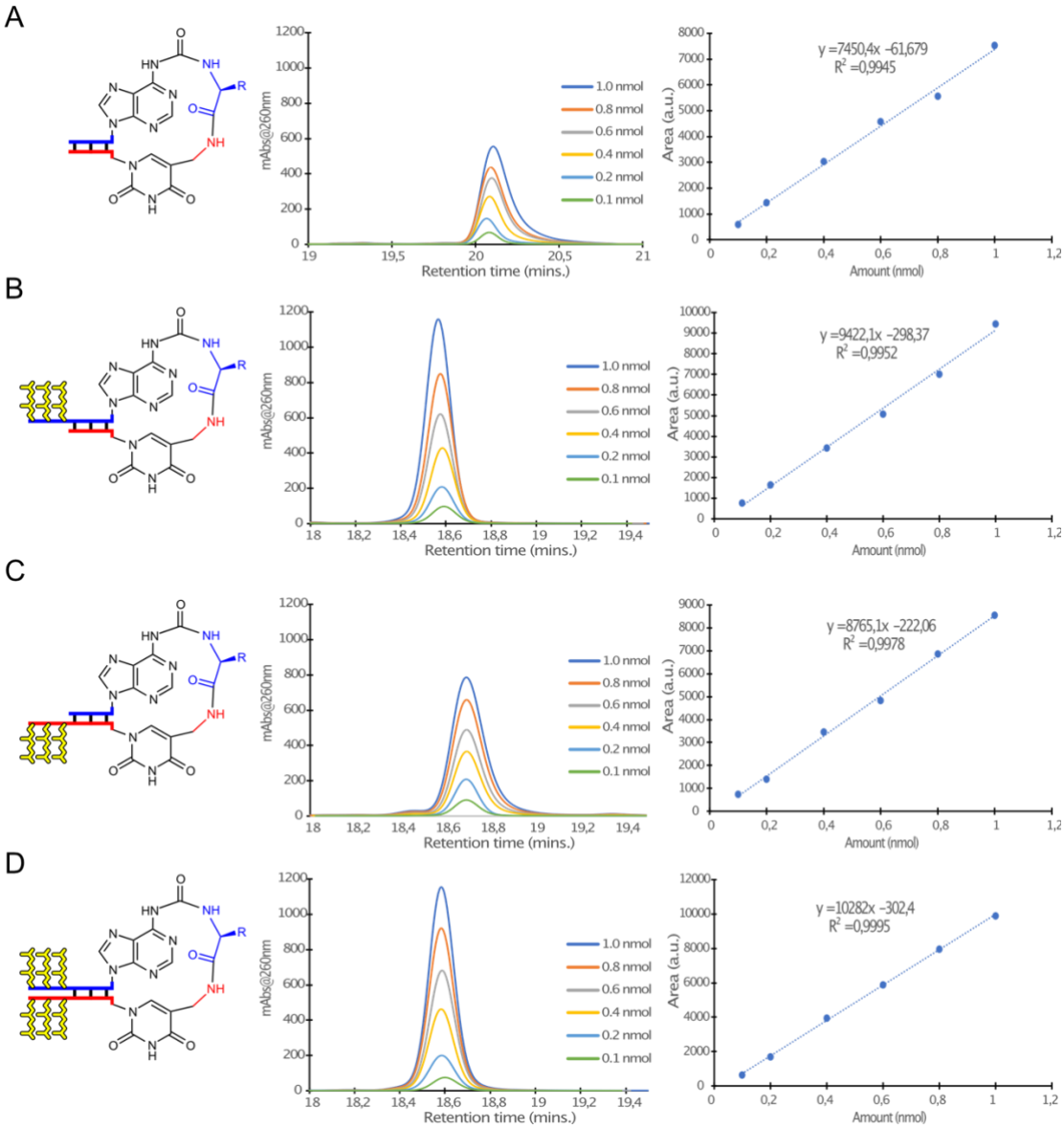


Fig. 2.33. Calibration curves for determination of coupling yields. *Calibration curves bases on canonical RNA hairpins with sequences equivalent to (A)* ***ON37-nm^{5}U:ON38-g^{6}A****, (B)* ***ON36-g^{6}A-geS:ON37-nm^{5}U****, (C)* ***ON38-g^{6}A:ON39-nm^{5}U-ges^{2}U*** *and (D)* ***ON36-g^{6}A-geS:ON39-nm^{5}U-geS****. Adapted from ref.*[227]

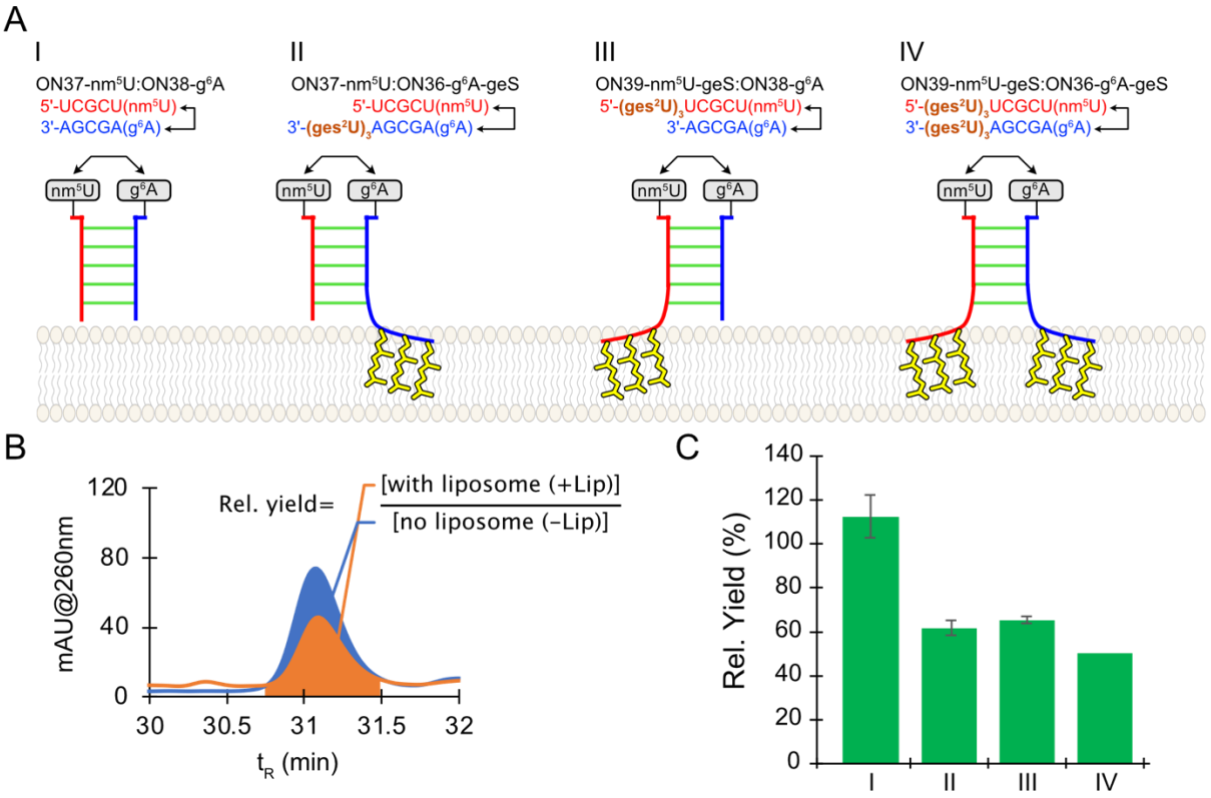


Fig. 2.34. Peptide couplings of duplexes with ges^{2}U tail on different strands. *(A) Depiction of 5 coupling reactions to investigate the lipid-anchoring (B) Relative yields of each coupling reaction were calculated by dividing the peak area of formed RNA-peptide from +Lip by that of -Lip. (C) Relative yields calculated for coupling reactions I-VI in (A). Adapted from ref.*[227]

To begin with, we checked if the presence of liposome has any effect on non-geranylated RNA duplex (Fig. 2.34A & C -I). We reacted **ON37-nm^{5}U:ON38-g^{6}A** in -Lip and +Lip and detected very similar yields (relative yield: 113±18±%). Next, we investigated systems of dsRNAs with the three ges^{2}U tail on either side of the duplex. In the reactions of duplexes **ON37-nm^{5}U:ON36-g^{6}A-geS** (Fig. 2.34A & C -II) and **ON39-nm^{5}U-ges^{2}U:ON38-g^{6}A** (Fig. 2.34A & C -III), we found that their peptide couplings are generally suppressed in +Lip, with resulting relative yields of 62±10% and 65±3%, respectively. Putative reasons for this effect are: the restricted conformation of the *aa*^{6}A/nm^{5}U nucleotides and limited accessibility of the hydrophilic activator DMTMM•Cl when the duplex is anchored at the water-lipid interphase, or the reaction takes place mainly in the bulk solution where the concentration of oligomers is very small, decreasing the coupling rate. When both sides of the duplex are geranylated in the reactions of **ON39-nm^{5}U-geS:ON36-g^{6}A-geS**, the relative yield was further decreased to 50±2% (Fig. 2.34A & C -IV).

We then investigated the potential chemoselectivity of peptide coupling caused by the proxy lipid phase with derivatives of *aa*^{6}A charged with different amino acids (Fig. 2.35). To obtain a more unbiased result, we employed **ON39-nm^{5}U-ges^{2}U** as the universal nm^{5}U oligo and reacted it with **ON38**s charged with different *aa*^{6}A nucleotides (Fig. 2.35A). Apart from the single amino acid-modifed *aa*^{6}A, we also synthesized *peptide*^{6}A charged with repeating amino acid sequences with the mentioned on-bead coupling method, forming reaction series of

ON39-nm^{5}U-ges^{2}U:ON38-[n]*aa*^{6}A with n = length of the repeating peptide (1-4) and *aa* = glycine (g), alanine (a), valine (v), leucine (l) or phenylalanine (f). We compared these reactions in -Lip and +Lip setups and calculated their relative yields. The results were shown in Fig. 2.35B. Starting from reactions of **ON39-nm^{5}U-ges^{2}U:ON38-1*aa*^{6}A**, we observe minor difference of relative yields of all amino acids with the order of Phe ≈ Gly > Leu ≈ Val > Ala, ranging from Phe 68±15 to Ala 45±1%. However, we observed greater relative yield deviation as the peptide length increases, resulted in a greater difference between peptides formed by more and less lipophilic amino acids. In the coupling reactions of tetrapeptidal RNAs (**ON38-4*aa*^{6}A**:**ON39-nm^{5}U-ges^{2}U**), we observed an alignment of relative yields in accordance with the lipophilicity of the peptide, resulting in the order of 4Gly > 4Ala > 4Val ≈ 4Leu > 4Phe, relative yields from Gly 54±3% to Phe 0% (undetectable). It is noteworthy that we did not observe any trend if we only look at the absolute coupling yields (Table 2.4), except that di- and tri-peptidyl RNAs usually coupled better as the width of a base pair is similar to the length of 2-3 peptide bonds. We encountered solubility issue when we tried to synthesized 5Phe and 5Leu RNAs and therefore, stopped at tetrapeptidyl RNAs.

To demonstrate selectivity in a direct competition experiment, we set up coupling experiments with a mixture of 1:1:1 **ON38-4a^{6}A:ON38-4l^{6}A:ON39-nm^{5}U-geS** (Fig. 2.35C). Here, we observed that the ratio of the formed RNA-peptide hairpins from 4l^{6}A to 4a^{6}A was initially 1:1.3 in -Lip setup, which was then increased to 1:1.9 in the +Lip case, showing that the liposomes can amplify subtle differences in the peptide coupling yields. In summary, our model suggested a potential scenario where short peptides and RNAs could be co-selected by a lipid phase, by anchoring the peptide coupling on liposome surface with an RNA modification.

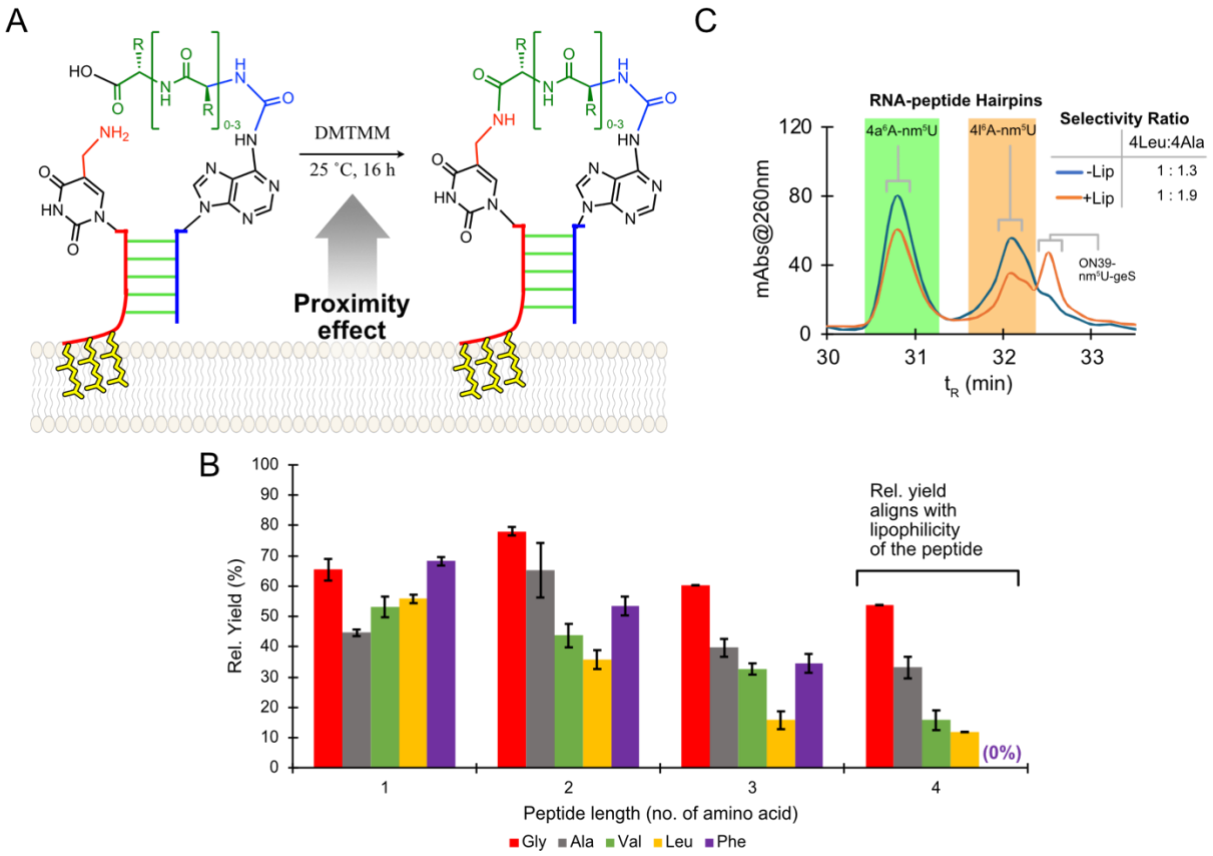


Fig. 2.35. Template-directed peptide couplings on liposome surface. *(A) Schematic of lipid surface template directed coupling by incubating 10 µM dsRNA with the corresponding modifications in 100 mM NaCl, 100 mM MES buffer pH 6.0, and 50 µM DMTMM•Cl in the absence (-Lip) or presence (+Lip) of liposomes. (B) Relative yields calculated from coupling reactions with aa^{6}A charged with different amino acids or their repeating oligopeptides and reacted with* ***ON39-nm^{5}U-geS****. The coupling reactions of tetrapeptidyl RNA features decreasing relative yields with increasing lipophilicity of the reacting peptides. (C) Competition experiment between* ***ON17-4a^{6}A*** *and* ***-4l^{6}A*** *with* ***ON39-nm^{5}U-geS****. The ratios of the formed RNA-peptides hairpins of 4a^{6}A to 4l^{6}A are 1:1.3 and 1:1.9 in -Lip and +Lip respectively. Adapted from ref.*[227]

Name	Sequence information			Polymer	HPLC		MALDI-TOF	
	5' Mod.	Sequence (5'-3')	3' Mod.		0-40% B t_R (min)	0-80% B t_R (min)	*m/z* calcd. for $[M-H]^-$	found
ON22-geS	-	UUUUU(ges^{2}U)	-	RNA	40.4	-	1925.3	1924.7
ON22-5FAM-S	6FAM	UUUUU(s^{2}U)	-	RNA	26.1	13.7	2326.3	2325.5
ON22-5FAM-geS	6FAM	UUUUU(ges^{2}U)	-	RNA	-	22.8	2462.4	2463.7
ON23-S	-	UUUU(s^{2}U)(s^{2}U)	-	RNA	18.4	11.6	1805.1	1805.1
ON23-geS	-	UUUU(ges^{2}U)(ges^{2}U)	-	RNA	-	29.3	2077.4	2076.7
ON23-5FAM-S	6FAM	UUUU(s^{2}U)(s^{2}U)	-	RNA	23.1	13.5	2342.3	2344.2
ON23-5FAM-geS	6FAM	UUUU(ges^{2}U)(ges^{2}U)	-	RNA	-	30.7	2614.5	2616.1
ON24-5FAM-S	6FAM	UU(s^{2}U)UU(s^{2}U)	-	RNA	24.0	14.2	2342.3	2343.2
ON24-5FAM-geS	6FAM	UU(ges^{2}U)UU(ges^{2}U)	-	RNA	-	27.7	2614.5	2616.0
ON25-5FAM-S	6FAM	GCGA(s^{2}U)(s^{2}U)	-	RNA	24.3	14.2	2442.3	2443.1
ON25-5FAM-geS	6FAM	GCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	30.5	2714.6	2715.3
ON26-S	-	CAGCGA(s^{2}U)(s^{2}U)	-	RNA	18.1	11.1	2539.3	2540.4
ON26-geS	-	CAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	29.3	2811.6	2812.4
ON26-5FAM-S	6FAM	CAGCGA(s^{2}U)(s^{2}U)	-	RNA	25.7	15.2	3076.4	3078.9
ON26-5FAM-geS	6FAM	CAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	30.8	3348.7	3350.8
ON27-5FAM-S	6FAM	GUACAGCGA(s^{2}U)(s^{2}U)	-	RNA	23.9	13.8	4056.6	4059.0
ON27-5FAM-geS	6FAM	GUACAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	29.3	4328.8	4331.5
ON28-5FAM-S	6FAM	AUCGUACAGCGA(s^{2}U)(s^{2}U)	-	RNA	25.2	14.6	4996.7	4998.4
ON28-5FAM-geS	6FAM	AUCGUACAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	29.3	5268.9	5275.5
ON29-3FAM	-	UCGCUG	6FAM	RNA	Ordered			
ON30-S	-	CAGCGA(s^{2}U)(s^{2}U)(s^{2}U)	-	RNA	20.6	11.9	2861.3	2862.3
ON30-geS	-	CAGCGA(ges^{2}U)(ges^{2}U)(ges^{2}U)	-	RNA	-	35.1	3269.7	3271.4
ON31-3FAM	-	ACGGU	6FAM	RNA	Ordered			

ON32-S	-	(s^2U)(s^2U)UCGCUG	-	RNA	19.4	11.7	2493.3	2493.6
ON32-geS	-	(ges^{2}U)(ges^{2}U)UCGCUG	-	RNA	-	28.9	2765.5	
ON33-S	-	(s^2U)(s^2U)UUGCCG	-	RNA	19.2	11.8	2493.3	2493.7
ON33-geS	-	(ges^{2}U)(ges^{2}U)UUGCCG	-	RNA	-	28.5	2765.5	
ON34-S	-	(s^2U)CGCUG	-	RNA	17.6	10.9	1865.2	1865.5
ON34-geS	-	(ges^{2}U)CGCUG	-	RNA	-	21.9	2001.4	
ON35-S	-	(s^2U)(s^2U)CGCUG	-	RNA	19.3	11.8	2187.2	2187.6
ON35-geS	-	(ges^{2}U)(s^2U)CGCUG	-	RNA	-	22.5	2459.5	
ON36-g^{6}A-S	-	(g^{6}A)AGCGA(s^2U)(s^2U)(s^2U)	-	RNA	20.0	12.1	2986.3	2987.6
ON36-g^{6}A-geS	-	(g^{6}A)AGCGA(ges^{2}U)(ges^{2}U)(ges^{2}U)	-	RNA	-	34.5	3394.7	3395.1
ON37-nm^{5}U	-	UCGCU(nm^{5}U)	-	RNA	13.9	9.2	1839.3	1839.3
ON38-g^{6}A	-	(g^{6}A)AGCGA	-	RNA	18.1	11.2	2020.3	2020.1
ON38-2g^{6}A	-	(2g^{6}A)AGCGA	-	RNA	18.6	11.5	2077.4	2077.4
ON38-3g^{6}A	-	(3g^{6}A)AGCGA	-	RNA	18.7	11.5	2134.4	2134.2
ON38-4g^{6}A	-	(4g^{6}A)AGCGA	-	RNA	18.6	11.4	2191.4	2191.8
ON38-t^{6}A	-	(t^{6}A)AGCGA	-	RNA	18.6	11.4	2064.4	2064.4
ON38-2t^{6}A	-	(2t^{6}A)AGCGA	-	RNA	19.2	11.7	2165.4	2165.4
ON38-3t^{6}A	-	(3t^{6}A)AGCGA	-	RNA	19.3	11.7	2266.5	2266.7
ON38-4t^{6}A	-	(4t^{6}A)AGCGA	-	RNA	19.5	11.8	2367.5	2367.7
ON38-a^{6}A	-	(a^{6}A)AGCGA	-	RNA	19.1	11.6	2034.4	2034.6
ON38-2a^{6}A	-	(2a^{6}A)AGCGA	-	RNA	18.8	11.7	2105.4	2105.5
ON38-3a^{6}A	-	(3a^{6}A)AGCGA	-	RNA	19.0	11.8	2176.4	2176.6
ON38-4a^{6}A	-	(4a^{6}A)AGCGA	-	RNA	19.2	12.0	2247.5	2247.5
ON38-v^{6}A	-	(v^{6}A)AGCGA	-	RNA	22.1	13.2	2062.6	2062.4
ON38-2v^{6}A	-	(2v^{6}A)AGCGA	-	RNA	24.2	14.2	2161.4	2161.5
ON38-3v^{6}A	-	(3v^{6}A)AGCGA	-	RNA	25.7	15.0	2260.5	2260.5
ON38-4v^{6}A	-	(4v^{6}A)AGCGA	-	RNA	27.9	16.2	2359.6	2359.6

ON38-l^{6}A	-	(l^{6}A)AGCGA	-	RNA	24.1	14.3	2076.4	2077.6
ON38-2l^{6}A	-	(2l^{6}A)AGCGA	-	RNA	28.5	16.6	2189.5	2189.4
ON38-3l^{6}A	-	(3l^{6}A)AGCGA	-	RNA	32.6	18.8	2302.6	2302.5
ON138-4l^{6}A	-	(4l^{6}A)AGCGA	-	RNA	37.3	21.3	2415.7	2415.6
ON38-f^{6}A	-	(f^{6}A)AGCGA	-	RNA	24.5	14.5	2110.4	2111.7
ON38-2f^{6}A	-	(2f^{6}A)AGCGA	-	RNA	31.0	17.8	2257.5	2257.8
ON38-3f^{6}A	-	(3f^{6}A)AGCGA	-	RNA	36.3	20.7	2404.5	2405.2
ON38-4f^{6}A	-	(4f^{6}A)AGCGA	-	RNA	42.3	23.9	2551.6	2551.8
ON39-teocnm^{5}U-S	-	(s^{2}U)(s^{2}U)(s^{2}U)UCGCU(teocnm^{5}U)	-	RNA	32.9	19.2	2949.3	2950.4
ON39-teocnm^{5}U-geS	-	(ges^{2}U)(ges^{2}U)(ges^{2}U)UCGCU(teocnm^{5}U)	-	RNA	-	34.3	3357.7	3359.5
ON39-nm^{5}U-geS	-	(ges^{2}U)(ges^{2}U)(ges^{2}U)UCGCU(nm^{5}U)	-	RNA	-	32.4	3213.7	3214.3
CON7	-	UCGCUUAAGCGA	-	RNA	Ordered			
CON8	-	UCGCUUAAGCGAUUU	-	RNA				
CON9	-	UUUUCGCUUAAGCGA	-	RNA				
CON10	-	UUUUCGCUUAAGCGAUUU	-	RNA				

Table 2.3. Oligonucleotides described in section 2.3. *Left to right: oligo's nomenclature, sequence information, polymer type and its characterization data including HPLC retention time (gradient 0-40% & 0-80% buffer B) and MADI-TOF mass spectrometry data. Modifications aa^{6}A are named after the single letter IUPAC code for the corresponding amino acid. Information for all oligos in this thesis is summarized in the appendix.*

Synthesised from		HPLC	MALDI-TOF		Peptide coupling yield (%)		
aa^{6}A strand	nm^{5}U strand	0-80%B t_R (min)	*m/z* cald. for $[M-H]^-$	*m/z* found	-Lip	+Lip	Rel. Yield
ON38-g^{6}A	ON37-nm^{5}U	12.4	3841.6	3842.7	28.1 ± 2.0	30.9 ± 2.6	112.6 ± 18.0
ON36-g^{6}A-geS	ON37-nm^{5}U	33.8	5217.0	5224.4	30.5 ± 3.1	19.5 ± 2.6	61.7 ± 9.6
ON38-g^{6}A	ON39-nm^{5}U-geS	38.9	5217.0	5222.4	24.2 ± 0.3	16.9 ± 0.7	65.4 ± 3.4
ON36-g^{6}A-geS	ON39-nm^{5}U-geS	31.1	6591.4	6601.3	23.8 ± 1.9	13.4 ± 1.0	50.3 ± 1.7
ON38-2g^{6}A	ON39-nm^{5}U-geS	31.3	5274.0	5273.2	31.7 ± 1.5	25.4 ± 1.1	78.0 ± 4.5
ON38-3g^{6}A	ON39-nm^{5}U-geS	31.3	5331.0	5333.6	33.6 ± 1.4	21.5 ± 0.6	60.4 ± 2.8
ON38-4g^{6}A	ON39-nm^{5}U-geS	31.3	5388.1	5379.6	28.3 ± 0.7	16.7 ± 0.5	53.8 ± 2.6
ON38-a^{6}A	ON39-nm^{5}U-geS	31.4	5231.0	5229.3	25.8 ± 0.8	13.3 ± 0.2	44.7 ± 1.1
ON38-2a^{6}A	ON39-nm^{5}U-geS	31.3	5302.0	5307.2	36.2 ± 1.6	24.5 ± 1.9	65.2 ± 9.1
ON38-3a^{6}A	ON39-nm^{5}U-geS	31.3	5373.1	5378.2	37.9 ± 3.1	16.9 ± 1.0	39.7 ± 2.9
ON38-4a^{6}A	ON39-nm^{5}U-geS	31.3	5444.1	5449.6	34.2 ± 3.6	13.3 ± 1.1	33.2 ± 3.6
ON38-v^{6}A	ON39-nm^{5}U-geS	30.8	5259.0	5268.4	24.1 ± 0.5	14.2 ± 0.5	53.1 ± 3.5
ON38-2v^{6}A	ON39-nm^{5}U-geS	30.9	5358.1	5364.1	42.2 ± 1.6	20.3 ± 2.1	43.8 ± 3.9
ON38-3v^{6}A	ON39-nm^{5}U-geS	31.2	5457.2	5464.7	44.1 ± 1.1	15.7 ± 0.4	32.6 ± 1.9
ON38-4v^{6}A	ON39-nm^{5}U-geS	31.2	5556.2	5565.8	19.7 ± 2.3	5.6 ± 0.1	15.8 ± 3.3
ON38-l^{6}A	ON39-nm^{5}U-geS	30.8	5273.1	5278.7	22.4 ± 0.3	13.9 ± 0.2	55.9 ± 0.5
ON38-2l^{6}A	ON39-nm^{5}U-geS	31.3	5386.1	5393.9	40.2 ± 1.3	16.3 ± 1.6	35.8 ± 5.8
ON38-3l^{6}A	ON39-nm^{5}U-geS	31.7	5499.2	5508.6	26.1 ± 0.5	6.8 ± 0.5	15.8 ± 2.0
ON38-4l^{6}A	ON39-nm^{5}U-geS	32.2	5612.3	5616.4	21.3 ± 0.3	5.3 ± 0.4	11.9 ± 1.9
ON38-f^{6}A	ON39-nm^{5}U-geS	30.9	5307.0	5313.1	29.0 ± 0.4	20.8 ± 0.4	68.3 ± 1.4
ON38-2f^{6}A	ON39-nm^{5}U-geS	31.2	5454.1	5464.6	30.5 ± 1.5	17.7 ± 0.2	53.5 ± 3.1
ON38-3f^{6}A	ON39-nm^{5}U-geS	31.7	5601.2	5588.5	30.6 ± 1.5	12.6 ± 0.6	34.5 ± 3.0
ON38-4f^{6}A	ON39-nm^{5}U-geS	32.2	5748.2	5739.7	28.9 ± 2.5	n.d.	n.d.

Table 2.4. Yields of peptide coupling reactions. *(Left to right) Sequence information of the reacted aa^{6}A and nm^{5}U RNAs, and HPLC, MALDI-TOF and calculated yields of the formed RNA-peptide hairpins.*

2.4 Polymerization of modified RNAs by a triplet polymerase ribozyme

As we learnt that *aa*^{6}A destabilizes duplex formation, we explored the possibility of potential replication of modified RNAs via fragment ligation using a ribozyme, as fragments have higher melting temperatures than nucleoside triphosphate monomers and are more capable to tolerate unpaired nucleotides. Furthermore, we asked if the replacement of uridines to s^{2}U in the templates and the substrates could potentially improve the fidelity of the ribozyme's template-copying, as s^{2}U prevents wobble base pair with guanosine. The data in this section are unpublished and are resulted from a collaboration with the Holliger's group in the MRC-LMB, Cambridge. (Information on all oligos described in 2.4 is listed in p. 71 and section 4.2)

2.4.1 Synthesis of modified triplet triphosphates

While canonical triplet triphosphates (NNNTPs) can be prepared by *in vitro* transcription, the production of NNNTPs with non-canonical nucleotides can only resort to chemical synthesis, as non-base pairing modifications, like (m^{6})*aa*^{6}A, stall transcriptases and base pairing modifications, like s^{2}U, suffer from very low yields (unpublished results from the Holliger's lab). Multiple synthetic methods were described in the literature, however, the most efficient is the one described by the Meier group in 2015.[234,235] A *cyclo*Sal phosphoramidite (**41**) was employed on the solid phase RNA/DNA synthesis to activate the 5'OH group of the terminal nucleotide and immediately reacted with pyrophosphate (added as tetrabutylammonium solution) on the synthesizer (Fig. 2.36).

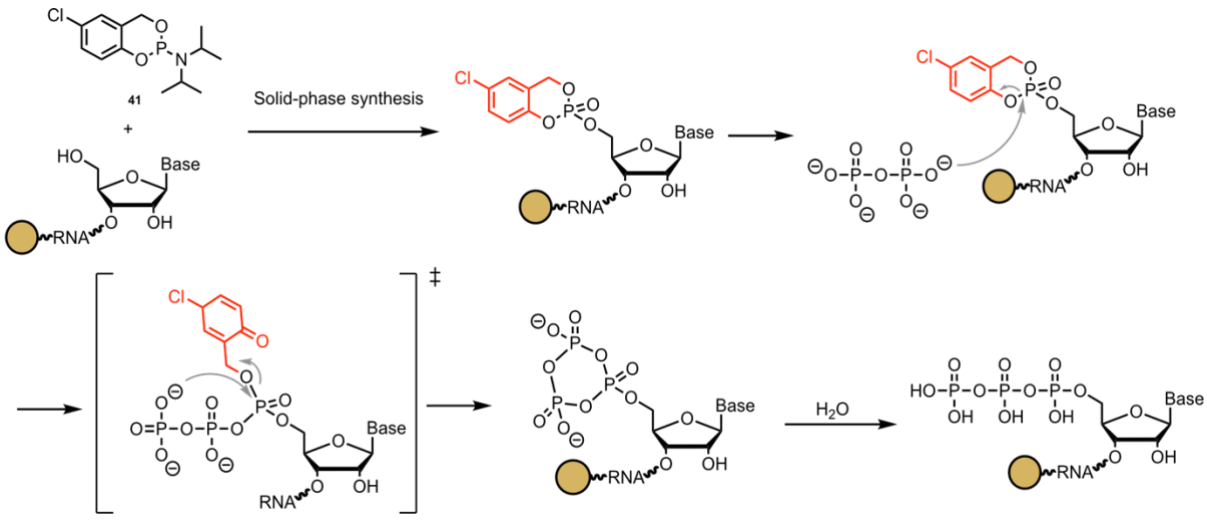
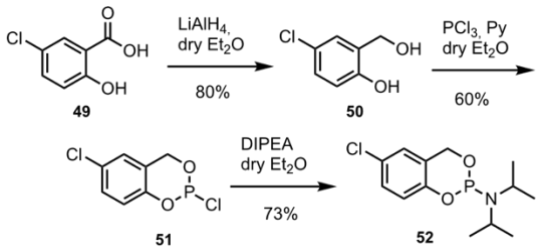


Fig. 2.36. Solid phase synthesis of oligonucleotide 5' triphosphates. *The 5'OH group was first coupled with the cycloSal phosphoramidite (**41**) and further reacted with tetrabutylammonium pyrophosphate to form the final 5'triphosphate oligonucleotides, which then proceeded with standard cleavage and deprotection processes.*



Scheme. 2.10. Synthesis of *cyclo*Sal-phosphoramidite. *The synthesis route was reported previously by Meier.*[234]

The synthesis of the *cyclo*Sal phosphoramidite (Scheme 2.10) started with the reduction of 5-chlorosalicyclic acid (**49**-**50**) and phosphitylation with PCl_3, forming highly unstable **51**. It can then readily react with DIPEA to form the desired *cyclo*Sal phosphoramidite (**52**), which has high solubility in acetonitrile and can be coupled as normal phosphoramidite on the DNA/RNA synthesizer with a modified coupling cycle lacking the capping steps. The phosphoramidite **52** can be stored in the freezer for up to 9 months in an Ar-filled vial and is extremely sensitive to moisture. A series of 6 modified NNNTPs were synthesized by solid phase synthesis, namely pppC(s^2U)A, ppp(s^2U)GC, pppC(m^6A)G, ppp(m^6A)GG, ppp(m^6g^6A)GG and pppC(mnm^5U)C (Fig. 2.37 & Table 2.5). Theoretically, phosphoramidites of non-canonical nucleosides can be incorporated in the 3' end if a universal solid support (e.g. Glen UnySupport™ and Universal Suport III, Glen research) was employed. However, it was discovered that these linkers experienced side reactions with the phosphitylation chemistry and therefore, the synthesis of NNNTPs was restricted to commercially available CPG beads with a 3' pre-coupled canonical nucleotide. The npe and teoc deprotection reactions for m^6g^6A and mnm^5U, respectively, proceeded smoothly with the triphosphate without any adjustment. To investigate the s^2U containing NNNTPs in more detail, pppCG(s^2U) was prepared via *in vitro* transcription. The pppNN(s^2U) are the only s^2U-containing NNNTPs that can be efficiently synthesized by transcriptases.

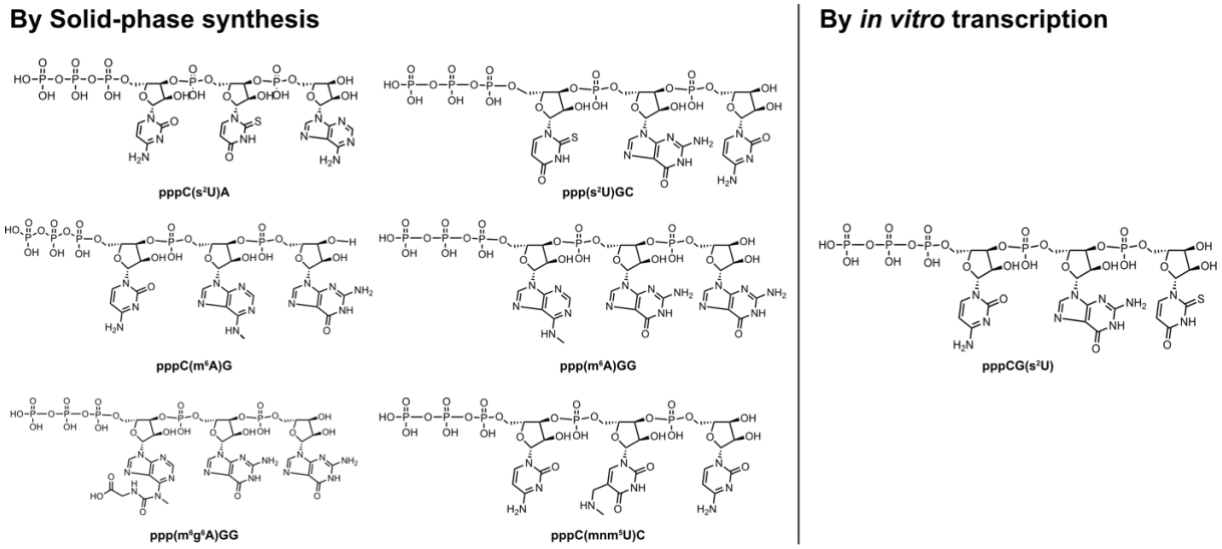


Fig. 2.37. Modified NNNTPs described in this study. *Chemical structures of NNNTPs employed in this study, characterized by their incorporated non-canonical nucleotides like s^2U, m^6A, m^6g^6A and mnm^5U in various positions.*

Name (Sequence)	HPLC	MALDI-TOF	
	0-20% B t_R (min)	*m/z* calcd. for $[M-H]^-$	found
pppC(s^2U)A	22.0	1133.0	1132.9
ppp(s^2U)GC	19.8	1149.0	1149.0
pppC(m^6A)G	24.0	1170.1	1170.0
ppp(m^6A)GG	24.5	1210.1	1210.5
ppp(m^6g^6A)GG	24.7	1311.1	1311.1
pppC(mnm^{5}U)C	14.1	1136.1	1136.0
pppCG(s^2U)	*in vitro* transcribed		

Table 2.5 Characterization data of NNNTPs prepared. *Name (sequence), HPLC retention time (0-20% buffer B) and MALDI-TOF mass spectrometry data of synthesized NNNTPs.*

2.4.2 Primer extension with modified triplet triphosphates

With the non-canonical NNNTPs in hand, we challenged the 5TU+t1.5 ribozyme that was prepared by the Holliger group to perform 3 times primer-extensions with the synthesized triplets with their respective template strands (**tP10UCG$_3$**, **tP10CUA$_3$**, **tP10CAG$_3$**, **tP10AGG$_3$** & **tP10CUC$_3$**). We employed a 5'Cy3-labeled primer (**P10**) and incubated it (0.5 µM) with the ribozyme (0.5 µM), the respective NNNTPs (5 µM) and template strand (0.5 µM) in the reported eutectic condition (200 mM $MgCl_2$, 50 mM Tris-HCl pH 8.3, -7 ˚C, 1 d)[183] and analyzed the results by denaturing PAGE (Fig. 2.38). By comparing the primer-extensions of the modified NNNTP with their respective canonical NNNTPs, we observed an effect of ribozyme stalling after the first triplet extension (1st Ext.) in almost all modified NNNTPs, except pppC(mnm^{5}U)C, which is the only one with a major groove modification and proceeded through all three extensions. In a separate experiment, we also tested the *in vitro* transcribed non-canonical triplet, pppCG(s^2U), with an 11-time extension template (**tP10CGU$_{11}$**). To our surprise, its polymerization proceeded well beyond the 1st Ext., despite a reduced reactivity when compared with the canonical pppCGU (Fig. 2.39). These results suggested that while the ribozyme polymerase can perform primer-extension with these modified NNNTPs, all of them suffered from a reduced reactivity. Furthermore, most polymerizations were stalled after the 1st Ext., and this stalling effect is apparently minor-groove sensitive and position sensitive, as major groove modified pppC(mnm^{5}U)C and 3'-end modified pppCG(s^2U) could proceed through the second (2nd Ext.) and third (3rd Ext.) triplet extensions.

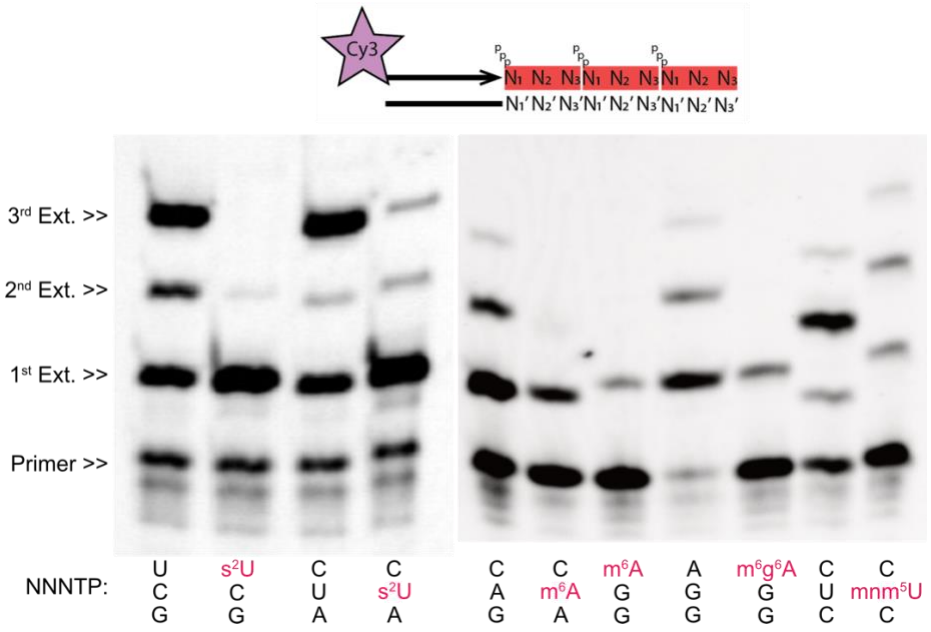


Fig. 2.38. Template-extension with non-canonical NNNTPs. *Ribozyme 5TU+t1.5 was challenged to perform primer-extension with modified NNNTP for three times and compared its results with the respective canonical NNNTP controls. (Modified nucleotides highlighted in red).*

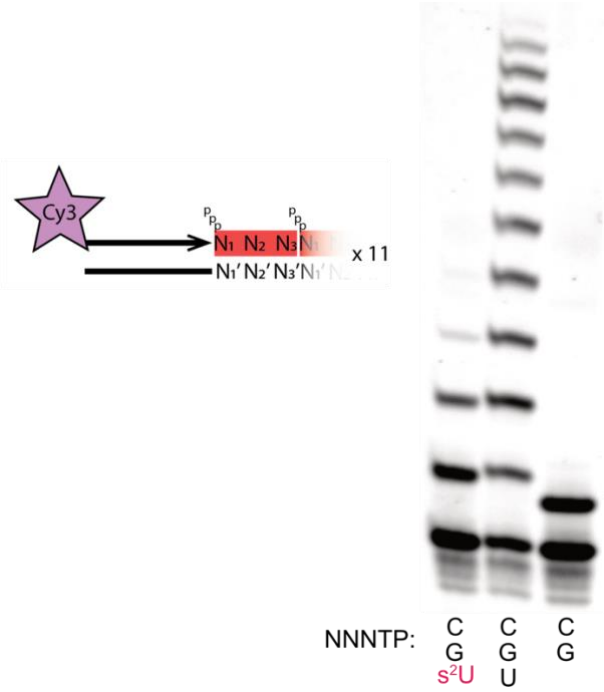


Fig. 2.39. Template-extension with pppCG(s^2)U. *Unlike other modified NNNTPs, pppCG(s^{2}U) could be incorporated beyond the 1st Ext., although with reduced reactivity compared to the canonical pppCGU. A negative control with pppCG was performed to show that the extension is sequence sensitive.*

To better understand the stalling effect, we looked extensively into the s^2U-modified triplets, as we have the NNNTPs with s^2U in all three different positions of a triplet, namely ppp(s^2U)CG, pppC(s^2U)A and pppCG(s^2U). We designed template strands **tP10UCGc3flk**, **tP10CUAc3flk** & **tp10CGUc3flk**, so that now the incorporation of modified NNNTPs is positioned at the 2nd Ext., flanked by CCC triplet incorporations at 1st Ext and 3rd Ext. (*i.e.* Primer **P10** → 1st Ext.:CCC → 2nd Ext.:NNN → 3rd Ext. CCC). We, again, incubated the reaction mixtures in eutectic phase and observed in the following PAGE analysis that the incorporations of ppp(s^2U)CG and pppC(s^2U)A but not pppCG(s^2U) stalled the ribozyme completely from incorporating the subsequent pppCCC at the 3rd Ext. (Fig. 2.40). This led us to suspect that s^2U modifications exist at the primer strand -2 and -3 but not -1 positions upstream of the ligation site would greatly inhibit the extension. This hypothesis, indeed, is in agreement with a reported result in upstream modification's influence on the template extension reactivity of the Z-polymerase ribozyme, an ancestor of 5TU+t1.5 (Fig. 2.41).[236] According to their result, the Z-polymerase ribozyme experienced a major reactivity drop when s^2U was incorporated at the -2 and -3 positions but almost no change at the -1 position upstream of the ligation site. Since both Z-polymerase and 5TU+t1.5 were evolved from the class I ligase and share the same reaction core, they should have a similar s^2U tolerance in the primer duplex. Examining the reported crystal structure of class I ligase's catalytic core revealed that the scaffold J1/3 of the reaction core docks at the minor groove of the template duplex and interact with the minor groove of nucleotides at these positions.[237] Hence, a replacement of U to s^2U at positions -2 and -3 could significantly destabilize the ribozyme-primer binding, as the thiol group is a poor H bond acceptor and therefore, the ribozyme falls off and the extension stalled.

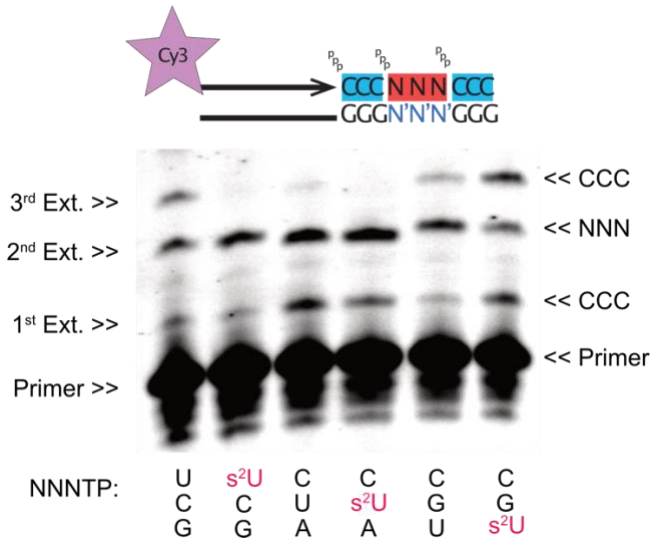


Fig. 2.40. Primer extension of s^2U-modified NNNTPs with CCC flanks. *PAGE analysis of the reactions showed that only pppCG(s^2U) did not stall the subsequent pppCCC incorporation.*

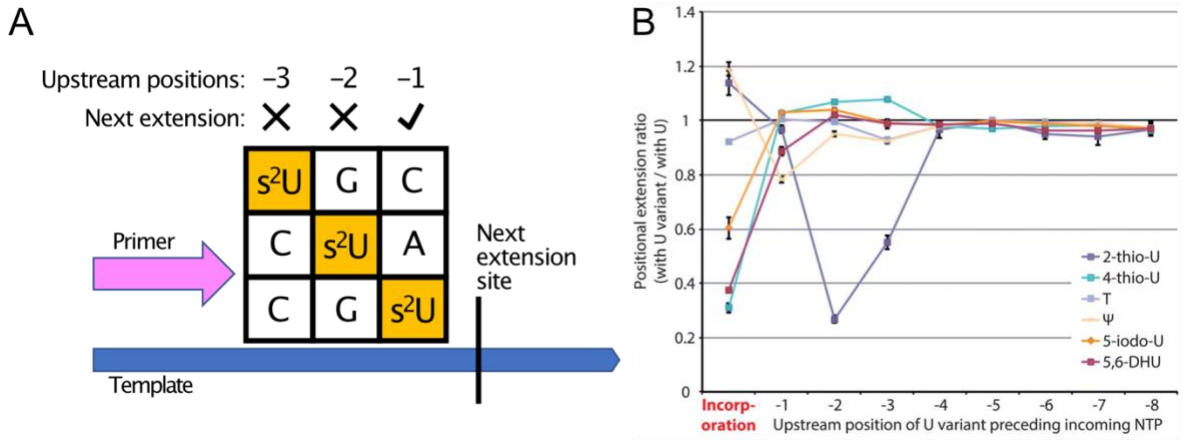


Fig. 2.41. Modification tolerance of ribozyme polymerases in upstream position. *(A) Upstream positions of s^2U in the primer strand of the tested NNNTPs after their incorporations. If the s^2U nucleotide is at the 3rd position of the NNNTP, like pppCG(s^2U), it will always end up at the -1 position of the primer strand. (B) Influence of various U modifications at different upstream positions of the primer strand on the extension efficiency of the Z-polymerase. A major drop was reported when s^2U was located at the upstream -2 and -3 but not -1 positions, (B) is reprinted from ref.*[236]

We next asked if this effect is the same when the s^2U nucleotides are located at the template strand but not the NNNTPs. We, therefore, prepared s^2U-modifed template strands (**tP10*NNN*$_3$_s^2U** & **tP10*NNN*c3flk_s^2U**) by using an s^2U-substituted NTP mixture for *in vitro* transcription. We performed primer-extension experiments with pppAGC, pppGAC and pppCGA, respectively, with their corresponding s^2U-substituted templates and compared their results with that of the canonical templates (Fig. 2.42). Surprisingly, we did not observe any extension stalling effect regardless of the position of the s^2U nucleotides at the template strand, showcasing an asymmetric structural interaction with the extending primer but not its template strand from the ribozyme.

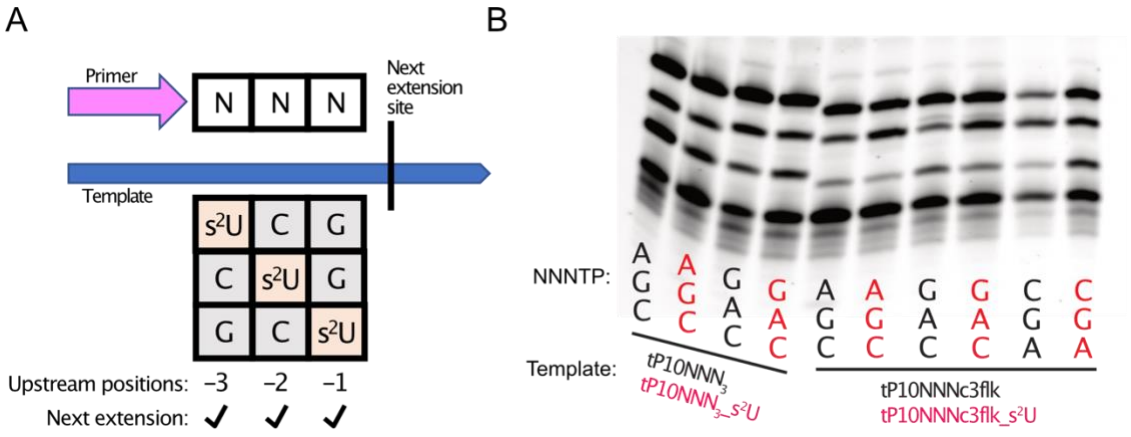


Fig. 2.42. Primer extension of canonical NNNTP with s^2U-modified template strands. *(A) Upstream positions of s^2U in the tested s^2U-modified template strands. (B) PAGE analysis*

revealed no extension stalling for s^2U-modified template strands (red, wells 2, 4, 6, 8, 10) when compared to the canonical ones (black, wells 1, 3, 5, 7, 9).

2.4.3 Effects on template-copying fidelity

We wondered if an s^2U-substituted template strand can lead to a higher primer-extension fidelity for the ribozyme than a canonical one, since s^2U can greatly improve A's incorporation accuracy by being unable to form a wobble base pair with G. Using the above canonical and s^2U-substituted, CCC-flanked template strands **tP10AGCc3flk** & **tP10AGCc3flk_s^2U**, we designed 8 independent experiments to analyze if the s^2U-modified template can increase the rate of 'correct incorporation' of pppAGC and reduce the 'wobble incorporation' of pppGGC (Fig. 2.43). Serving as positive controls, we provided the ribozyme with only the correct triplet pppAGC in sets 1 & 2, which use the canonical and s^2U-modified template strands, respectively. Repeating the experiments with only the wobble triplet pppGGC formed the negative controls of sets 3 & 4. We then designed competition experiments between the pppAGC and pppGGC incorporations. In sets 5 & 6, the respective canonical and s^2U-modifed templates were provided a 1:1 mixture of pppAGC and pppGGC. In the last two sets 7 & 8, we stressed the ribozyme for wobble incorporations by repeating the experiments with a 1:10, pppAGC:pppGGC mixture. We then excised the product from the unreacted primer **P10** from denaturing PAGE (Fig. 2.43B) and analyzed the results by NGS (see section 4. Experimental).

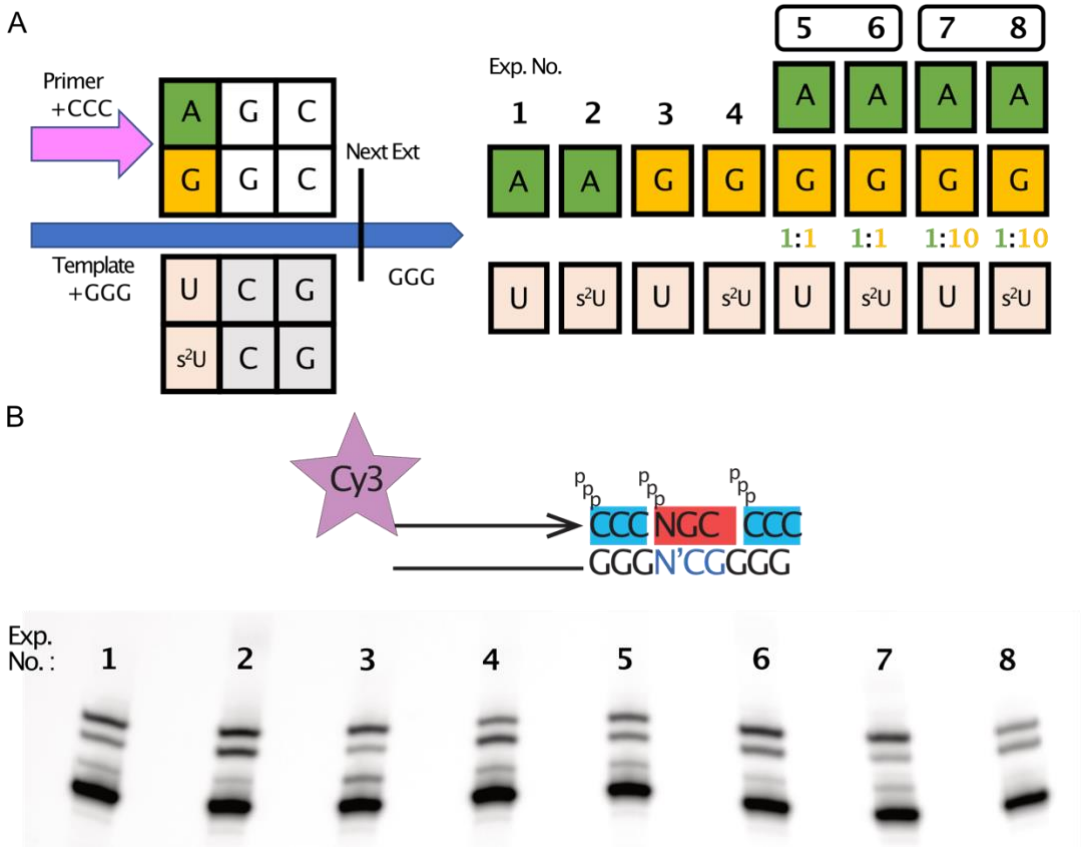


Fig. 2.43. Competition experiments of s^2U-modified template strands. *(A) The ribozyme was provided with either a canonical (1, 3, 5, 7) or an s^2U-substituted (2, 4, 6, 8) template for*

primer-extension. In the positive controls (1, 2), the ribozyme was only given the correct triplet pppAGC (green), whereas in the negative controls (3, 4) only the wobble triplet pppGGC (orange). In the competition experiments, the ribozyme was provided with a 1:1 (5, 6) or 1:10 (7, 8) pppAGC:pppGGC mixture to determine if the s^2U templates can reduce the wobble incorporation. (B) PAGE result of the reactions (1-8). The extension products were extracted for NGS analysis.

The NGS result is displayed in Fig. 2.45. Even though we tried to avoid excising the primer band in the gel (Fig. 2.44B), we still obtained around 3000-5000 reads of it on average in all setups (Fig. 2.44A). As the gel image in Fig. 2.43B showed that the product with only the 1st Ext. (primer **P10** +CCC) had the faintest bands in all experiment setups, it also accounted for least number of reads in the NGS result (Fig. 2.44A). Serving as an internal positive control, this indicates that the ribozyme could finish the 1st Ext. of pppCCC and proceed along the template strand to form products after the 2nd Ext. and 3rd Ext. in all experiments. The main investigation lies in the 2nd Ext., where a canonical U in the template strand was given to the odd number setups (Exp. No.: 1, 3, 5, 7) and an s^2U in the template strand was given to the even number setups (Exp. No.: 2, 4, 6, 8). A higher read counts of products stalled at the 2nd Ext. (primer **P10** +CCC +NNN) in all setups suggest that the ribozyme activity was slightly inhibited after this extension. Still, in most cases, full products after the 3rd Ext. (primer **P10** +CCC +NNN +CCC) accounted for most of the read count percentage (Fig. 2.44B). We, therefore, focus on the incorporation distribution of pppAGC and pppGGC in products stalled at the 2nd Ext. (Fig. 2.44C) and after the 3rd Ext. (Fig. 2.44D) in all setups. In the positive (setups 1, 2) and negative (setups 3, 4) controls, we observed only the incorporation of pppAGC and pppGGC, respectively, as only one of them was provided to the ribozyme in these cases. When the ribozyme was presented with a 1:1 mixture of pppAGC and pppGGC (setups 5, 6), we observed that the correct incorporation of pppAGC was strongly favored in all cases, while the wobble incorporation of pppGGC was slightly more suppressed in the s^2U-substituted template strand (setup 6) than the canonical one (setup 5) in both the 2nd Ext. and 3rd Ext. products. When the ribozyme was given a 1:10 pppAGC to pppGGC mixture, we observed a strong preference of the wobble incorporation of pppGGC in the canonical template strand (setup 7) but not in the s^2U-substituted template strand (setup 8). In the products stalled at the 2nd Ext., a weak favor for the correct incorporation of pppAGC was still preserved even under the concentration stress when an s^2U-substituted template strand was used. In the population of products after the 3rd Ext., we observed a strong favor of the correct incorporation over the wobble incorporation, suggesting that duplexes with a correct incorporation at the 2nd Ext. were more stable and allowed the ribozyme to proceed to the last CCC extension more efficiently than the one with a s^2U:G base pair.

Our results here demonstrated that molecular fossils, like s^2U, could benefit a self-replicating system in an early RNA world by improving its fidelity in template-copying. Even though many of them could not be incorporated as efficiently as the canonical ones, sparse incorporations of these modifications in specific location of an early sequence could play an important role in primitive gene regulations by improving replication fidelity (like s^2U) or stalling specific sequence from being replicated (like m^6g^6A and other described ribozyme-stalling modifications).

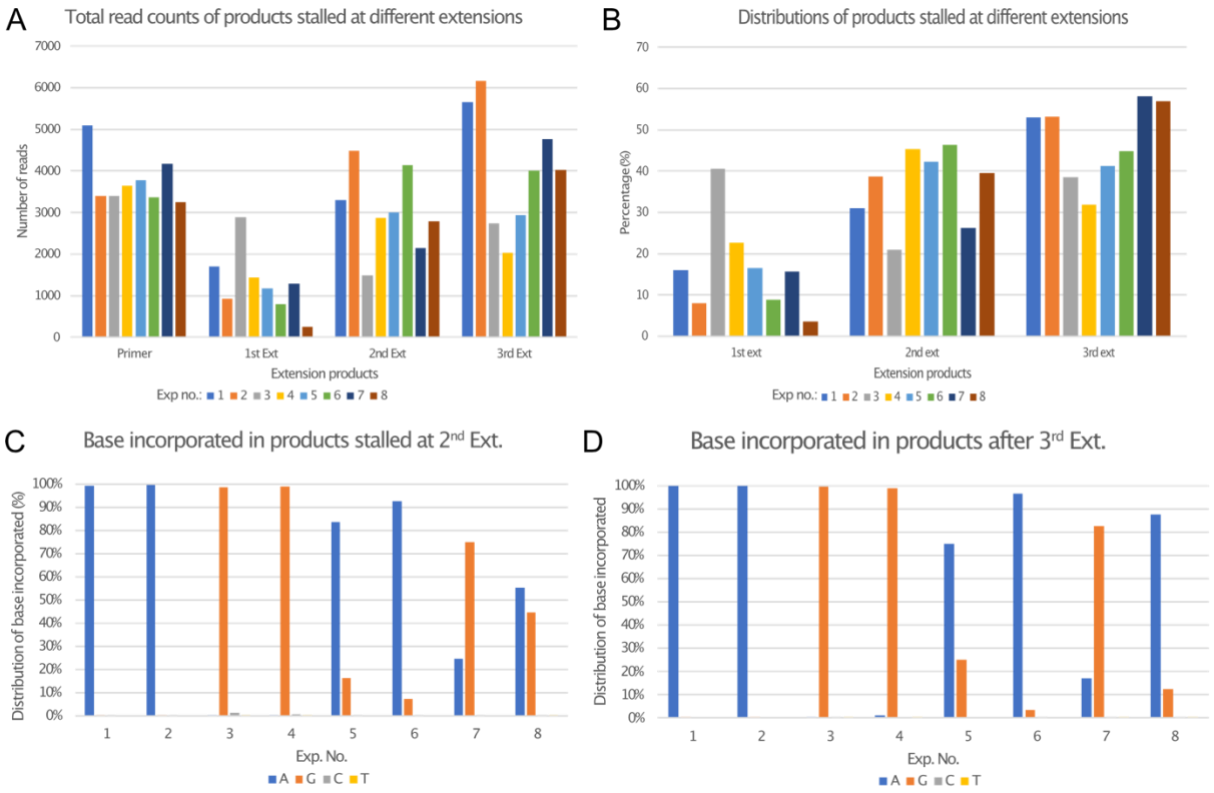


Fig. 2.44. NGS results of s^2U effect on template-copying fidelity. *(A) Number of reads of primer and each extension products of each experiment setups (1-8, color coded). (B) Percentage of total read of each extension product among all extension products in each setup. (C & D) Percentage of the incorporation at the site of interest in products (C) stalled at the 2nd Ext. and (D) after the 3rd Ext. in all setups.*

Name	3' Mod.	Sequence	5' Mod.	Polymer
P10	Cy3, Biotin	CUGCCAACCG	-	RNA
tP10CAG$_3$	-	CUGCUGCUGCGGUUGGCAG	-	RNA
tP10AGG$_3$	-	CCUCCUCCUCGGUUGGCAG	-	RNA
tP10CUC$_3$	-	GAGGAGGAGCGGUUGGCAG	-	RNA
tP10CUA$_3$	-	UAGUAGUAGCGGUUGGCAG	-	RNA
tP10UCG$_3$	-	CGACGACGACGGUUGGCAG	-	RNA
tP10CGU$_{11}$	-	**(ACG)$_{11}$**CGGUUGGCAG	-	RNA
tP10UCGc3flk	-	GGGCGAGGGCGGUUGGCAG	-	RNA
tP10CUAc3flk	-	GGGUAGGGGCGGUUGGCAG	-	RNA
tP10CGUc3flk	-	GGGACGGGGCGGUUGGCAG	-	RNA
tP10AGC3	-	GCUGCUGCUCGGUUGGCAG	-	RNA
tP10GAC3	-	GUCGUCGUCCGGUUGGCAG	-	RNA
tP10AGCc3flk	-	GGGGCUGGGCGGUUGGCAG	-	RNA
tP10GACc3flk	-	GGGGUCGGGCGGUUGGCAG	-	RNA
tP10CGAc3flk	-	GGGUCGGGGCGGUUGGCAG	-	RNA
tP10AGC$_3$_s^{2}U	-	GC(s^{2}U)GC(s^{2}U)GC(s^{2}U)CGG(s^{2}U)(s^{2}U)GGCAG	-	RNA
tP10GAC$_3$_s^{2}U	-	G(s^{2}U)CG(s^{2}U)CG(s^{2}U)CCGG(s^{2}U)(s^{2}U)GGCAG	-	RNA
tP10AGCc3flk_s^{2}U	-	GGGGC(s^{2}U)GGGCGG(s^{2}U)(s^{2}U)GGCAG	-	RNA
tP10GACc3flk_s2U	-	GGGG(s^{2}U)CGGGCGG(s^{2}U)(s^{2}U)GGCAG	-	RNA
tP10CGAc3flk_s2U	-	GGG(s^{2}U)CGGGGCGG(s^{2}U)(s^{2}U)GGCAG	-	RNA
(tx)tP10AGC$_3$	-	CTGCCAACCGAGCAGCAGCCTATAGTGAGTCGTATTAATTTCGC GGGCGAGATCGATC	-	DNA

(tx)tP10GAC$_3$	-	CTGCCAACCGGACGACGACCTATAGTGAGTCGTATTAATTTCGC GGGCGAGATCGATC	-	DNA
(tx)tP10AGCc3flk	-	CTGCCAACCGCCCAGCCCCTATAGTGAGTCGTATTAATTTCGCG GGCGAGATCGATC	-	DNA
(tx)tP10GACc3flk	-	CTGCCAACCGCCCGACCCCTATAGTGAGTCGTATTAATTTCGCG GGCGAGATCGATC	-	DNA
(tx)tP10CGAc3flk	-	CTGCCAACCGCCCCGACCCTATAGTGAGTCGTATTAATTTCGCG GGCGAGATCGATC	-	DNA
5T7 promoter	-	GATCGATCTCGCCCGCGAAATTAATACGACTCACTATAG	-	DNA
(tx)CGU_triplet	-	ACGUTATAGTGAGTCGTATTAATTTCGCGGGCGAGATCGATC	-	DNA
HDVlig	-	GGGTCGGCATGGCATC	C_3 spacer	DNA
HDVrec	-	GATGCCATGCCGACCC	-	DNA
kyleF_P10	Cy3, Biotin	d(GGATTCACTGCGATAGAGT)r(CCUGCCAACCG	-	DNA-RNA chimera
kyleF	-	GGATTCACTGCGATAGAGTC	-	DNA
Pxx_kyleF	-	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGT GTGCTCTTCCGATC**NNN**ATCACGGGATTCACTGCGATAGAGTC	-	DNA
p513HDVba	-	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA CGCTCTTCCGATCT**NNN**AGTCAAGATGCCATGCCGACCC	-	DNA

Table 2.6. Oligonucleotides described in section 2.4. *Left to right: oligo's nomenclature, sequence information and polymer type. Canonical oligonucleotides were ordered from IDT or Sigma, and s^2U-substituted RNA strands were in vitro transcribed from their respective DNA template strands whose names begin with (tx). For Pxx_KyleF & p513HDVba, the NNN are the barcodes for sequencing.*

2.5 Exploring functional RNA-peptides

In this last section, we will discuss our preliminary data of an ongoing research in the Carell group to search for biophysically or biochemically active RNA-peptides that can be anchored on a liposome surface and potentially functionalize it. In the first attempt, we tried to re-construct a known cyclic peptide sequence at the RNA-peptide hairpin and hoped that it can retain some degree of its binding properties. In the second attempt, we tried to conjugate a tripeptide catalyst onto an nm^{5}U-containing RNA oligo and anchor it onto a liposome surface by its complementary ges^{2}U-modified strand and investigate if the lipid membrane would change its catalytic properties. While both ideas only achieved very limited success, the syntheses of the materials and their biochemical properties were briefly investigated and reported here. (Information on all oligos described in 2.5 is listed section 4.2)

2.5.1 RNA conjugation of an ion-binding cyclic peptide

Cyclic peptides are polypeptide chains with a ring structure that greatly restrict its conformational freedom for optimal biochemical activities. The cyclizations are commonly achieved by linking one end to another with an amide bond or other chemically stable bonds, such as lactone, ether and disulfide. While complex cyclic peptides can be identified in nature, numerous small, artificial cyclic peptides of approximately 5-14 amino acid sequences were created as biochemical tools and potential therapeutic compounds.[238-240] Inspired by the structure of the RNA-peptide hairpin formed between (m^6)*aa*^{6}A and (m)nm^{5}U containing RNAs, we wondered if the conjugated peptide can resemble the properties of a cyclic peptide when it is weakly circularized by the base pair of an RNA duplex. Such RNA-peptide hairpin could then be anchored on the liposome surface with the (ges^{2}U)$_3$ tail and potentially allow functionalization. We, therefore, searched for small cyclic peptides consist of amino acid sequence whose side chains are compatible with our RNA on-bead coupling method, and possess basic biochemical properties that can be easily tested on our RNA-peptide construct. A cyclic non-canonical hexapeptide consists of repeating proline and 3-aminobenzoic acid (Abz) units was reported as a strong binder to multiple anions (Fig. 2.45A).[241,242] Since the width of a base pair is roughly 2-3 peptide bond long, we wondered if an RNA-peptide composed of an Abz-Pro-Abz unit could retain some weak anion-binding capabilities.

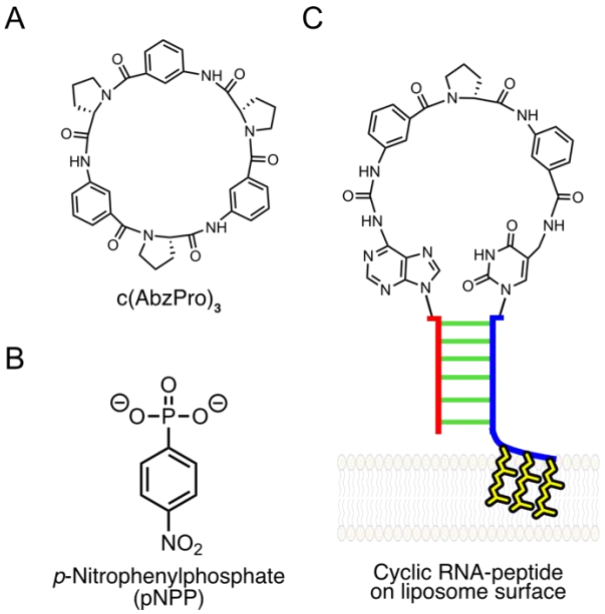
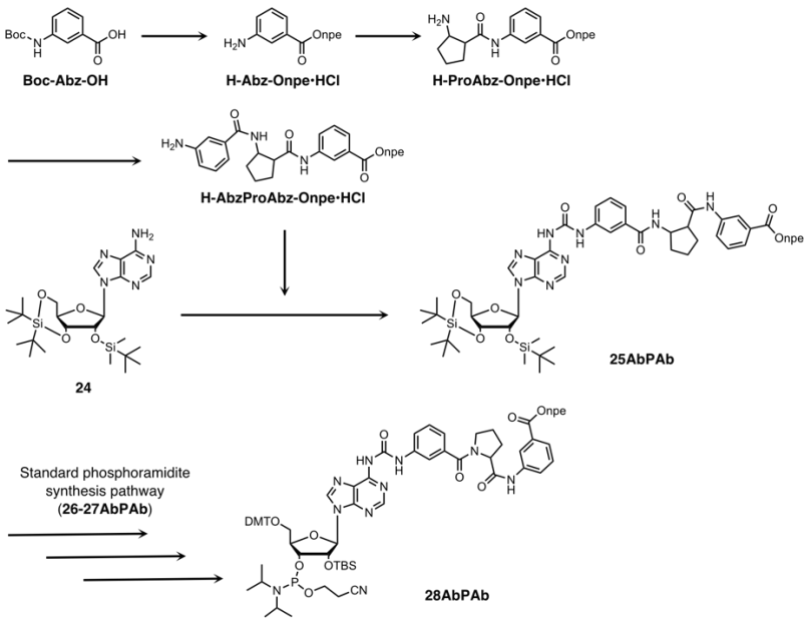


Fig. 2.45. Cyclic RNA-peptide on liposome. *Chemical structures of (A) cyclic hexapeptide (AbzPro)₃ and (B) p-nitrophenylphosphate (pNPP). (C) Synthesized lipid-anchoring cyclic RNA-peptide for potential pNPP capture on liposome surface.*

We, therefore, synthesized the phosphoramidite of Abz-Pro-Abz conjugated *aa*^{6}A (**28AbPAb**) according to the mentioned synthesis procedure with npe-protected Abz-Pro-Abz (Scheme 2.11), and performed its incorporation in an oligonucleotide synthesizer. After purification, we tested its template-directed peptide coupling with ges^{2}U- and nm^{5}U-modifed **ON39-nm^{5}U-geS** and detected a reasonable yield of the respective RNA-peptide hairpin formation (Fig. 2.45C).



Scheme 2.11. Synthesis of (AbPAb)^{6}A phosphoramidite. *Tripeptide analog H-AbzProAbz-Onpe was synthesized from conventional peptide coupling chemistry and coupled to protected adenosine **24** for its phosphoramidite synthesis (Scheme 2.9).*

We then investigated if this ges^{2}U modified RNA-peptide can bind to a UV-detectable anion, like *p*-nitrophenylphosphate (pNPP, Fig. 2.45B), and accumulate it on liposome surface (Fig. 2.46). To a 10 mM Egg PC liposome suspension, we added 20 µM of the RNA-peptide with 20 and 100 µM of pNPP, respectively, and incubated the samples for 30 mins. The mixtures were then spun down and resuspended with a pNPP-free buffer and were analyzed by HPLC. We compared the sample with a positive control, which contains the same amount of pNPP (2 nmol, '2 pNPP only') in the initial incubation mixture, and a negative control where the experiment was repeated in the absence of pNPP ('2 RNA-peptide only'). While expecting a signal of bound pNPP in the samples of 20 µM RNA-peptide with 20 and 100 µM pNPP in between these 2 controls, we detected no pNPP signal. We, therefore, assumed that the Abz-Pro-Abz conjugated RNA-peptide hairpin does not have any ion-binding properties as its parent cyclic peptide.

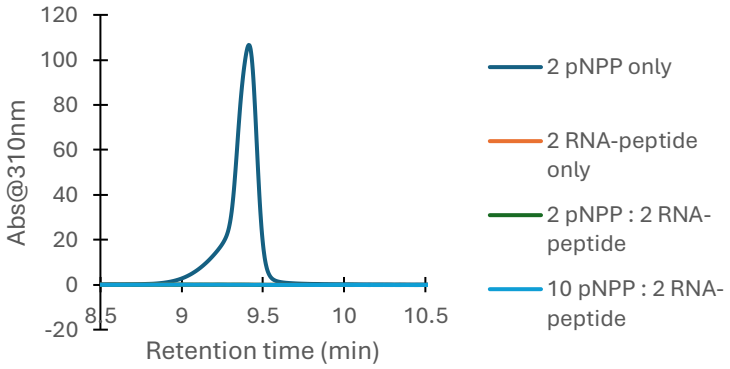


Fig. 2.46. Cyclic RNA-peptide anion capture experiments. *Overlay of HPL-chromatograms of (top to bottom) positive control with only 2 nmol pNPP; negative control with the RNA-peptide and liposome; experiment setup with 2 nmol pNPP, 2 nmol RNA-peptide and liposome; and with 10 nmol pNPP, 2 nmol RNA-peptide and liposome. The RNA-peptide did not show any binding properties. Absorbance was detected at 310 nm, which is the maximum absorbance for pNPP.*

2.5.2 RNA conjugation of a catalytic tripeptide

We next searched for small peptides with catalytic properties and compatible side chains to functionalize our RNA-liposome system. Tripeptide *D*-Pro-Pro-Glu-NH_2 (pPE-NH_2) was discovered to catalyze a 1,4-addition reaction between an aldehyde and a nitroolefin.[243,244] Using butanal and nitrostyrene as model compounds, it was demonstrated that the catalysis could lead to almost full conversion in organic solvents like 10% *i*PrOH in chloroform in the presence of *N*-methylmorpholine. To allow the catalysis to happen in water, a further study showed that replacing the terminal carboxamide to an *n*-dodecyl amine chain could promote the formation of emulsions, which dissolve the hydrophobic substrate nitrostyrene and facilitates the catalytic reaction in aqueous medium (Fig. 2.47).[245,246] Since the N-terminus of *D*-Pro is involved in the catalytic cycle,[247] we wondered if the tripeptide can be conjugated

onto an nm^{5}U oligo via its C-terminus and be anchored onto a liposome surface by its complementary ges^{2}U modified strand.

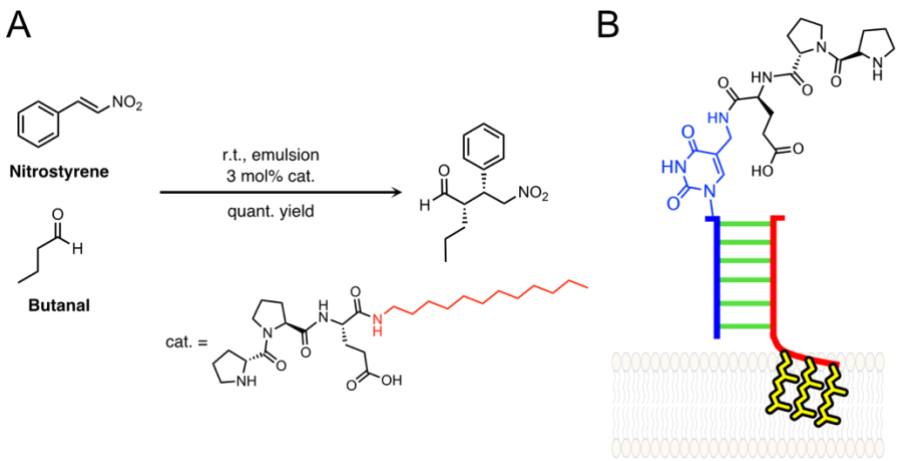


Fig. 2.47. Tripeptide catalyst on liposome. *(A) Reported[245] tripeptide catalyst D-Pro-Pro-Glu-NH-C$_{12}$ forms emulsion in water to dissolve nitrostyrene and catalyze its 1,4-addition with butanal in quantitative yield. (B) Proposed lipid anchoring of the tripeptide catalyst on liposome surface by modified RNA.*

With our on-bead coupling method, we successfully synthesized a D-Pro-Pro-Glu conjugated nm^{5}U RNA oligo (**ON37-pPE^{5}U**). With the help of the Huc group, we obtained the reported pPE-NH$_2$ from solid-phase peptide synthesis as a positive control. We then compared the 1,4-addition reaction among ssRNA **ON37-pPE^{5}U**, dsRNA **ON37-pPE^{5}U:ON30-geS** in the presence of liposome. We expected the **ONpPE^{6}A:ON30-geS** duplex could preserve the catalytic activity to some extent, as the liposome could dissolve the nitrostyrene while the ges^{2}U counterstrand could bring the pPE-conjugated nm^{5}U oligo onto the lipid surface. Since the reaction is concentration-dependent, and the required catalyst concentration is in the mM range, we setup 5 µL reaction mixtures with 25 nmol RNA-peptide to achieve such concentration. However, we did not observe any formation of product in any RNA-peptide reactions in our UPLC analysis, regardless the presence of liposome (Fig. 2.48). Being separated by the RNA duplex, the tripeptide was likely to be too far away from the lipid phase to facilitate the reaction, or the catalyst has its reactivity greatly reduced due to some secondary interactions with the RNA construct. All in all, our investigation here showcased that the nm^{5}U modification can conjugate a variety of canonical and non-canonical peptides and be brought to a proxy lipid phase by a ges^{2}U-modified complementary counterstrand. However, the RNA-peptide compound used did not display catalytic properties toward the investigated reaction.

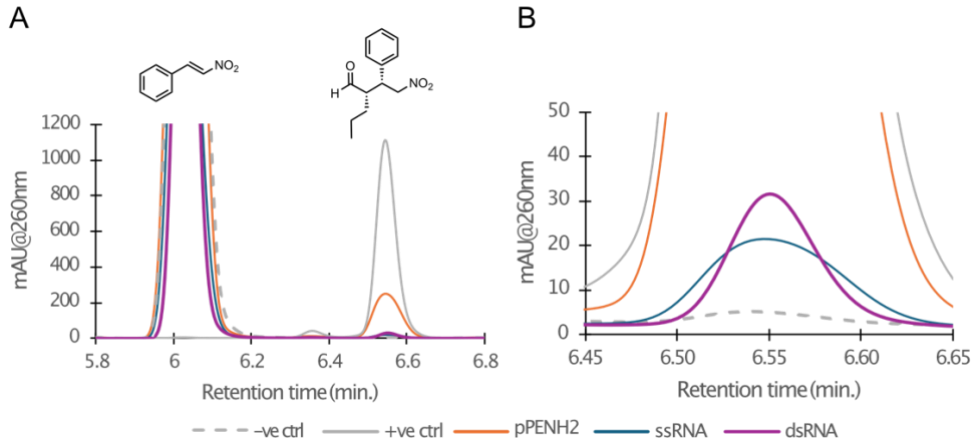


Fig. 2.48. RNA-peptide catalysis experiments. *(A) Overlay of UPL-chromatograms showing starting material, nitrostyrene, at 6 min. and intended product at 6.6 min. The -ve ctrl contains no catalyst with only the starting materials and liposome. A catalytic reaction was performed with the reported*[246] *catalyst pPE-NH*$_2$ *in 10% iPrOH in chloroform, showing a conversion with quant. yield. The product has a lower absorbance coefficient than the starting material. The ssRNA and dsRNA experiments contain the starting materials, liposome and ssRNA* ***ON37-pPE^{5}U*** *or dsRNA* ***ON37-pPE^{5}U:ON30-geS*** *respectively, both of them showed close to 0% conversion when compared to the -ve ctrl in the (B) zoomed-in region of the product area.*

3. Discussion

3.1 Historic debates that led to chemical evolution

The historic debate of life's origin resulted in the current theory of chemical evolution, in which small precursors, like HCN derivatives, were driven by natural forces in their environments to slowly react to form life's basic building blocks, like amino acids, monosaccharides and nucleos(t)ides. The accumulation of these building blocks allowed them to oligomerize by spontaneous chemical activations and thereafter, their complexity in structure and functionalization slowly developed as the length of these biopolymers increased. Among those, RNAs possess a wide range of catalytic capabilities while at the same time allow Watson-Crick base pairing for inheriting genetic information via template-copying. This combines phenotype and genotype in one molecule and allowed RNA sequences with life-essential physicochemical properties to be naturally selected and amplified by the principles of Darwinian evolution in the 'RNA world'.

The parallel formation of non-canonical and canonical nucleic acids from multiple proposed prebiotic pathways suggested that non-canonical nucleotides were likely to be incorporated into early RNA molecules. Coincidentally, many of these non-canonical nucleobases and nucleosides formed by prebiotic pathways overlap with the modifications found in nowadays tRNAs and rRNAs, who play crucial roles in translation regulations. These modifications are highly conserved across all lifeforms on Earth and are considered to be living 'molecular fossils' of a pre-LUCA system. Examining their chemical structures reveals a wide range of chemical functional groups, which could be of significant importance in early RNA's functionalization.

One of the greatest mysteries of the origin of life that cannot be explained by the RNA world model is the origin of translation. Translation is a template-directed peptide synthesis that takes place in the context of a specific genetic code dictionary. Without such dictionary, the result of translation would be the same as a random oligomerization of amino acids. The universality of the genetic codes in all life on Earth indicates a primitive, pre-LUCA system for co-differentiation between RNAs and amino acids. Today, the genetic code dictionary is defined by the aaRS-catalyzed tRNA aminoacylation, in which non-canonical nucleotides serve as identity determinants for accurate aaRS-tRNA recognitions. This inspired us to ask if these modifications also involved in a primordial RNA-amino acid pairing system. However, to correspond a modification with an amino acid, the system would have needed to first introduce this modification on specific RNAs. This leads to a three-layered paradox: How could modifications first be introduced onto RNA specifically? Then, how could these modified RNAs correspond to specific amino acids? Finally, how did these amino acid-loaded RNAs undergo template-directed peptide-synthesis?

3.2 "Molecular fossils" in tRNA

One can try to understand the origin of translation by studying tRNA, one of the key components of the translation machinery. Today, more than 170 RNA modifications were

identified, in which >80% of them were found on tRNAs, giving an average of 13 modification sites on each unique tRNA molecule.[248-250] These modifications were found to be highly conserved across all kingdoms of life,[251] and many of them were hypothesized to form simultaneously in several prebiotic pathways discussed above. Hence, the molecular fossil theory postulates that non-canonical nucleotides containing a wide range of chemical diversity must have incorporated into primordial RNA oligomers and served as essential components for an early RNA functionalization and therefore, preserved throughout the evolution. Today, most of these modifications hold significant roles in translation fidelity,[252,253] stress responses[254-256] and aaRS recognitions.[226,257] If we examine the tRNA molecules impartially, one can observe that the aminoacylated adenosine located at the 3' end of the acceptor stem overhang is not the only nucleotide that is conjugated to an amino acid (Fig. 3.1). In fact, amino-acid modified adenosines containing a carboxylic group, like g^6A,[258] t^6A,[259] and m^6t^6A,[260] were identified at position 37 adjacent to the anticodon, and modified uridines containing an amino group, like nm^5U and mnm^5U, can be found at the position 34, as the last letter of the anticodon.[261-263] Our model demonstrated that if they are put in close proximity, a template-directed peptide synthesis could take place, forming a peptide crosslink from the urea moiety of the *aa*^{6}A to the amine of the (m)nm^5U. This peptide synthesis model has an advantage over conventional prebiotic peptide synthesis models that utilize the 2'/3'-aminoacylated adenosine,[264,265] as most amino acid esters have half-lives of <500 mins at neutral pH,[266,267] whereas the half-life of urea is around 3.6 years.[268,269] The prebiotic plausibility of *aa*^{6}A and nm^5U were also discussed by Carell's[218] and Miller's[270] groups, respectively. In particular, Miller described the potential co-emergence of (m)nm^5U and other functionalized nucleobases from the reactions of 5-hydroxymethyluracil with amino acids and other prebiotic molecules, which could be fossils of how the RNA world transitioned into a DNA-protein world.[270]

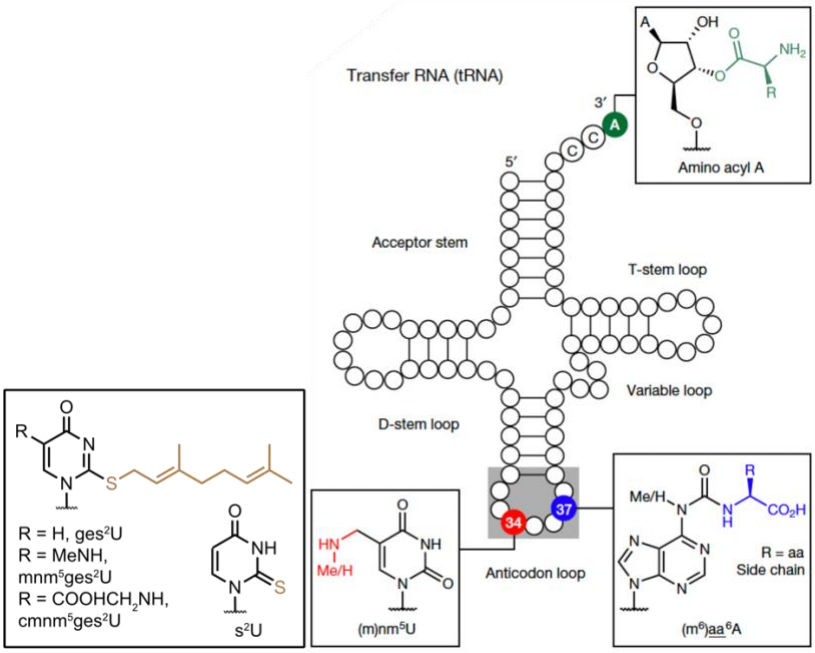


Fig. 3.1. Structure of tRNA and modifications described in this thesis. *Apart from the 3' acceptor end, amino acid modified (m^6)aa^{6}A and amine modified (m)nm^{5}U can be found at positions 34 and 37 respectively. Moreover, geranylated nucleotides like ges^{2}U and its analogs can be modified from s^{2}U and be found at position 34. Modified from ref.*[209]

The first part of this thesis (section 2.1) described the chemical synthesis of RNA oligonucleotides containing these modified nucleotides, namely (m)nm^{5}U and *aa*^{6}A.[208] In 1981, Caruthers reported a solid-phase synthesis method for nucleic acids utilizing nucleoside phosphoramidites,[271] which was found to be very efficient and adopted till today with minor adjustments (Fig. 3.2).[272] Briefly, the synthesis cycle involves an acid-labile 4,4-dimethoxytrityl (DMT) protecting group on the 5'-OH group of the ribose that get rapidly removed prior to the coupling. After that, the phosphoramidite moiety of another nucleoside will be activated and promptly coupled to the free 5'-OH group. The phosphate backbone and exocyclic amines of the nucleobases are protected orthogonally by base-labile groups throughout the cycle and are removed by ammonia after iteration of the coupling cycle to obtain the desired oligomer. In the case of RNA, the 2'-OH groups are commonly protected by the *t*-butyldimethylsilyl (TBS) group, which is eventually removed by fluoride. To invent novel phosphonamidites for *aa*^{6}A and (m)nm^{5}U, the carboxylic and amino groups must be protected with suitable groups that are compatible with the solid-phase synthesis cycle. This means they must be inert throughout the solid-phase synthesis, while removable without side-reaction afterwards. To this end, we described our general strategy to protect the amine of (m)nm^{5}U with the 2-(trimethylsilyl)ethoxycarbonyl (teoc) group and the carboxylic group of *aa*^{6}A with *p*-nitrophenylethyl (npe) group. These protecting groups can be removed orthogonally by $ZnBr_2$ and non-nucleophilic organic bases, like 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU), respectively. With a few additional protecting groups for specific amino acid side chains on the *aa*^{6}A, we successfully synthesized a series of RNA oligomers with *aa*^{6}A charged with different amino acids. We then investigated the basic physicochemical properties of these oligomers, like their capabilities to base-pair with canonical nucleotides.

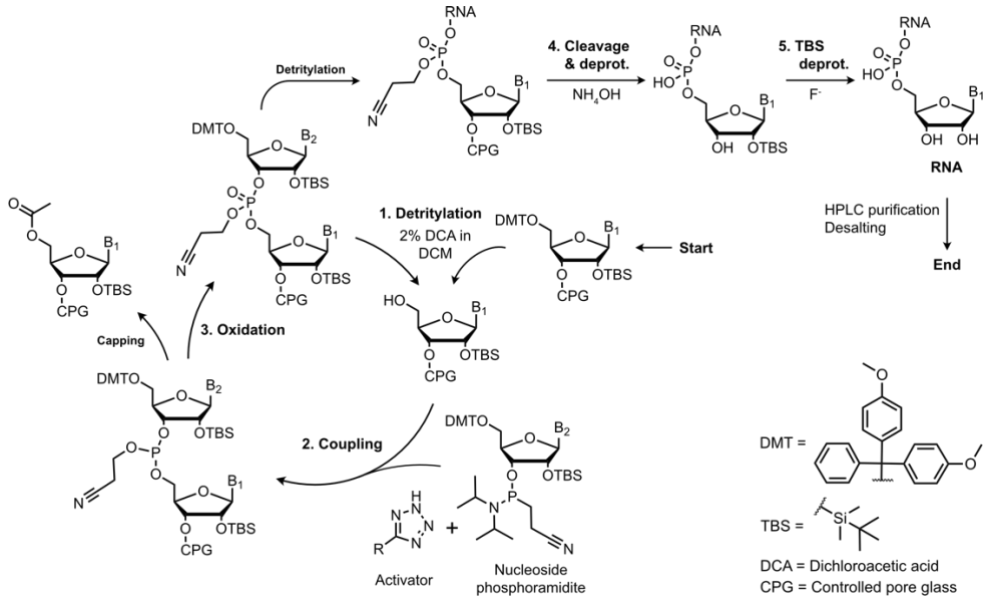


Fig. 3.2. Solid-phase synthesis of RNA oligonucleotides. *The cycle starts with detritylation of the 5'O-DMT group. The free 5'OH will react with the activated phosphoramdite of the subsequent nucleoside. Unreacted 5'OH will be capped by acylation. Phosphite backbone will be oxidized to phosphate in the last step. The cycle iterates until desired sequence acquired. Protecting groups on nucleobases and phosphate backbone will be removed and the oligo*

cleaved off from CPG bead by ammonia. For RNA, extra deprotection is needed to remove the 2' O-TBS group. The design of novel phosphoramidite for modified nucleotides must contain be compatible with this cycle.

In section 2.2, we described how template-directed peptide synthesis can be performed on RNA duplexes containing (m^6)*a*a^6A and (m)nm^5U.[209] We showed that a system of peptide elongation can be achieved by putting (m^6)*a*a^6A and (m)nm^5U in close proximity based on RNA duplex formation. The peptide coupling cycle started with the incubation of the modified RNA duplex with a carboxylic acid activator, like 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium (DMTMM), to form a hairpin-like product with a peptide bond bridging the two RNA strands. Incubation at high temperature and slightly acidic buffer allows the cleavage of the urea linkage of *a*a^6A in the hairpin, resulting in two RNA oligos with one bearing an m^6A and the other an amino acid-coupled (m)nm^5U. Iteration of this cycle allows subsequent peptide growth on the (m)nm^5U strand. Prebiotically plausible amino acid analogs, like aminonitriles,[224] were also loaded on *a*a^6A and tested in this system. Taken together, we demonstrated an alternative possibility for peptide synthesis based on amino acid modified nucleobases independent of the phosphate or the 2'/3'-OH group, featuring a more robust linkage independent from the conventional ester or phosphate anhydride bonds.

In section 2.3, we showed how 2-thiouridine (s^2U) and *S*-geranyl-2-thiouridine (ges^{2}U) modified RNAs can be introduced into this peptide synthesis system to add primitive selectivity on the peptide coupling in the presence of liposomes. In the later stage of the RNA world, functionalized RNAs might start to colonize the surface of liposomes to gain specificity of their chemical reactions. This transition phase is named by Cavalier-Smith in 2001 as *"obcell"*, indicating an early protocell stage where RNAs and amino acids with specific lipophilicity were co-accumulated and reacted selectively on the surface of lipid membranes, before an efficient mechanism was evolved to allow RNAs to be encapsulated into the interiors.[273,274] Modern day protein enzymes provide specific microenvironment in their reactive sites designated to their substrate's molecular size and physicochemical properties, like hydrophobic pockets, to acquire interaction specificity. Early RNAs, however, would have needed to acquire such microenvironment from an external source, like liposomes, before the emergence of proteins. Prebiotic lipidation of RNA to negatively charged or zwitterionic membranes has been previously reported in the presence of cations,[275,276] specific peptides[277,278] or phosphate backbone modifications.[279,280] Despite being efficient, these mechanisms lipidate all RNAs without sequence-specificity. Alternatively, several liposome-binding RNA aptamers were also reported, like the RNA10 from the Yarus' group.[281,282] Yet, their sequences are typically 80-120 nt. long, suffering from a questionable prebiotic plausibility. We, therefore, explored the possibility of lipidation of short RNAs functionalized by nucleobase modifications. At position 34 of bacterial tRNA anticondons for Lys, Glu and Gln, geranylated thiouridines were discovered,[283] which are promising molecular fossil candidates for RNA lipidation (Fig. 3.1). The prebiotic origin of s^2U was recently demonstrated,[284] whereas that for zwitterionic phosphatidylcholine lipids that we employed in the model was also recently suggested.[285] We showed an RNA-liposome system that allows sequence-specific geranylation of s^2U to ges^{2}U, and displayed how template-directed peptide synthesis described in the previous section can happen on ges^{2}U-modified RNA duplex with vague chemoselectivity in the presence of

zwitterionic phosphatidylcholine liposome, based on the peptide's lipophilicity. This suggested how an RNA and peptide co-recognition system could happen on the liposome surface based on their physicochemical properties.

In the fourth part (section 2.4), we summarized our preliminary results, obtained in a collaboration with the Holliger's group to polymerize non-canonical RNAs by the NNNTP polymerase ribozyme 5TU+t1.5, an engineered version of the reported RNA polymerase ribozyme.[286] Non-canonical nucleotide-containing NNNTPs that cannot be *in vitro* transcribed by transcriptases were prepared by solid phase synthesis according to a method reported by Meier.[234] Their compatibility with the triplet polymerase ribozyme was examined with various primer-extension experiments. Due to the synthetic limitations, most NNNTPs modified by m^6A, s^2U, g^6A and nm^5U were on the 1st or 2nd position. The initial result suggested that the ribozyme has higher tolerance towards modifications located at the major groove, such as nm^5U, than the others. Specific interest was put on the investigation of s^2U, as it has a potential to avoid wobble base pairs and improves the fidelity of the ribozyme's template-copying. Triplet triphosphates with an s^2U incorporated at three different positions were investigated with the ribozyme. Furthermore, its effects on the fidelity and efficiency on the ribozyme's template-copying was investigated by next-generation sequencing (NGS). The obtained result suggested how modified RNAs could be polymerized under specific conditions and influence ribozyme's catalysis.

In the last part (section 2.5), we summarized preliminary attempts to discover functional RNA-peptide chimeras and discuss possible directions of future investigations in this area. We described the synthesis of an RNA-peptide hairpin with an unnatural peptide sequence mimicking a reported short cyclic peptide, a hexamer repeat of *L*-proline and 3-aminobenzoic acid, that binds to anions with high affinity.[241,242] While the synthesis was successful, its binding to aromatic anions was not observed in water. In this respect, we next conjugated a linear catalytic tripeptide, *D*-Pro-Pro-Glu, which was reported to catalyze a conjugation between butanal and *p*-nitrostyrene in emulsion,[245,246] to an nm^5U-modified RNA. After the successful synthesis of the RNA-peptide, the further investigation of its hybridization with a ges^2U-containing complementary strand on the liposome surface showed that its catalytic activity was abolished. These results suggested that the catalytic activity of RNA-peptide would potentially require a higher order of secondary structures that position the amino acids in a suitable arrangement leading to active sites that are yet to be discovered.

The remaining discussion will mainly surround the lipid-anchored peptide synthesis model presented in this thesis. The model demonstrated that a proxy lipid phase can change the microenvironment of the RNA and influence its template-directed peptide synthesis. Interestingly, while the liposome catalyzes reactions driven by lipophilic reactants, like the geranylation with geBr, it inhibits those driven by hydrophilic reactants, like DMTMM-induced peptide coupling. In the study, we employed model reactants like geBr and DMTMM•Cl for geranylation and amino acid activations, respectively. To demonstrate the generality of the observed lipid-anchoring effect, the geranylation and peptide synthesis must be re-investigated in detail with prebiotically plausible geraniol and carboxylic acid activating mechanisms or moieties. Taking all into account, our model suggested how the three pillars of translation – specific RNA modification, co-recognition of RNA and amino acid/peptide and template-directed peptide synthesis – could co-evolve in a single system when functional

RNAs started to colonize the surface of liposomes in a late-RNA, early-protocell world. And one can imagine that liposomes formed by different lipid compositions could attract different RNAs and peptides with varying sequences and modifications, which slowly evolved to form the basis of the first genetic code dictionary.

3.3 Evolution history of translation and the genetic code

Yet, how should we place our proposed model in the grand timeline of the evolutionary history of the genetic code? Starting from a completely random amino acid-codon assignment, the early life systems must be first capable of vaguely distinguishing different groups of amino acids, instead of an individual one. This means that a single codon was assigned to more than one amino acids via a rudimentary amino acid-codon co-recognition process. Meanwhile, the codon-anticodon recognition of the first ribosome must have been loose and error-ridden, with its accuracy easily influenced by external stresses. At this period, the probability of translating any RNA entirely correctly was close to zero. Hence, life must have been relied on "statistical proteins" – a population of proteins translated from the same genetic information that carry varying peptide sequences but share similar physicochemical or catalytic (if any) properties.[287] Within this population, only a proportion of the synthesized proteins were functional, and the translation system evolved step-wise to reduce the probability of synthesizing relatively non-functional proteins and increase that of the functional ones. Eventually, the translation process became specific, and life started to rely on specific proteins coded by their corresponding genes under a universal genetic code. Even still, no two proteins in a cell are identical today.

However, these evolutionary steps could not be done by simply having "a better translation apparatus". As pointed out by Woese, it is easier to understand the reasoning here in reverse:[287] If a modern cell makes errors in translation above a certain threshold, and some of the faulty proteins are destined to become the next generation of translation apparatus, like a ribosome, then the new ribosomes could be more prone to make translational mistakes than its predecessor. This will in turn lead to the synthesis of similarly faulty ribosomes at an even faster rate. After cycles, this will cascade into a catastrophic situation in which the cell line translates in a completely error-ridden fashion, synthesizing random, non-functional peptides and becomes non-viable. What happened in the evolution of the genetic code could be the reverse of this sequence of events. This means, before the primitive cell could produce a more accurate set of translation apparatus, it first needed to translate more accurately with the given translation apparatus at any evolutionary time-point.

Woese, therefore, pointed out that the ancestral genetic code assignment must be necessarily ambiguous and the ancestral translation be necessarily error-prone,[287] and the processes of ambiguity-reduction on the genetic code and error-minimization on translation must have happened outside of the translation apparatus. It has to be emphasized here that codon-ambiguity is not translation error.[288] While errors can be caused by a less robust ribosome failing to ensure a correct codon-anticodon pairing, codon-ambiguity implies a more promiscuous assignment between codons and amino acids. However, a non-ambiguous genetic code assignment and an accurate translation did appear on Earth. This implies that both ambiguity-reduction and error-minimization must have co-evolved parallelly to allow a single

codon only codes for one and only one amino acid today. This co-evolution theory is, in fact, supported by modern biological evidence – the number of ribosomal proteins increased throughout evolution. In effect, with 57 in Bacteria, 68 in Archaea and 78 in Eukaria.[289] While some are universal and necessary for the functions, some have a stimulating effect but are fundamentally disposable.[290,291] As pointed out by Barbieri,[292] this follows the engineering principle that the accuracy of an automaton's function increases proportional to its size measured roughly by the number of its components.[293] However, in the evolution of translation, this increase could be perpetuated only if a higher number of statistical protein families could reappear in the descendent, and this could only happen if the ambiguity of the genetic code was reduced. This is because the phenotypes of a group of statistical proteins could only reappear in a descendent if the statistical differences between the two generations were not cancelled out by the ambiguity of the genetic code.[292] Therefore, the increase in the number of ribosomal proteins reinforced the ambiguity-reduction theory of the genetic code evolution, and the added ribosomal proteins, in return, contributed to the error-minimization of translation by making the ribosome heavier, more robust, and more resistant in response to the external environment, like thermal stress, pH stress, or the presence of interfering chemicals.

Woese proposed how the effects of translation errors could be minimized by codon arrangement at the initial stage of translation evolution.[287] The theory was deduced from two observations – (1) the error rate in the second letter of a codon is the lowest, followed by the first letter and then the third letter, which has the highest error rate, and (2) codons with a purine as the second letter generally has a lower error rate than those with a pyrimidine.[287] If one examine the nowadays codon assignment (Table 3.1), a general trend can be observed that most synonymous codons are in fact only differ in the third letter. Moreover, amino acids that possess a highly functional side chains (e.g. Lys, Glu, Arg, Cys, etc.) are coded by codons with purines as the second letter (i.e. NAN and NGN codons), whereas all relatively non-functional amino acids (e.g. Leu, Ile, Val, Ala, Phe, etc.) are coded by codons with pyrimidines as the second letter (i.e. NUN and NCN codons), especially NUN codons as U is the least accurate letter due to its potential Wobble base-pair with G. From a biological perspective, the differences of chemical properties among amino acids are not equally important. For instance, Ile and Leu are isomers. In a mutation scenario where a complete substitution of Leu to Ile happened in the genetic code, most proteins should conserve most of their functions. Thus, the distinction between Leu and Ile must have come later in evolution, as there was no high selection pressure for it in the earlier stage. In an ambiguous genetic code assignment system, it is of a higher importance to first evolve the capability to distinguish functional amino acids that involve directly in the catalytic process from "spacer" amino acids that merely place the previous in the correct positions or to provide a hydrophobic environment. Since the translation apparatus could not be improved by itself, this suggests that one of the key steps of the genetic code evolution was to assign functional amino acids towards the least error-prone codons, which are the codons with a purine as the second letter, leaving the more error-prone codons, whose second letter is a pyrimidine, for the relatively less functional amino acids.[287] Once translation was improved to some extent by this maneuver, the less accurate first letters can then be employed to further divide amino acid groups into subgroups, and the least accurate third letters to form the synonymous codons. This could reduce the effective translation errors to its minimum – reading mistakes made in the most error-prone letter would likely result in

the incorporation of the same amino acid, and mistakes made in the second most error-prone letter would likely result in the incorporation of an amino acid that possess similar chemical properties as the correct one – improving the effective translation accuracy of statistical proteins in the early translation system without the need of a more advanced ribosome.

First Letter (I)	Second Letter (II)								Third Letter (III)
	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG	STOP	UGG	Trp	G
C	CUU	Leu	CCC	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Table 3.1. Codon assignment table. *A general tendency can be observed that most synonymous codons differ only in the third letter while functional and "spacer" amino acids are separated by the second letter.*

However, later *in silico* simulations did not fully support Woese's theory. Novozhilov *et al.* pointed out that out of 10^{84} possible amino acid-codon assignments, the genetic code only went through half way up the optimality ladder for error-minimalization.[294] As pointed out by others, it was also possible that the genetic code did not only evolve around the translation errors, but also other parameters like protein structure stabilization,[295] introduction of additional information into the genetic code (like splicing and RNA secondary structures),[296] and facilitation of the correct protein folding.[297] Alternatively, it was also possible that this partial optimization passed the initial error-making threshold required for the earliest life form to translate the first generation of ribosomal protein families, which allowed translations to happen significantly more accurately and relieved the selection stress for further optimization of codon assignment.

While Woese's theory suggested how error-minimalization could have happened in the earliest translation system, this thesis suggested one of the many possibilities of how the ambiguity of the genetic code could be reduced in its earliest stage. As discussed earlier, early

phospholipids that can spontaneously assemble into liposomes must have co-existed with early oligopeptides and RNA oligonucleotides, and their interactions in the prebiotic world were inevitable. While the presence of liposome suppresses the coupling efficiency in our model, it is arguable that the effect could be reversed if a hydrophobic carboxylic acid activation mechanism was employed instead. For instance, prebiotic activations of amino acids must have also activated fatty acids parallelly to from amino acid-fatty acid conjugates due to their commonly available carboxylic groups. It has been shown that such conjugates, if composed of positively charged amino acids, can recruit RNA on the lipid surface.[278] Hence, similar mechanism would also give rise to fatty acid-cysteine conjugates, whose accumulation on the lipid membrane could have catalyzed cysteine-induced amino acid activation.[298] Furthermore, a reduction of peptide synthesis efficiency might not necessarily be a disadvantage. One of the challenges in evolution is parasite avoidance. It is known that nucleic acids with GC-rich templates[299,300] or peptides in the presence of β-sheet amyloids[301] can oligomerize rapidly. However, not all GC-rich sequences are functional, and big amyloid structures suffer from poor solubility. Hence, the capability to control the cancerous growth of biopolymers with undesired sequences and the degradation of them to recycle the building blocks should be of similar importance as the capability to catalyze their polymerization in evolution. From this perspective, the spontaneous hydrolysis of RNA that leads to its innate instability could be an advantage, so that ultra-stable parasitic duplexes or G-quadruplexes could eventually be degraded and their building blocks be reused for the formation of more functional sequences. In the early peptide synthesis systems, the growth of long hydrophobic peptide sequences also needed to be avoided to prevent precipitation. All amino acids contain a carboxylic group, which makes their terminals hydrophilic. Thus, their activations were likely to happen in an aquaeous environment and our model that suppresses the peptide growth of more hydrophobic peptides over the others showed how the selection pressure of insolubility avoidance could have driven the first ambiguity-reduction process – RNA sequences with hydrophobic nucleotides, like ges^{2}U, were selected in favor to correlate with hydrophilic amino acids in the presence of a lipid phase. This coincides with the fact that today ges^{2}U modifications are only found in tRNAs that code for Glu, Gln and Lys, all of which are hydrophilic amino acids.[283]

Another noteworthy discussion is the relevance of these "molecular fossil" modifications to the formation of the earliest translation system and the genetic code. Alongside rRNA and ribosomal proteins, tRNAs and aaRS are the oldest molecules in cell suggested by phylogenetic analysis.[291,302-304] This does not only suggest that the assignment of genetic code is one of the oldest features of life dated back to LUCA, but immense efforts were put throughout the evolution to conserve the genetic code. The key here is not just that it is universal, but its absence of change. As discussed earlier, the assignment of the genetic code was only optimized partially. Hence, the conservation of the genetic code is not a physical constraint, but a biological constraint. Throughout the entire evolutionary timeline, the ribosome and most of the components of life evolved. Nevertheless, under the same flux of selection pressure experienced by these systems, the genetic code did not change. This implies that there are biological mechanisms that actively and continuously restore its original structure.[292] Conventional Darwinian evolution states that life acquires new functions to adapt to new selection pressure imposed by the environment. In contrary, new mechanisms, manifested in gene replication or gene repair pathways, were constantly invented from Bacteria

to Eukaria as a dynamic effort to maintain the imperative genetic code and conserve it in time. The genetic code is, therefore, maintained against the selection pressure imposed by the external environment in a non-Darwinian fashion. Thus, tRNAs, as the closest molecules to the genetic code on which amino acids are loaded directly, must contain the most relevant trail on how the genetic code specificity was first acquired.

All tRNAs are 75-90 nt. long oligonucleotides with a basic cloverleaf structure that is highly conserved, which indicates that they were derivatized from a common ancestor. An ambiguity genetic code means that at the early stage, the primitive translation apparatus was using tRNAs that could accept more than one amino acids. Hence, if we analogize the aaRS-tRNA relationship as a lock-and-key system, the last universal common ancestor of all tRNAs must have been a master key that could fit into all locks at the beginning of the genetic code evolution.[305,306] Only when evolution proceeded did the aaRS-tRNA pairs start to derivatize and acquired interaction specificity. In most master-key systems, the master key is usually the simplest key. Complexity comes when you want a key to open one and only one lock but not the others.[305] During this derivatization, non-canonical nucleotides must have first introduced onto tRNAs, which then stimulated the evolution of their cognate aaRS to recognize them specifically. This is because the refinement of aaRS binding sites required the translation system to synthesize better statistical proteins beyond the statistical error caused by an ambiguous genetic code, but tRNAs with certain non-canonical nucleotides could correlate with certain amino acids based on their physicochemical properties.

3.4 Possible ambiguity-reduction theories on tRNA

Under this background, a few theories were proposed on how tRNAs and amino acids could have formed the earliest recognition interactions. Initially, Woese reasoned this to nucleobase-amino acid H bond interactions.[307] All nucleobases and amino acids contain H bond acceptors and donors. However, the more hydrophobic is the amino acid, the lesser the energy penalty for it to break its H bonds with water and form new ones with the base, as it is compensated by the new hydrophobic interactions. In a series of paper chromatography experiments using pyridine, 2-methylpyridine and 2,6-dimethylpyridine as model nucleobases, Woese obtained that the polarity index of amino acids that could vaguely correlate with the codon assignment table.[307] However, it is unsure if the nucleobases and the amino acids could reach such high concentrations in the prebiotic world so that the effect did not become negligent due to water dilution. It is also unsure how other co-existing precursor molecules would affect this type of interactions in the same period.

Most of the remaining theories can be categorized as stereochemical theories, which state that the ancient codons or anticodons of tRNAs must have the capability to bind to their cognate amino acids specifically in an aptamer-like fashion.[308,309] One of the most notables, and also most explored experimentally, is the Hopfield theory (Fig. 3.3).[206] Hopfield examined the modern tRNA cloverleaf structure and deduced that it is less stable than a potential hairpin structure that it could form. Hence, he proposed that before the cloverleaf structure had evolved, tRNAs must have existed as a hairpin, with the TψC loop as the terminal bulge and the right half of the molecule – from the acceptor stem 3'-end adenosine to the right half of the anticodon

loop – forming a stable duplex stem (Fig. 3.3). In this structure, the 3'-end adenosine would be positioned at the anticodon pocket, with the remaining left half of the molecule, including the D loop and the left halves of the anticodon loop and acceptor stem, folded as an ancient aptamer that specifically binds to the cognate amino acid, facilitating its specific aminoacylation.[206] One may argue that the remaining sequence of the aptamer mutated throughout the evolution to form the cloverleaf structure and the anticodon is the remain of the ancient aptamer. In fact, quadruplet codons do exist in nature,[310-312] which suggests that the codons might have been shortened throughout evolution. Another similar theory proposed by de Duve suggested that a potential "second genetic code" might have existed at the acceptor stem that predates the modern one.[313] The reason is that the anticodon loops is too distant to impose any effect on the aminoacylation at the acceptor stem. Hence, an abandoned genetic code adjacent to the aminoacylation site must have facilitated the early specific amino acid loading.[313] The main problem of this theory is, even if it was proven to be correct, it still could not answer how the secondary code could be related to the modern ones, so the core question would remain fundamentally unanswered. Nevertheless, based on these ideas, Sutherland's and Jäschke's groups performed experiments on a library of model acceptor stems with randomized 3-letter code upstream of the amino acid loading site.[137] As discussed in the introduction section, the comparison of sequencing results showed that when the library was incubated with an individual amino acid, a preferred loading sequence was observed. However, when this sequence was incubated with a mixture of amino acids, it failed to conjugate with their specific amino acids preferentially in most cases.[137] One must keep in mind that in this study, only the canonical nucleotides were tested due to the constraints of the current sequencing technology, as many non-canonical nucleotides cannot properly base-pair or be read by polymerases. However, the ancient tRNA aptamer, if existed, must have contained a plethora of non-canonical nucleotides. This is because, again, prebiotic pathways must have given rise to a mixture of canonical and non-canonical nucleic acid materials concurrently, which were oligmerized into fragments by spontaneous chemical activations. Instead of the modern-day polymerization that occurs as a single-nucleotide addition, the ancient tRNAs were more likely to be synthesized by ligation of these fragments directed by a loosely binding template, with non-canonical nucleotide bulges existing intermittently between proper base-pairs. The Hopfield theory can, therefore, neither be justified nor falsified until one day a library consists of both canonical and non-canonical nucleotides can be efficiently produced and sequenced.

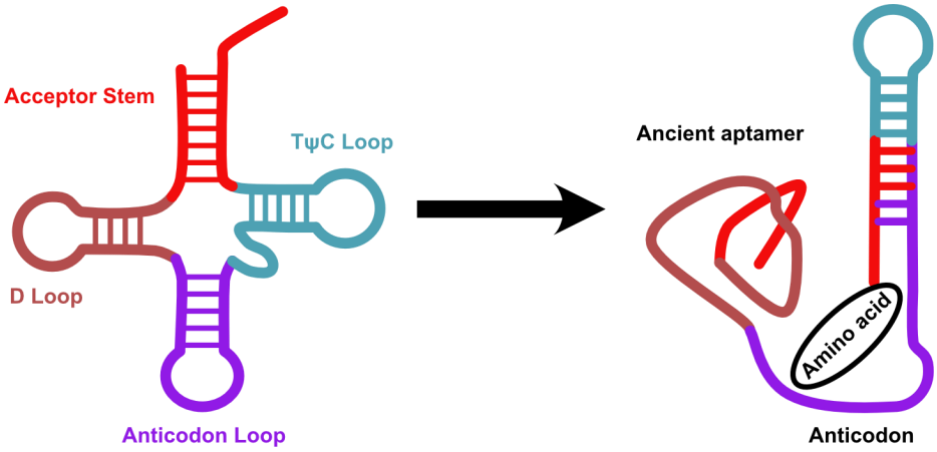


Fig. 3.3. Hopfield's stereochemical theory. *Proposed theory by Hopfield that early tRNAs could have folded into a hairpin structure with the aminoacylation site facing the anticodon which was part of an ancient aptamer or ribozyme.*[206]

The model presented in this thesis argues that the second genetic codes could have been tRNA modifications instead of a sequence. With the aid of contemporarily co-existing lipids, ancestors of tRNA composed of hydrophobic non-canonical nucleotides could recruit lipid membranes to selectively suppress the elongation of hydrophobic peptides heavier than the others, thus, favoring its co-accumulation with hydrophilic peptides. One must keep in mind that a nucleotide is roughly 2-4 times bigger than an amino acid. Hence, if the earliest genetic code was established based on the overall physicochemical properties of an amino acid-RNA conjugate, the "amino acid" part must have been an oligopeptide so that it is big enough to affect the overall properties of the molecule. Hence, it makes sense that the trend in peptide coupling yields was only observed in the reactions of longer peptides presented in our data. In many prebiotic pathways, amino acids were used as building blocks to form nucleotides. Thus, at the time of the RNA world, the concentration of amino acids in the primordial soup must be drastically higher than the RNA molecules to support the continuous synthesis of RNA's building blocks. These amino acids were also likely to be oligomerized by chemical activations, with the ones that share a similar hydrophobicity more likely to oligomerize with each other due to potential interactions. Hence, the early ambiguous genetic code might only be able to distinguish groups of peptide fragments ambiguously instead of groups of amino acids based on their hydrophobicitiy. The high concentration of amino acids over RNA also means that it is thermodynamically possible for amino acids to react with available functional groups within an RNA oligonucleotide, like the exocyclic amines and 2'-hydroxyl groups of a nucleotide, forming amino acid-modified nucleotides like *aa*^{6}A and the aminoacylated A, respectively. But why did translation happen on the aminoacylated A but not the other amino acid-conjugated nucleotides like *aa*^{6}A? During the template-directed replication of RNA, an amino acid-conjugation at the nucleobase would only hamstring a nucleotide's capability to base-pair at a regional level. Below a certain level, it would weaken but not completely blocking the template extension process. However, if an amino acid was conjugated at the 2'/3'- hydroxyl group, it may block the elongation process completely and terminate the replication. Hence, the earliest ribosome was perhaps the earliest detoxification system that efficiently remove amino acids

conjugated at the 2'/3'- hydroxyl group. In a competitive template extension scenario with high concentration of activated amino acids, if a part of a template could fold into a ribozyme that can efficiently remove any aminoaclyated amino acid during the process any allowed smoother template extension, it surely could have a selection advantage over other templates that could not.

3.5 The finalized model

The presented model pictures that the ancestral tRNA might well start as a short, modified RNA fragment with potential geranylation sites serving as the "second genetic code". This was caused by tolerating intermittent incorporations of (m)^6*aa*^{6}A, (m)nm^{5}U and (ge)s^{2}U during the fragment ligation process of primitive RNA replication. A template-directed peptide synthesis cycle could then be formed between (m)^6*aa*^{6}A and (m)nm^{5}U, with the potential to co-accumulate hydrophilic peptides with (ge)s^{2}U modifications in the presence of liposomes. At this period, multiple template-directed peptide synthesis mechanisms could have co-existed. In addition, we showed that the geranylation of s^{2}U could be facilitated by liposome in the presence of an activated geraniol and be introduced site-specifically, allowing a possible mechanism to be evolved to geranylate s^{2}U based on sequence recognition. As evolution went by, the length and structural complexity of tRNAs increased, which may require more and more hydrophobic nucleotides like ges^{2}U to allow its lipidation. However, the aminoacylation process would eventually be taken over by the first generation of statistical aaRS proteins and the contribution of hydrophobic nucleotides gradually reduced, so as their number in the tRNA. Eventually, all peptide synthesis pathways would be unified by the ribosomal translation that takes place at the 3'-end aminoacylated A. However, the frequent co-occurrence of ges^{2}U and hydrophilic peptides might have irreversibly shaped the evolution landscape of the binding sites of the early statistical aaRS proteins to recognize the ges^{2}U for the aminoacylation of hydrophilic amino acids. Until today, like the conservation of the genetic code, new proteins were invented to actively maintain the (ge)s^{2}U or its further modified se^{2}U at precise positions, which remain as essential aaRS recognition sites for a faithful aminoacylation process. While our model suggested how ges^{2}U modifications could have selected for tRNAs coding for hydrophilic amino acids, other selection models must have existed to select other modifications for other tRNAs specifically, which can be further explored.

4. Experimental

4.1 Materials and methods

General information of materials and instrumentations

Chemical and anhydrous solvents were acquired from commercial sources like *Sigma Aldrich, Carbosynth, TCI, ABCR, Acros Organics* or *VWR* and used freshly without further purification. Routine ^{1}H NMR, ^{13}C NMR and ^{31}P NMR were measured on a Bruker AVANCE NEO™ 500 MHz spectrometer. Deuterated solvents used are indicated in the characterization and chemical shifts (δ) are reported in ppm. All NMR *J* values are given in Hz. NMR spectra were analysed using MestReNova software version 14.3.0. High Resolution Mass Spectra (HRMS) were measured on a Thermo Finnigan™ LTQ-FT MS™ 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 solution.

General information on oligonucleotide synthesis

Phosphoramidites of canonical nucleotides (Bz-rA-CE, Dmf-rG-CE, Ac-rC-CE and rU-CE) were purchased from *LGC Biosearch Tech* or *Sigma-Aldrich*. Oligonucleotides (ONs) consists of only canonical nucleotides and 6-FAM labels were ordered from Ella Biotech (Germany). Otherwise, oligonucleotides were synthesised on the corresponding 1.0 µmol CPG cartridges from LGC Biosearch Tech. High Load Glen UnySupport™ was used as solid supports for strands containing non-canonical nucleotides at the 3' end. Oligos were synthesised on an Applied Biosystems 394 DNA/RNA Synthesizer with a standard phosphoramidite chemistry with a 'DMT-OFF' setting. Solution 2% DCA in DCM was used 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.

Synthesis of (m^6)*aa*^{6}A and (m)nm^{5}U modified oligonucleotides

For the deprotection of the *para*-nitrophenylethyl (npe) group in oligos modified by (m^6)*aa*^{6}A, the solid support beads were suspended in 1 mL 10% DBU in THF at r.t. for 2 h. After that, the supernatant was discarded, and the beads were washed with 1 mL THF thrice. The beads were then dried and proceeded to cleavage and 2'TBS deprotection.

For the deprotection of the 2-(trimethylsilyl)ethoxycarbonyl (teoc) group in oligos modified by (m)nm^{5}U but not 2-thiouridine, the solid support beads were suspended in a saturated solution of $ZnBr_2$ in 1mL of 1:1 $MeNO_2$/*i*PrOH and incubated at r.t. overnight. After

that, the supernatant was removed and the beads were washed with 0.1 M EDTA in water (1 mL) and water (1 mL), once each.

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

The solid support beads were suspended in a 1:1 aqueous solution mixture (0.8 mL) of 30% NH_4OH and 40% $MeNH_2$. The suspension was heated at 65˚C (10 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 0.2 mL water once. 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 90 mins. Finally, the ON was precipitated by adding 3 M NaOAc in water (25 µL) and *n*-butanol (1 mL). The mixture was kept at -80˚C for 2 h and centrifuged at 4˚C for 1 h. The supernatant was removed and the white precipitate was lyophilized.

For **ON1-geS** which underwent direct ges^{2}U incorporation and oligos with peptide conjugation, the beads were suspended with (0.8 mL) of 30% NH_4OH at r.t. for 14 h. The supernatant was collected, and the beads were washed and dried similarly as above. Followed by that, the crude was reacted with triethylamine trihydrofluoride (125 µL) in DMSO (100 µL) overnight. The ON was then precipitated, centrifuged and lyophilised as above.

Purification of oligos 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: unless otherwise stated, 0-40% of B for ONs without *S*-geranyl-2-thiouridine and 0-80% of B for ONs with *S*-geranyl-2-thiouridine 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: 0-20%, 0-30%, 0-40% or 0-80% of B in 45 min, similar as above. Flow rate = 1 mL·min^{-1}. Finally, the triethylammonium acetate was removed using a Sep-Pak® Plus Short C18 cartridge from Waters and dried with a SpeedVac concentrator.

Determination of the concentration and mass of the synthesized oligos

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 their closest canonical counterparts.

The synthesized oligonucleotides (2-3 μL) was desalted on a ø 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).

Post-synthetic geranylation for 2-thiouridine containing oligos

S-geranyl-2-thiouridine containing ONs were geranylated from their corresponding 2-thiouridine oligos according to a reported method with minor modifications.[228] Briefly, dried and desalted ON was dissolved in 66% EtOH to achieve a final concentration of 2 mM. Geranyl bromide (250 eq.) was added to the solution and the mixture was vortexed thoroughly to fully dissolve the geranyl bromide. After that, Et_3N (250 eq.) was added and the solution was incubated at r.t. with 1400 rpm on a thermoshaker for 6 h. The reaction was then quenched by adding 5X volume of 20% EtOH in HPLC buffer A and dried with a SpeedVac concentrator. The crude was then resuspended in water, filtered, purified by HPLC and desalted according to the above method.

Post-geranylation teoc-deprotection for oligos containing both ges^2U and nm^5U

The geranylated, HPLC purified, desalted and dried oligo was cooled on ice and cold 1 M TBAF in THF solution was added to get a 0.5 - 1mM solution of the oligo. The solution was then incubated at 10 ˚C for 1.5 h. After that, it was cooled in an ice bath for 5 min and diluted with 5 mL cold HPLC buffer A. The solution was lyophilized overnight and purified by HPLC and desalted.

Coupling of amino acids and peptides to oligos on the solid support

Oligos with more than one amino acids attached to the $(m^6)aa^6A$ modification were prepared with this on-bead coupling method. Briefly, the npe group was first removed as described above. The beads were then suspended in 0.6 mL DMF solution containing the corresponding npe-protected amino acid or peptide (100 μmol), DMTMM•BF_4 (100 μmol) and DIPEA (200 μmol) at r.t. for 1 h. It was then washed with DMF twice and proceed to another npe deprotection protocol. It was then carried on by the aforementioned oligo purification process.

Oligos with more than one amino acid attached to the $(m)nm^5U$ were prepared similarly. After the teoc-deprotection of the initial $(m)nm^5U$, the beads were suspended in a solution of Boc-protected amino acid or peptide (100 μmol), DMTMM•BF_4 (100 μmol) and DIPEA (200 μmol) at r.t. for 1 h. It was then washed with DMF twice. To deprotect the Boc group, the beads were then incubated in 1:1 TFA/DCM solution at r.t. for 30 mins. After that, the supernatant was discarded, and the beads washed with DCM twice. They were then dried and proceeded with standard cleavage and deprotection protocol.

UV-melting measurements

For results in section 2.1, samples were prepared to a concentration of 10 mM sodium phosphate buffer (pH7), 150 mM NaCl and 4 μM dsRNA, with a total volume of 80 μL. For results in section 2.3, samples were prepared to a concentration of 100 mM Tris-HCl buffer, 100 mM NaCl and 10 μM dsRNA, with a total volume of 400 μL. UV melting experiments were performed on an Jasco V-650 UV-Vis Spectrophotometer, using 1-cm path length quartz cuvette. The samples were scanned from 5 to 85 °C with a temperature incremental rate of 1.0

˚C/min. The temperature was hold at 85 °C for 5 mins before a reversed scan was performed, scanning from 85 to 5 °C with the same rate. The denaturing transition was monitored at 260 nm. Baseline corrected data was then smoothed over 5 points and its first derivative was plotted in Microsoft Excel. The final melting temperature was obtained by averaging the melting temperatures in the 5 repeated forward scans.

HPLC calibration curves using canonical oligos (**CON1-10**)

Canonical oligonucleotides (**CON1-10**) were used for the development of HPLC calibration curves. Separate stock solutions of **CON1-10** 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 **CON1-10** 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 $mL \cdot min^{-1}$). 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.

Coupling reactions between (m^6)*aa*^{6}A and (m)nm^{5}U modified oligos

For results in section 2.2, 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. The samples were then analyzed by HPLC and MALDI-TOF. The buffer and pH used for each activator are listed as follow:

1. DMTMM•Cl, MES buffer pH 6
2. EDC/Sulfo-NHS, MES buffer pH 6
3. MeNC, MES buffer pH 6
4. DTT (50 mM), borate buffer pH 8 (for **ON10-Gcnm**)

For results in section 2.3, stock solutions of MES pH 6 buffer (400 mM), NaCl (1 M) and DMTMM•Cl solution (500 mM) were prepared in water in advance. For setting up a reaction, equimolar of *aa*^{6}A and nm^{5}U containing ONs (1 nmol each) were mixed in water and dried in a SpeedVac concentrator. Subsequently, the oligos were re-dissolved in a 2-fold concentrated reaction buffer (i.e. 200 mM NaCl, 200 mM MES pH=6.0 buffer), with or without 10 mM Egg PC liposome suspension (see section below). The reaction was then diluted to a final concentration of 100 mM NaCl, 100 mM MES pH=6.0 and 10 μM dsRNA, and 50 mM DMTMM•Cl was added from a fresh solution to achieve a final concentration of 50 μM. It was then incubated in a thermomixer at 25°C for 16 h and then analysed by HPLC. Samples with liposome were injected directly to the HPLC without additional treatment.

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

The crudes of the coupling 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: unless otherwise stated, 0-40% or 0-80% of B in 45 min. Flow rate = 1 mL·min^{-1}. Injection: 100 µL (1 nmol). 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 oligos. 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 was expected that double strands and/or secondary structures were disrupted under the HPLC conditions used.

For results in section 2.3, relative yields of coupling reactions on liposome surface were calculated as follow:

$$Relative\ yield = \frac{[Peak\ area\ of\ RNA\ peptide\ hairpin\ in\ w/\ liposome]}{[Peak\ area\ of\ RNA\ peptide\ hairpin\ in\ w/o\ liposome]} \times 100\%$$

The isolated products were desalted with a Sep-Pak tC18 1 cc Vac Cartridge (Waters) and analysed by MALDI-TOF mass spectrometry.

Urea cleavage of RNA-peptide hairpins

The RNA-peptide (0.5 nmol) was added to an Eppendorf tube. Acetate buffer (pH4) or MES buffer (pH6), NaCl and water were added to the ON's solution and the reaction was heated in a Thermocycler at specified conditions. Concentration of the components in the reaction mixture: 10-50 µM of oligonucleotide, 100 mM of buffer and 100 mM of NaCl.

Continuous peptide synthesis cycle

The continuous peptide synthesis described in section 2.2.6 was performed with 15 nmol of **ON20** as starting mnm^{5}U strand. 15 nmol of m^6*aa*^{6}A strand **ON10-Gm** or **ON10-Vm** 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.

Kinetic study on peptide coupling between m^6*aa*^{6}A and mnm^{5}U

The peptide coupling reactions were carried out under identical conditions to those described above using DMTMM•Cl as activator. The kinetic 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:

dsRNA → RNA-peptide hairpin; k_{app}

The initial concentration of the double strand was refined as variable but constrained between 30 and 50 µM. 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.[265] 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.

Preparation of liposome suspension

Egg PC (Avanti Polar Lipids) was dissolved in chloroform to achieve a concentration of 20 mM. This solution was prepared as stock and stored in -20 ˚C. To produce liposome suspension, a required volume of the chloroform solution was taken, and a lipid film can be formed by vaporizing the chloroform in a round bottom flask under reduced pressure. Unless otherwise stated, the film was then rehydrated by a 2X reaction buffer and suspended completely in the solution through vortexing.

The liposome suspension was then pressed through a mini-extruder (Avanti Polar Lipids) equipped with a ø 100 nm Nuclepore™ Track-Etched membrane (Whatman) for 15 times and a small sample of the suspension was analysed by dynamic light scattering (DLS, Wyatt DyanPro NanoStar). For confocal microscopy, the extrusion and DLS steps were skipped. For flow cytometry and confocal microscopy, 1% and 0.1 % mol of Nile Red (Sigma-Aldrich) was added to stain the liposome respectively. The suspension was then diluted to a final concentration of 10 mM Egg PC, 100 mM NaCl and 100 mM MES buffer pH 6.0, plus the desired concentration of oligo as stated at each experiment.

Confocal microscopy of lipid binding oligos

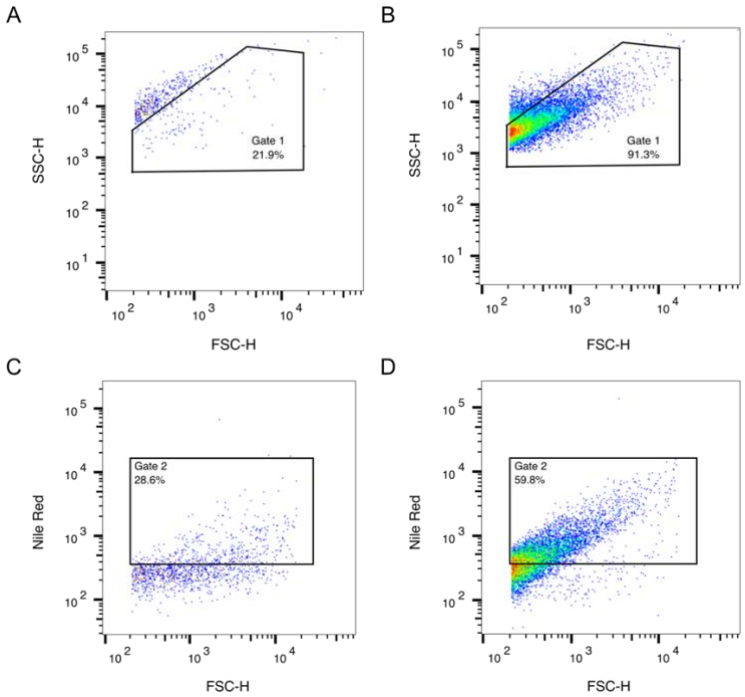
Samples were prepared with liposomes and 10 µM of FAM-labeled ONs. Without further incubation, 10 µL of the sample was then transferred to a sample slide.

The slides were then imaged on a Leica TCS SPE confocal microscope with an inverted microscopic stage. For the detection of Nile Red stained liposome, images were acquired with laser excitation of 532 nm and emission from 550-600 nm. For the detection of FAM-labelled ONs, images were acquired with laser excitation of 488 nm and emission from 500-550 nm.

Flow cytometry of lipid binding oligos

Extruded liposomes were prepared according to the method described above, with a final concentration of 2.5 µM of each of the RNAs in 400 µL. Samples with liposome, were then spun down for 3 h in 8 ˚C. The supernatant was carefully removed and 400 µL of fresh buffer (100 mM NaCl and 100 mM MES buffer pH 6.0) was used to resuspend the liposome. This washing process was repeated once before the cytometry measurement. Apart from samples containing RNA, controls of pure buffer, unstained liposome, negative control of stained liposome (marked as '-ve' below), and stained liposome + 0.1% Triton™ X-100 were prepared in a similar manner for gate selections.

Flow cytometry experiments had been performed on a LSR Fortessa™ cell analyzer (Becton Dickinson). Buffer only and Triton™ controls had been run for 5 min to ensure proper exclusion of background. In all samples at least 5000 events have been recorded. Nile Red fluorescence was measured on the 'PE-Cy5-5' setting and the fluorescein signal was measured on the 'FITC' setting provided by the instrument software. Gates 1 & 2 were created based on the background of the controls as shown below and population within Gate 2 were analyzed with FlowJo™ version 10.9.0 (Becton Dickinson).



Supplementary Fig. 1. *Flow cytometry data of (A) Buffer control, (B) Unstained liposome, (C) Nile Red stained liposome + 0.1% Triton™ X-100, and (D) Nile Red stained liposome (-ve). (A)&(B): Gate 1 was selected by excluding the background events caused by the buffer (100 mM NaCl and 100 mM MES buffer pH 6.0). (C)&(D): Within Gate 1, Gate 2 was created by excluding background Nile Red signal that is not caused by its localization in the lipid phase.*

Liposome-catalyzed geranylation of s^2U oligos

Vesicle solution was prepared by mentioned thin film rehydration with an aqueous solution containing NaCl (100 mM) and borate buffer (pH = 10.0, 100 mM) giving a final EggPC concentration of 40 mM. The oligos for the reaction (2 nmol each) were put into a 1.5 mL eppendorf tube and were dried in vacuo. Vesicle solution (100 µl) was added to the dried oligos and geBr (2.0 µmol, 0.397 µL) was added to the mixtures followed by vortexing. The reaction mixtures were incubated at 1000 rpm at r.t. over a period of 2 h. It was then injected to HPLC, and peaks were collected and analysed by MALDI-TOF mass spectrometry.

Stability study of ges^2U hydrolysis in the presence or absence of liposome

Liposome solution was prepared as mentioned above containing NaCl (200 mM) and acetate (pH 7, 100 mM) or glycine buffer (pH 10, 100 mM), giving a final Egg PC concentration of 20 mM. The oligos for the reaction (5.5 nmol each) were put into a 1.5 mL eppendorf tube and dried in vacuo. Vesicle solution (275 µl) was added to the dried ONs followed by addition of rehydration buffer (275 µl), to achieve a final concentration of 100 mM NaCl, 50 mM buffer and 10 mM Egg PC. For the -Lip setup, Egg PC addition was excluded. The samples were vortexed and shaken with a Thermomixer (1000 rpm) at r.t. for 4 days. After 0, 24, 48, 72 and 96 h an aliquot of 100 µl were removed from the samples and analysed by HPLC.

Synthesis of modified triplet triphosphate on oligo synthesizer.

The *cyclo*Sal phosphoramdite (**41**) and tetrabutylpyrophosphate DMF solution were synthesized and prepared according to literature.[234,235] The *cyclo*Sal phosphoramidite was dissolved in anhydrous acetonitrile with a concentration of 30 mg mL^{-1} and attached to one of the phosphoramidite ports of the synthesiser. The synthesis was carried out first by standard 1.0 µmol RNA synthesis method with the corresponding RNA phosphoramidites on CPG solid support and ended on the coupling of the *cyclo*Sal-chlorophosphoramidite with a special coupling setting of 600 sec coupling time, 60 sec oxidation time and with capping and DMT deprotection steps skipped. The cartridge was then immediately reacted with 1 mL 0.5 M tetrabutylpyrophosphate DMF solution by syringe pushing overnight. The beads were then washed with 5 mL of DMF, 10 mL of deionized water and 5 mL of acetonitrile, and cleaved with 1 mL of 30% NH_4OH and 40% $MeNH_2$ for 2 h at r.t. The solution was then dried by speed-vacuum and the 2'TBS deprotection was carried out with 1 mL of 1 M tetra-*n*-butylammonium fluoride in THF at r.t. overnight. The crude product was precipitated in 0.3 M sodium acetate (pH 5.5) and 80% cold ethanol. The pellet was re-dissolved in water and purified by HPLC.

In vitro transcription of s^2U template RNAs and trinucleotide triphosphates

Prior to transcription, the respective DNA template (tx) strand was mixed with the 5T7 promoter strand (1 µM each) and underwent 2 PCR cycles (94 ˚C, 2mins.; [94 ˚C, 30 s; 50 ˚C, 30 sec; 72 ˚C, 1 min.] x2; 72 ˚C, 2 mins.) with GoTaq® Hotstart Readymix (Promega) and purified with QiaQuick nucleotide removal kit (Qiagen) to obtain the desired dsDNA template.

Transcriptions for template RNAs for ribozyme-catalyzed primer extension were performed on ca. 15 ng/µL dsDNA using MegaShortScript enzyme and buffer (Thermo Fisher) according to the manufacturer's protocol with 4 mM of NTP mixture (for s^2U oligos, UTP was replaced by s^2UTP) and additionally 20 mM GMP to form 5'monophosphate on the template strand to reduce aberrant ligation by the ribozyme. The reaction mixture was incubated in 37 ˚C overnight. After that, the DNA template was digested by TURBO™ DNase (0.2 U/µL, Thermo Fisher) in 37 ˚C for 30 mins. The reaction mixture was then purified by 10 % denaturing PAGE and the desired bands were excised under UV shadowing. The gel pieces were crushed and suspended in 400 µL of 10 mM Tris-HCl buffer (pH7.4). The suspension was slowly frozen in dry ice bath and quickly thawed in 50 ˚C. The suspension was then shaken at r.t. overnight. Next, the gel pieces were filtered and discarded. To the solution, 10% volume of 3 M sodium acetate (pH5.5) was added and diluted with absolute ethanol up to 2 mL. The solution was then centrifuged for 1 h to remove the ethanol and obtain the RNA pellets at the bottom of the Eppendorf tube.

The transcription of triplet triphosphates was described in detail in a previous paper from the Holliger's group.[183] Briefly, a solution of the three required NTPs (CTP, GTP and s^2UTP. 2.4 mM each), **5T7 promoter** (2 µM), and triplet template strand **(tx)CGU_triplet** (2 µM) were transcribed with the MegaShortScript enzyme and buffer (Thermo Fisher). The **(tx)CGU_triplet** strand encoded for ACGU instead of CGU, as the transcriptase tends to skip the first incorporation in extremely short template. The problem was avoided by introducing a spacer base where its respective triphosphate is absent in the reaction mixture, in this case A and ATP.[183] The DNA template was then digested by TURBO™ DNase (0.1 U/µL) in at 37 ˚C for 30 mins. The crudes were then purified by 30% denaturing PAGE and extracted from their excised gel pieces as described above.

Primer extension experiments of 5TU+t1.5 ribozyme

A 5 µL solution of ribozyme 5TU+t1.5 (0.5 µM, provided by the Holliger's group), the respective NNNTP (5 µM), template RNA strand (0.5 µM), Cy3 and biotin dual labeled primer **P10** and Tween® 20 (0.05%) was mixed and denatured in 80 ˚C for 2 mins., annealed in 17 ˚C for 10 mins. and cooled down to 4 ˚C. Next, $MgCl_2$ and Tris-HCl (pH8.3) were added to the reaction mixture to reach respective final concentrations of 200 mM and 50 mM, and a final reaction volume of 10 µL. The reaction was frozen in dried ice and incubated in a -7 ˚C glycerol bath overnight.

The reaction was quenched by adding 1X volume of 0.5 M EDTA solution. To isolate the 5' Cy3 and biotin dual labeled extended primers, 5 µL/reaction of MyOne™ Dynabeads™ Streptavidine C1 (Invitrogen) was washed with 10X volume of BWBT buffer thrice. The beads were then suspended in 40X volume of BWBT buffer (200 µL/reaction) and added to the reaction mixture. The mixture was then mixed at r.t. for 30 mins., and then washed with equal volume of BWBT buffer once, NaBET buffer twice, and then BWBT buffer twice. The beads were then resuspended with 40 µL of 25 mM EDTA in 95% formamide. The mixture was heated at 95 ˚C for 5 mins., before being separated by a 20% denaturing PAGE and imaged on a Typhoon FLA-9000 imager (GE Healthcare, US) with appropriate wavelength.

BWBT buffer: 200 mM NaCl, 10 mM Tris-HCl (pH7.4), 1 mM EDTA, 0.1% Tween® 20
NaBET buffer: 50 mM NaOH, 1 mM EDTA, 0.05% Tween® 20

Sequencing of primer-extension fidelity test

The ribozyme primer-extension was carried out as described above, except that an extended dual labeled primer, **kyle3-P10**, was used instead. After the denaturing PAGE analysis, the gel was wrapped around with a piece of transparent wrapping film. An ink-marker was used to draw marks in the blank area on the wrapping film and the wrapped gel was once again scanned by the phosphorimager. The scanned image was printed out with actual size and the gel was placed on top of the paper with its position aligned by the ink-marks created earlier. The bands of each reaction was then excised from the wrapped gel according to the aligned positions of the printed image. The wrapping film was then carefully removed from the excised bands, which were crushed in respective Eppendorf tubes.

MyOne™ DynaBeads™ was washed and resuspended in BWBT buffer as described above. To each gel piece sample, 200 µL of bead suspension was added and they were mixed in 4 ˚C overnight. The gel piece was then separated from the bead suspension by a Sysmex CellTrics® 50 µm filter (Sysmex) and washed with 400 µL BWBT buffer for 4 times. Together with the washings, the beads were isolated on a magnetic rack and washed with 400 µL of BWBT buffer once, NaBET buffer once, and then BWBT buffer twice.

For the library preparation of each reaction, adaptor **HDVlig** was 5' adenylated with 5' DNA adenylation kit (NEB) in 65 ˚C for 2 h. The product was then extracted from neutral phenol/chloroform and precipitated in 72% ethanol. Next, a 20 µL adaptor ligation reaction mixture consisted of PEG_{8000} (15%), Tween® 20 (0.04%), 5' adenylated **HDVlig** (2 µM), T4 RNA ligation buffer and T4 RNA ligase 2 truncated KQ (NEB) was added to each bead sample. The ligation was mixed at r.t. for 1 h and then 10 ˚C overnight.

The ligation was quenched by adding 5 µL of 0.5 M EDTA solution and 180 µL of BWBT buffer to each bead suspension sample. The beads were then washed with 20 µL BWBT buffer once and resuspended in 22.5 µL water. To this suspension, a 5 µL reaction mixture of forward primer **kyleF** (5 µM), reverse primer **HDVrec** (5 µM) and reaction buffer and enzymes from the SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* DNA polymerase

reaction kit (Thermo Fisher). The mixture was then proceeded through 24 cycles of PCR (50 ˚C, 60 mins ; 94 ˚C, 2 mins; [94˚C, 15 sec; 50 ˚C, 30 sec; 68 ˚C, 30 sec] x 24; 68 ˚C, 5 mins; 15 ˚C, forever). Afterwards, the quality of the PCR product was checked by a 4 % agarose gel electrophoresis. The PCR product was then purified by PCR purification kit (Qiagen).

To ligate the adaptors and barcodes, primers **Pxx_kyleF** and **p513HDVba** (1 µM each) were added to 5 µL of the purified dsDNA, and the samples were proceeded through 15 cycles of PCR with GoTaq® Hotstart Readymix (Promega) with conditions described above. The barcoded PCR samples from different experiment setups were mixed and 30 µL of it was purified by 4 % agarose gel. The correct bands were excised, extracted with the QIAquick® Gel extraction kit (Qiagen), and quantified with Qubit fluorometer (Thermo Fisher). Sequencing was performed on Illumina MiSeq™ platform with a v2 reagent NanoKit-300 cycles (Illumina). A total of 32 fmol of dsDNA with 15 mol % of ϕ_x reference sequence (ca. 4.8 fmol) containing the fidelity experiments and other barcoded samples were sequenced according to manufacturer's protocol.

Binding study of (AbPAb)^{6}A RNA-peptide

The RNA-peptide was produced by a peptide coupling reaction between **ON38-AbPAb^{6}A** and **ON39-nm^{5}U-geS** as described above. After that, 100 µL samples containing 10 mM Egg PC liposome, 100 mM MES buffer (pH 6), 100 mM NaCl, 20 µM RNA-peptide and 20 µM (2 pNPP : 2 RNA-peptide) or 100 µM (10 pNPP : 2 RNA-peptide) pNPP were incubated at r.t. for 30 mins. Negative (2 RNA-peptide only) and positive (2 pNPP only) controls contained only 20 µM of RNA-peptide or pNPP, respectively, with the liposomes and buffer. The controls and samples were then analyzed by HPLC with setting described above with a 0 → 80% B gradient.

RNA-peptide catalysis study with pPE^{5}U modification

The on-bead coupling method was employed on the solid supports of **ON37-nm^{5}U** with commercially available Boc-protected amino acid building blocks to obtain **ON37-pPE^{5}U**. To set up the reactions, 5 µL reaction mixtures containing 5 mM of ssRNA **ON37-pPE^{5}U** or dsRNA **ON37-pPE^{5}U:ON30-geS**, *trans*-ß-nitrostyrene (500 mM), butanal (750 mM), *N*-methylmorpholine (20 mM) and Egg PC liposome (10 mM) were incubated at r.t. for 16 h. Negative control contained only the starting materials and liposome. For the positive control, 5 mM of the reported tripeptide catalyst pPE-NH_2 (provided by the group of Prof. Dr. Ivan Huc from solid phase peptide synthesis) was reacted with *trans*-ß-nitrostyrene (500 mM), butanal (750 mM) and *N*-methylmorpholine (20 mM) in a solvent system of 10 % *i*PrOH in chloroform overnight. For the 'pPENH_2' setup, the same reaction was repeated in water instead of organic solvent. After that, the samples were diluted with 95 µL of 50 % acetonitrile in water and injected into a *Dionex UltiMate 3000* micro UHPLC-System (pump, auto sampler, column compartment and diode array detector) with RP-column chromatography methods using a *Hypersil GoldC18 selectivity* column (100 × 2.1mm) with a gradient of 0-64% acetonitrile with 0.01% formic acid.

4.2 Sequence information

Name	Sequence Information			Polymer	HPLC			MALDI-TOF (m/z)		Section
	5' Mod.	Sequence (5'-3')	3' Mod.		0-30%B t_R (min)	0-40%B t_R (min)	0-80%B t_R (min)	calcd. for $[M-H]^-$	found	
ON1	-	GUC(t^6A)ACCUGA	-	RNA	26.5	-	-	2956.4	2955.6	2.1
ODN1	-	GTC(t^6dA)ACCTGA	-	DNA	25.1	-	-	2841.5	2841.1	2.1
ON2	-	GUC(g^6A)ACCUGA	-	RNA	25.2	-	-	2913.4	2912.8	2.1
ON3	-	GUC(v^6A)ACCUGA	-	RNA	22.5	-	-	2955.5	2954.9	2.1
ON4	-	GUC(h^6A)ACCUGA	-	RNA	23.7	-	-	2993.4	2992.9	2.1
ON5	-	GUC(d^6A)ACCUGA	-	RNA	21.0	-	-	2971.4	2970.9	2.1
ON6	-	GUC(f^6A)ACCUGA	-	RNA	29.8	-	-	3003.5	3002.6	2.1
ON7	-	GUC((*D*-Phe)^{6}A)ACCUGA	-	RNA	31.2	-	-	3003.5	3002.4	2.1
ON8	-	AUCG(t^6A)CUACG(t^6A)AUCGC(t^6A)ACCG	-	RNA	31.5	-	-	7109.1	7107.8	2.1
ON9	-	AGAUGUG(s^6A)(d^6A)(h^6A)GAGAUGA	-	RNA	25.3	-	-	6042.9	6041.7	2.1
ON10-Gm	-	(m^6g^6A)AUCGCU	-	RNA	-	18.8	-	2277.4	2278.4	2.2
ON10-Am	-	(m^6a^6A)AUCGCU	-	RNA	-	20.2	-	2291.4	2290.0	2.2
ON10-Vm	-	(m^6v^6A)AUCGCU	-	RNA	-	22.2	-	2319.4	2317.8	2.2
ON10-Lm	-	(m^6l^6A)AUCGCU	-	RNA	-	24.3	-	2333.4	2331.6	2.2
ON10-Tm	-	(m^6t^6A)AUCGCU	-	RNA	-	18.9	-	2321.4	2320.0	2.2
ON10-Pm	-	(m^6p^6A)AUCGCU	-	RNA	-	18.0	-	2317.4	2316.8	2.2
ON10-Fm	-	(m^6f^6A)AUCGCU	-	RNA	-	24.5	-	2368.4	2365.4	2.2
ON10-Mm	-	(m^6m^6A)AUCGCU	-	RNA	-	23.2	-	2351.4	2350.4	2.2
ON10-Dm	-	(m^6d^6A)AUCGCU	-	RNA	-	17.2	-	2335.4	2334.3	2.2
ON10-Gcnm	-	(m^6gcn^6A)AUCGCU	-	RNA	-	21.2	-	2258.4	2258.5	2.2

ON10-V	-	(v^6A)AUCGCU	-	RNA	-	20.6	-	2305.4	2302.2	2.2
ON11	-	(m^6v^6A)CUAUUGAGU	-	RNA	-	22.3	-	3300.5	3301.1	2.2
ON12	-	(m^6v^6A)AUCGCUGUACCCUAUUGAGU(m^6g^6A)	-	RNA	-	23.1	-	7231.0	7233.7	2.2
ON13	-	(m^6v^6A)AUCGCUGUAC	-	RNA	-	23.2	-	3604.6	3603.4	2.2
ON14	-	(m^6g^6Am)AUCGCU	-	2'OMe RNA	-	23.8	-	2375.5	2374.4	2.2
ON15	-	(m^6g^6Am)AUCG	-	2'OMe RNA	-	23.6	-	1736.4	1735.1	2.2
ON16	-	(m^6g^6Am)AU	-	2'OMe RNA	-	23.1	-	1058.2	1058.2	2.2
ON17-m	-	GUACAGCGAU(mnm^5U)	-	RNA	-	17.4	-	3530.5	3529.7	2.2
ON17	-	GUACAGCGAU(nm^5U)	-	RNA	-	17.8	-	3516.5	3515.9	2.2
ON17-Vm	-	GUACAGCGAU($vmnm^5$U)	-	RNA	-	18.6	-	3629.6	3627.2	2.2
ON17-teoc	-	GUACAGCGAU(teoc-$vmnm^5$U)	-	RNA	-	37.7	-	3773.7	3776.9	2.2
ON18	-	GUACAGCGAU($gmnm^5$U)ACUCAAUAG(nm^5U)	-	RNA	-	18.7	-	6806.0	6806.4	2.2
ON19	-	GUACAGCGAU($vmnm^5$U)ACUCAAUAGG	-	RNA	-	19.9	-	6858.0	6857.7	2.2
ON20	-	GUACAGCGAU(mnm^5U)	-	2'OMe RNA	-	23.0	-	3670.5	3670.4	2.2
ON21	-	(m^6g^6A)AGCCCU	-	RNA	-	19.5	-	2276.4	2275.7	2.2
CON1	-	AAUCGCU	-	RNA	-	23.6	-	2162.3	2162.0	2.2
CON2	-	GUACAGCGA	-	RNA	-	23.1	-	3487.5	3486.9	2.2
CON3	-	GUACAGCGAUUAAUCGCU	-	RNA	-	23.9	-	5712.8	5711.7	2.2
CON4	-	AAUCGCU	-	2'OMe RNA	-	23.3	-	2261.6	2260.1	2.2
CON5	-	GUACAGCGA	-	2'OMe RNA	-	18.8	-	4772.7	4772.8	2.2
CON6	-	GUACAGCGAUUAAUCGCU	-	2'OMe RNA	-	18.6	-	6998.0	6995.1	2.2
ON22-geS	-	UUUUU(ges^2U)	-	RNA	-	40.4	-	1925.3	1924.7	2.3
ON22-5FAM-S	6FAM	UUUUU(s^2U)	-	RNA	-	26.1	13.7	2326.3	2325.5	2.3

ON22-5FAM-geS	6FAM	UUUUU(ges^{2}U)	-	RNA	-	-	22.8	2462.4	2463.7	2.3
ON23-S	-	UUUU(s^{2}U)(s^{2}U)	-	RNA	-	18.4	11.6	1805.1	1805.1	2.3
ON23-geS	-	UUUU(ges^{2}U)(ges^{2}U)	-	RNA	-	-	29.3	2077.4	2076.7	2.3
ON23-5FAM-S	6FAM	UUUU(s^{2}U)(s^{2}U)	-	RNA	-	23.1	13.5	2342.3	2344.2	2.3
ON23-5FAM-geS	6FAM	UUUU(ges^{2}U)(ges^{2}U)	-	RNA	-	-	30.7	2614.5	2616.1	2.3
ON24-5FAM-S	6FAM	UU(s^{2}U)UU(s^{2}U)	-	RNA	-	24.0	14.2	2342.3	2343.2	2.3
ON24-5FAM-geS	6FAM	UU(ges^{2}U)UU(ges^{2}U)	-	RNA	-	-	27.7	2614.5	2616.0	2.3
ON25-5FAM-S	6FAM	GCGA(s^{2}U)(s^{2}U)	-	RNA	-	24.3	14.2	2442.3	2443.1	2.3
ON25-5FAM-geS	6FAM	GCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	-	30.5	2714.6	2715.3	2.3
ON26-S	-	CAGCGA(s^{2}U)(s^{2}U)	-	RNA	-	18.1	11.1	2539.3	2540.4	2.3
ON26-geS	-	CAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	-	29.3	2811.6	2812.4	2.3
ON26-5FAM-S	6FAM	CAGCGA(s^{2}U)(s^{2}U)	-	RNA	-	25.7	15.2	3076.4	3078.9	2.3
ON26-5FAM-geS	6FAM	CAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	-	30.8	3348.7	3350.8	2.3
ON27-5FAM-S	6FAM	GUACAGCGA(s^{2}U)(s^{2}U)	-	RNA	-	23.9	13.8	4056.6	4059.0	2.3
ON27-5FAM-geS	6FAM	GUACAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	-	29.3	4328.8	4331.5	2.3
ON28-5FAM-S	6FAM	AUCGUACAGCGA(s^{2}U)(s^{2}U)	-	RNA	-	25.2	14.6	4996.7	4998.4	2.3
ON28-5FAM-geS	6FAM	AUCGUACAGCGA(ges^{2}U)(ges^{2}U)	-	RNA	-	-	29.3	5268.9	5275.5	2.3
ON29-3FAM	-	UCGCUG	6FAM	RNA	-	Ordered				2.3
ON30-S	-	CAGCGA(s^{2}U)(s^{2}U)(s^{2}U)	-	RNA	-	20.6	11.9	2861.3	2862.3	2.3
ON30-geS	-	CAGCGA(ges^{2}U)(ges^{2}U)(ges^{2}U)	-	RNA	-	-	35.1	3269.7	3271.4	2.3

ON31-3FAM	-	ACGGU	6FAM	RNA	-	Ordered				2.3
ON32-S	-	(s^2U)(s^2U)UCGCUG	-	RNA	-	19.4	11.7	2493.3	2493.6	2.3
ON32-geS	-	(ges^{2}U)(ges^{2}U)UCGCUG	-	RNA	-	-	28.9	2765.5	2765.6	2.3
ON33-S	-	(s^2U)(s^2U)UUGCCG	-	RNA	-	19.2	11.8	2493.3	2493.7	2.3
ON33-geS	-	(ges^{2}U)(ges^{2}U)UUGCCG	-	RNA	-	-	28.5	2765.5	2765.5	2.3
ON34-S	-	(s^2U)CGCUG	-	RNA	-	17.6	10.9	1865.2	1865.5	2.3
ON34-geS	-	(ges^{2}U)CGCUG	-	RNA	-	-	21.9	2001.4	2001.4	2.3
ON35-S	-	(s^2U)(s^2U)CGCUG	-	RNA	-	19.3	11.8	2187.2	2187.6	2.3
ON35-geS	-	(ges^{2}U)(s^2U)CGCUG	-	RNA	-	-	22.5	2459.5	2459.5	2.3
ON36-g^{6}A-S	-	(g^{6}A)AGCGA(s^2U)(s^2U)(s^2U)	-	RNA	-	20.0	12.1	2986.3	2987.6	2.3
ON36-g^{6}A-geS	-	(g6A)AGCGA(ges^{2}U)(ges^{2}U)(ges^{2}U)	-	RNA	-	-	34.5	3394.7	3395.1	2.3
ON37-nm^{5}U	-	UCGCU(nm^{5}U)	-	RNA	-	13.9	9.2	1839.3	1839.3	2.3
ON38-g^{6}A	-	(g^{6}A)AGCGA	-	RNA	-	18.1	11.2	2020.3	2020.1	2.3
ON38-2g^{6}A	-	(2g^{6}A)AGCGA	-	RNA	-	18.6	11.5	2077.4	2077.4	2.3
ON38-3g^{6}A	-	(3g^{6}A)AGCGA	-	RNA	-	18.7	11.5	2134.4	2134.2	2.3
ON38-4g^{6}A	-	(4g^{6}A)AGCGA	-	RNA	-	18.6	11.4	2191.4	2191.8	2.3
ON38-a^{6}A	-	(a^{6}A)AGCGA	-	RNA	-	19.1	11.6	2034.4	2034.6	2.3
ON38-2a^{6}A	-	(2a^{6}A)AGCGA	-	RNA	-	18.8	11.7	2105.4	2105.5	2.3
ON38-3a^{6}A	-	(3a^{6}A)AGCGA	-	RNA	-	19.0	11.8	2176.4	2176.6	2.3
ON38-4a^{6}A	-	(4a^{6}A)AGCGA	-	RNA	-	19.2	12.0	2247.5	2247.5	2.3
ON38-v^{6}A	-	(v^{6}A)AGCGA	-	RNA	-	22.1	13.2	2062.6	2062.4	2.3
ON38-2v^{6}A	-	(2v^{6}A)AGCGA	-	RNA	-	24.2	14.2	2161.4	2161.5	2.3
ON38-3v^{6}A	-	(3v^{6}A)AGCGA	-	RNA	-	25.7	15.0	2260.5	2260.5	2.3
ON38-4v^{6}A	-	(4v^{6}A)AGCGA	-	RNA	-	27.9	16.2	2359.6	2359.6	2.3
ON38-l^{6}A	-	(l^{6}A)AGCGA	-	RNA	-	24.1	14.3	2076.4	2077.6	2.3

ON38-2l^{6}A	-	(2l^{6}A)AGCGA	-	RNA	-	28.5	16.6	2189.5	2189.4	2.3
ON38-3l^{6}A	-	(3l^{6}A)AGCGA	-	RNA	-	32.6	18.8	2302.6	2302.5	2.3
ON38-4l^{6}A	-	(4l^{6}A)AGCGA	-	RNA	-	37.3	21.3	2415.7	2415.6	2.3
ON38-f^{6}A	-	(f^{6}A)AGCGA	-	RNA	-	24.5	14.5	2110.4	2111.7	2.3
ON38-2f^{6}A	-	(2f^{6}A)AGCGA	-	RNA	-	31.0	17.8	2257.5	2257.8	2.3
ON38-3f^{6}A	-	(3f^{6}A)AGCGA	-	RNA	-	36.3	20.7	2404.5	2405.2	2.3
ON38-4f^{6}A	-	(4f^{6}A)AGCGA	-	RNA	-	42.3	23.9	2551.6	2551.8	2.3
ON39-teocnm^{5}U-S	-	(s^{2}U)(s^{2}U)(s^{2}U)UCGCU(teocnm^{5}U)	-	RNA	-	32.9	19.2	2949.3	2950.4	2.3
ON39-teocnm^{5}U-geS	-	(ges^{2}U)(ges^{2}U)(ges^{2}U)UCGCU(teocnm^{5}U)	-	RNA	-	-	34.3	3357.7	3359.5	2.3
ON39-nm^{5}U-geS	-	(ges^{2}U)(ges^{2}U)(ges^{2}U)UCGCU(nm^{5}U)	-	RNA	-	-	32.4	3213.7	3214.3	2.3
CON7	-	UCGCUUAAGCGA	-	RNA	Ordered					2.3
CON8	-	UCGCUUAAGCGAUUU	-	RNA						2.3
CON9	-	UUUUCGCUUAAGCGA	-	RNA						2.3
CON10	-	UUUUCGCUUAAGCGAUUU	-	RNA						2.3
P10	Cy3, Biotin	CUGCCAACCG	-	RNA	Ordered					2.4
tP10CAG$_{3}$	-	CUGCUGCUGCGGUUGGCAG	-	RNA						2.4
tP10AGG$_{3}$	-	CCUCCUCCUCGGUUGGCAG	-	RNA						2.4
tP10CUC$_{3}$	-	GAGGAGGAGCGGUUGGCAG	-	RNA						2.4
tP10CUA$_{3}$	-	UAGUAGUAGCGGUUGGCAG	-	RNA						2.4
tP10UCG$_{3}$	-	CGACGACGACGGUUGGCAG	-	RNA						2.4
tP10CGU$_{11}$	-	**(ACG)$_{11}$**CGGUUGGCAG	-	RNA						2.4
tP10UCGc3flk	-	GGGCGAGGGCGGUUGGCAG	-	RNA						2.4

tP10CUAc3flk	-	GGGUAGGGGCGGUUGGCAG	-	RNA		2.4
tP10CGUc3flk	-	GGGACGGGGCGGUUGGCAG	-	RNA		2.4
tP10AGC3	-	GCUGCUGCUCGGUUGGCAG	-	RNA		2.4
tP10GAC3	-	GUCGUCGUCCGGUUGGCAG	-	RNA		2.4
tP10AGCc3flk	-	GGGGCUGGGCGGUUGGCAG	-	RNA		2.4
tP10GACc3flk	-	GGGGUCGGGCGGUUGGCAG	-	RNA		2.4
tP10CGAc3flk	-	GGGUCGGGGCGGUUGGCAG	-	RNA		2.4
tP10AGC3_s^2U	-	GC(s^2U)GC(s^2U)GC(s^2U)CGG(s^2U)(s^2U)GGCAG	-	RNA	Transcribed from template strands (tx)	2.4
tP10GAC3_s^2U	-	G(s^2U)CG(s^2U)CG(s^2U)CCGG(s^2U)(s^2U)GGCAG	-	RNA		2.4
tP10AGCc3flk_s^2U	-	GGGGC(s^2U)GGGCGG(s^2U)(s^2U)GGCAG	-	RNA		2.4
tP10GACc3flk_s^2U	-	GGGG(s2U)CGGGCGG(s2U)(s2U)GGCAG	-	RNA		2.4
tP10CGAc3flk_s^2U	-	GGG(s^2U)CGGGGCGG(s^2U)(s^2U)GGCAG	-	RNA		2.4
(tx)tP10AGC_3	-	CTGCCAACCGAGCAGCAGCCTATAGTGAGT CGTATTAATTTCGCGGGCGAGATCGATC	-	DNA	Ordered	2.4
(tx)tP10GAC_3	-	CTGCCAACCGGACGACGACCTATAGTGAGT CGTATTAATTTCGCGGGCGAGATCGATC	-	DNA		2.4
(tx)tP10AGCc3flk	-	CTGCCAACCGCCCAGCCCCTATAGTGAGTC GTATTAATTTCGCGGGCGAGATCGATC	-	DNA		2.4
(tx)tP10GACc3flk	-	CTGCCAACCGCCCGACCCCTATAGTGAGTC GTATTAATTTCGCGGGCGAGATCGATC	-	DNA		2.4
(tx)tP10CGAc3flk	-	CTGCCAACCGCCCCGACCCTATAGTGAGTC GTATTAATTTCGCGGGCGAGATCGATC	-	DNA		2.4

5T7 promoter	-	GATCGATCTCGCCCGCGAAATTAATACGACTCACTATAG	-	DNA						2.4
(tx)CGU_triplet	-	ACGUTATAGTGAGTCGTATTAATTTCGCGGGCGAGATCGATC	-	DNA						2.4
HDVlig	-	GGGTCGGCATGGCATC	C3 spacer	DNA						2.4
HDVrec	-	GATGCCATGCCGACCC	-	DNA						2.4
kyleF_P10	Cy3, Biotin	d(GGATTCACTGCGATAGAGT)r(CCUGCCAACCG	-	DNA-RNA chimera						2.4
kyleF	-	GGATTCACTGCGATAGAGTC	-	DNA						2.4
Pxx_kyleF	-	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC**NNN**ATCACGGGATTCACTGCGATAGAGTC	-	DNA						2.4
p513HDVba	-	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**NNN**AGTCAAGATGCCATGCCGACCC	-	DNA						2.4
ON38-AbPAb^{6}A	-	(AbProAb^{6}A)AGCGA	-	RNA	-	28.0	-	2296.5	2298.6	2.5
ON37-pPE	-	UCGCU(pPE-nm^{5}U)	-	RNA	-	15.8	-	2162.4	2162.7	2.5

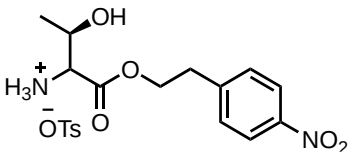
Supplementary Table 1. Sequence information and characterization data of all oligonucleotides described in this thesis.

4.3 Synthetic information and characterization data

Npe-protection of the carboxy group of amino acids

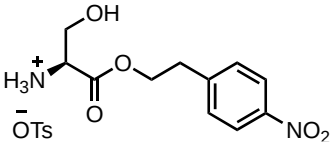
The reaction was performed according to literature.[210] Briefly, L-amino acid (1 eq.), 2-(4-nitrophenyl)ethanol (npe-OH, 3 eq.) and TsOH (3 eq.) were refluxed in toluene overnight in a Dean-Stark apparatus. The solution was cooled to room temperature and Et_2O was added. The oily residue was decanted, 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



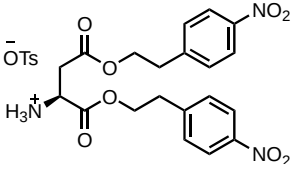
Yield: 70%;
IR: ṽ = 3401 (w), 2930 (s), 2892 (s), 2858 (s), 1730 (s), 1510 (vs), 1465 (s), 1300 (vs), 1258 (s), 1167 (s), 1010 (s), 895 (w), 832 (s) cm^{-1}
^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.2 – 8.1 (m, 5H), 7.58 (d, *J* = 8.8 Hz, 2H), 7.48 (d, *J* = 8.2 Hz, 2H), 7.11 (d, J = 8.2 Hz, 2H), 5.6 (br. s., 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-d6) 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 $[M+H]^+$.

Compound 5



Yield: 75%
IR: ṽ = 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

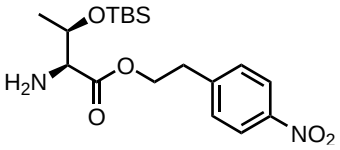


Yield: 82 %
IR: ṽ = 3402 (w), 2933 (s), 2894 (s), 2858 (s), 1730 (s), 1689 (s), 1514 (vs), 1469 (s), 1310 (vs), 1258 (s), 1167 (s), 1010 (s), 895 (w), 834 (s) cm^{-1}
^{1}H NMR (400 MHz, DMSOd$_6$) δ: 8.42 (s, 3H), 8.16 – 8.11 (m, 4H), 7.55 – 7.48 (m, 6H), 7.12 (d, *J* = 7.9 Hz, 2H), 4.37 (t, *J* = 6.3 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 $[M+H]^+$.

Protection of amino acid hydroxyl side chains with TBSCl

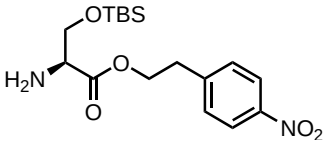
Npe-protected amin acid (1 eq.) was dissolved in pyridine and stirred with TBSCl (3 eq.) and 1H-imidazole (3 eq.) at r.t. overnight. The mixture was diluted with CH2Cl2 and washed successively with sat. $NaHCO_3$ solution and H_2O. The organic layer was dried, evaporated and purified by flash chromatography eluting with CH_2Cl_2/MeOH (10/1, *v/v*) to afford the target compound as an oil.

Compound 7



Yield: 94%
IR: ṽ = 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 $[M+H]^+$.

Compound 8



Yield: 96%
IR: ṽ = 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, CDCl3) 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, CDCl3) 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 $C17H29N2O5Si^+$: m/z = 369.1846 $[M+H]^+$; found: m/z = 369.1842 $[M+H]^+$.

Protection of carboxy group of Boc-protected amino acids

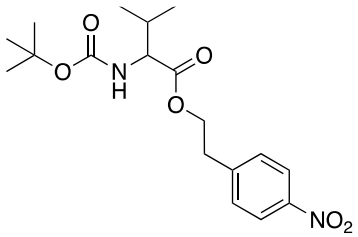
Boc-amino acid (1 eq.) was dissolved in CH_2Cl_2 under inert atmosphere and cooled to 0 ˚C. Then npeOH (1.3 eq.) and PPh_3 (1.3 eq.) were added followed by slow addition of DIAD (1.3 eq.). The reaction mixture was left to stir for 2 h at room temperature. Then the solution was washed with water, organic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

Deprotection of of Boc-protecting group

Npe-protected amino acid was dissolved in 4 M HCl/Dioxane mixture at 0 ˚C. The reaction mixture was left to stir for 2 h and afterwards was evaporated to dryness. Excessive Et_2O was added to precipitate the salt, which was then filtered, washed and dried under reduced atmosphere.

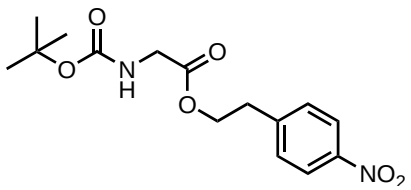
Later in the study, the protocol was modified to skip the column purification of the Boc- and npe-protected amino acid in the previous step. The reaction crude after the npe-protection was dissolved directly in 4 M HCl/Dioxane at 0 ˚C and precipitated afterwards.

Compound 12



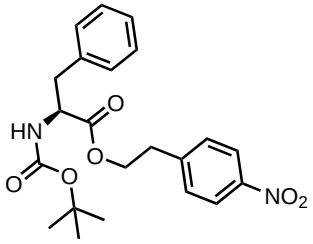
Eluent: Hex/EtOAc (4/1, v/v).
Yield: 96%
IR: ṽ = 3397 (w), 2975 (w), 1751 (s), 1709 (vs), 1519 (vs), 1391 (w), 1366 (w), 1345 (vs), 1159 (vs), 1056 (w), 905 (vs), 723 (vs) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.15 (d, *J* = 8.6 Hz, 2H), 7.38 (d, *J* = 8.6 Hz, 2H), 4.95 (d, *J* = 8.7 Hz, 1H), 4.38 (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



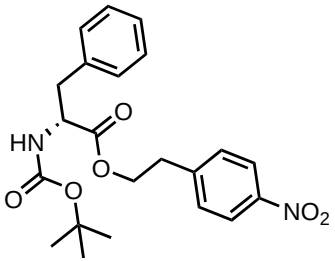
Eluent: Hex/EtOAc (3/1, v/v).
Yield: 85%
IR: ṽ = 3396 (w), 2978 (w), 2254 (w), 1750 (s), 1707 (vs), 1601 (w), 1518 (vs), 1391 (w), 1366 (w), 1345 (vs), 1250 (w), 1159 (vs), 1056 (w), 905 (vs), 727 (vs) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.10 (d, *J* = 8.6 Hz, 2H), 7.38 (d, *J* = 8.6 Hz, 2H), 4.97 (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 $[M+H]^+$.

Compound 14a



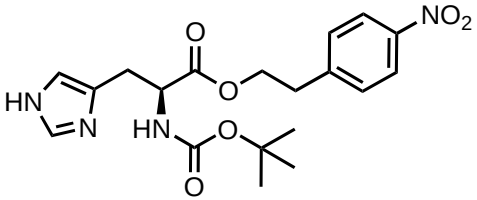
Eluent: Hex/EtOAc (4/1, v/v).
Yield: 81%
IR: ṽ = 3426 (w), 3356 (w), 1740 (w), 1709 (vs), 1602 (w), 1518 (vs), 1495 (s), 1344 (vs), 1249 (w), 1159 (vs), 1056 (w), 855 (s), 733 (s), 698 (vs) cm^{-1}
$_1$H NMR (400 MHz, $CDCl_3$) δ: 8.15 (d, *J* = 8.7 Hz, 2H), 7.31 (d, *J* = 8.7 Hz, 2H), 7.25 – 7.20 (m, 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



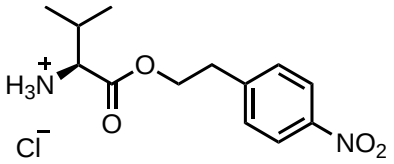
Eluent: Hex/EtOAc (4/1, v/v).
Yield: 85%
IR: ṽ = 3443 (w), 3375 (w), 2977 (w), 2929 (w), 1740 (w), 1709 (vs), 1602 (w), 1517 (vs), 1495 (s), 1343 (vs), 1248 (w), 1157 (vs), 1055 (w), 855 (s), 747 (s), 698 (vs) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.16 (d, *J* = 8.7 Hz, 2H), 7.33 (d, *J* = 8.7 Hz, 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.), 4-dimethylaminopyridine (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 NH4Cl 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: ṽ =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 $[M+H]^+$.

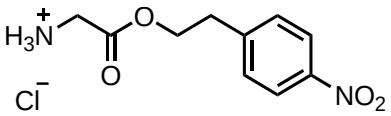
Compound 15/H-Val-Onpe•HCl



Yield: 99%
IR: ṽ = 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).

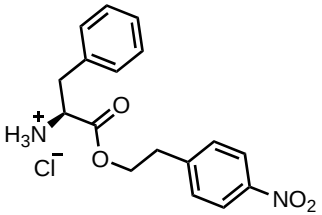
^{13}C NMR (101 MHz, DMSO-d_6) δ: 168.8, 146.3, 130.4, 123.4, 65.3, 57.2, 33.8, 29.2, 18.3, 17.4
HRMS (ESI): calculated for $C_{13}H_{19}N_2O_4^+$: m/z = 267.1339 $[M+H]^+$; found: m/z = 267.1139 $[M+H]^+$.

Compound 16/H-Gly-Onpe•HCl



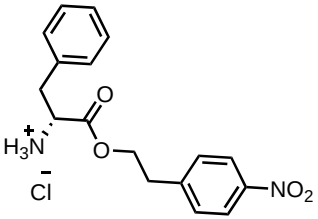
Yield: 99%
IR: ṽ = 2949 (w), 1746 (s), 1515 (vs), 1310 (vs), 1238 (vs), 1053 (w), 955 (s), 905 (s) 856 (s), 698 (s) cm^{-1}
^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.30 (br. s, 3H), 8.18 (d, J = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 4.44 (t, *J* = 6.4 Hz, 2H), 3.77 (s, 2H), 3.09 (t, *J* = 6.4 Hz, 2H)
^{13}C 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 $[M+H]^+$.

Compound 17/H-Phe-Onpe•HCl



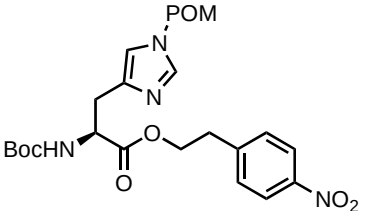
Yield: 99%
IR: ṽ = 3142 (w), 2988 (w), 2802 (w), 1740 (vs), 1601 (s), 1518 (vs), 1490 (vs), 1351 (vs), 1232 (vs), 1191 (s), 1102 (s), 981 (s), 856 (vs), 755 (vs), 736 (vs), 706 (vs) cm^{-1}
^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.74 (br. s, 3H), 8.16 (d, *J* = 8.7 Hz, 2H), 7.49 (d, *J* = 8.7 Hz, 2H), 7.33 – 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: ṽ = 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 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 r.t. 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% *i*-hexane in EtOAc to pure EtOAc). The pivalate protected histidine derivative was obtained as a yellow oil.

Yield: 55%

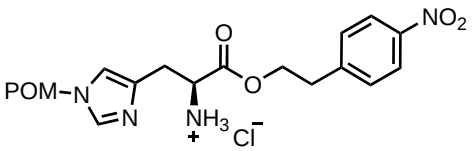
IR: 2850 (w), 2600 (w), 1738 (vs), 1624 (w), 1598 (w), 1573 (w), 1516 (vs), 1454 (w), 1414 (w), 1350 (vs), 1282 (w), 1191 (w), 1124 (vs), 1044 (w), 1008 (w), 854 (vs), 773 (w), 749 (w), 701 (w) cm^{-1}

^{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 21



Yield: 98%

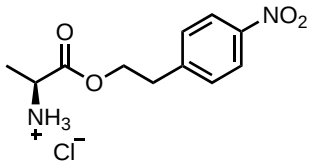
IR: ṽ = 2960 (w), 1737 (vs), 1624 (w), 1598 (w), 1516 (vs), 1454 (w), 1415 (w), 1351 (vs), 1282 (w), 1188 (w), 1124 (vs), 1045 (w), 1008 (w), 854 (s), 774 (w), 749 (w), 701 (w) cm^{-1}

^{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 $[M+H]^+$.

H-Ala-Onpe•HCl



Yield: 79% over two steps

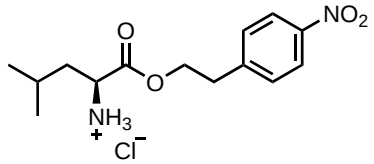
IR: ṽ = 2843 (m), 1730 (s), 1598 (m), 1345 (s), 1269 (w), 1233 (s), 1195 (m), 1115 (m), 820 (m), 746 (m) cm^{-1}

^{1}H NMR (400 MHz, DMSO-d_6): δ: 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 NMR (100 MHz, DMSO-d_6, 298 K) δ : 169.9, 146.4, 146.3, 130.4, 123.5, 65.3, 47.8, 33.9, 15.7

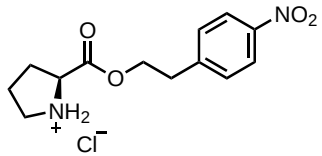
HRMS (ESI): calculated for $C_{11}H_{15}N_2O_4^+$: m/z = 239.1026 $[M+H]^+$; found: m/z = 239.1027$[M+H]^+$.

H-Leu-Onpe•HCl



Yield: 96% over two steps
IR: ṽ = 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) cm^{-1}
^{1}H NMR (400 MHz, DMSO-d_6) δ: 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 NMR (100 MHz, DMSO-d_6) δ: 169.8, 146.4, 130.4, 123.4, 65.3, 50.4, 33.8, 23.7, 22.2, 21.8
HRMS (ESI): calculated for $C_{14}H_{21}N_2O_4^+$: m/z = 281.1496 $[M+H]^+$; found: m/z = 281.1495 $[M+H]^+$.

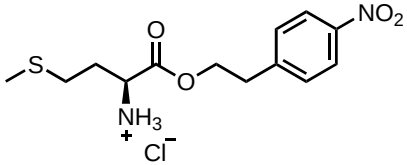
H-Pro-Onpe•HCl



Yield: 73% over two steps
IR: ṽ = 3400 (w), 2879 (w), 1649 (s), 1513 (s), 1432 (m), 1318 (s), 1159 (w), 1048 (m), 856 (m), 747 (m) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 175.4, 147.0, 145.6, 129.9, 123.9, 64.3, 59.8, 47.1, 35.0, 30.4, 25.6

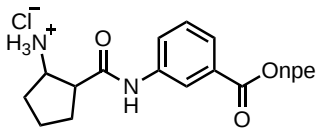
HRMS (ESI): calculated for $C_{13}H_{17}N_2O_4^+$: m/z = 265.1183 $[M+H]^+$; found: m/z = 265.1179 $[M+H]^+$.

H-Met-Onpe•HCl



Yield: 92% over two steps
IR: ṽ = 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) cm^{-1}
^{1}H NMR (400 MHz, DMSO-d_6) δ: 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 NMR (100 MHz, DMSO-d_6) δ: 169.1, 146.3, 146.3, 130.4, 123.5, 65.4, 50.8, 33.8, 29.3, 28.2, 14.1
HRMS (ESI): calculated for $C_{13}H_{19}N_2O_4S^+$: m/z = 299.1060 $[M+H]^+$; found: m/z = 299.1058 $[M+H]^+$.

H-ProAbz-Onpe•HCl

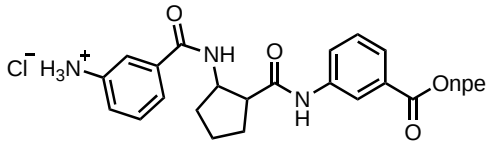


Boc-protected 3-aminobenzoic acid (1 eq.) was dissolved in CH_2Cl_2 under inert atmosphere and cooled to 0 ˚C. Then npeOH (1.3 eq.) and PPh_3 (1.3 eq.) were added followed by slow addition of DIAD (1.3 eq.). The reaction mixture was left to stir for 2 h at room temperature. Then the solution was washed with water, organic phase was dried over Na_2SO_4 and evaporated. The crude was then dissolved in 4 M

HCl in dioxane at 0 ˚C for 2 h. The solvent was then vaporized and H-Abz-Onpe•HCl was precipitated with excess Et_2O and dried. Without further purification, the obtained H-Abz-Onpe•HCl (1 eq.) was dissolved in CH_2Cl_2 and to it, Boc-proline (1.4 eq.), DIPEA (3.8 eq.) and activator PyCloP (1.4 eq.) were added. The reaction was stirred at r.t. overnight. After that, the mixture was washed with sat. $NaHCO_3$ solution once and brine twice. The organic solvent removed under reduced pressure. This crude was then dissolved in 4 M HCl in dioxane at 0 ˚C for 1 h, and the final product H-ProAbz-Onpe•HCl was precipitated with excess Et_2O and dried.

Yield: 87% overall
IR: ṽ = 3085 (w), 2953 (w), 2850 (w), 2693 (w), 2539 (w), 1690 (vs), 1598 (vs), 1564 (vs), 1516 (vs), 1446 (s), 1342 (vs), 1299 (vs), 1249 (vs), 1176 (s), 1084 (vs), 1042 (s), 849 (s), 754 (vs), 700 (s), 650 (s) cm^{-1}
^{1}H NMR (500 MHz, DMSO-d_6) δ: 11.29 (s, 1H), 8.32 (t, *J* = 1.9 Hz, 1H), 8.19 (d, *J* = 8.7 Hz, 2H), 7.87 (ddd, *J* = 8.2, 2.3, 1.1 Hz, 1H), 7.64 (dd, *J* = 8.1, 6.1 Hz, 3H), 7.49 (t, *J* = 7.9 Hz, 1H), 4.53 (t, *J* = 6.4 Hz, 2H), 4.45 (dd, *J* = 8.4, 6.7 Hz, 1H), 3.28 (dt, *J* = 9.1, 7.0 Hz, 3H), 3.19 (t, *J* = 6.4 Hz, 2H), 2.47 – 2.40 (m, 1H), 2.02 – 1.91 (m, 3H)
^{13}C NMR (126 MHz, DMSO-d_6) δ: 167.24, 165.28, 146.64, 146.32, 138.66, 130.41, 130.16, 129.45, 124.61, 124.05, 123.53, 119.92, 66.37, 64.72, 59.58, 45.71, 34.16, 29.71, 23.61
HRMS (ESI): calculated for $C_{20}H_{22}N_3O_5^+$: m/z = 384.1554 $[M+H]^+$; found = m/z = 384.1550 $[M+H]^+$

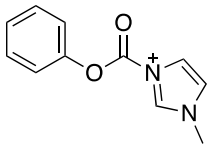
H-AbzProAbz-Onpe•HCl



H-Pro-Abz-Onpe•HCl (1 eq.) was dissolved in DMF under inert atmosphere. Next, Boc-protected 3-aminobenozic acid (1.2 eq.), NEt_3 (4 eq.), HOBt (1.2 eq.) and EDC•HCl (1.2 eq.) were added, and the solution was stirred at r.t. overnight. The reaction mixture was then washed with sat. $NaHCO_3$ solution once and brine twice. The organic solvent was removed under reduced pressure. The crude was then dissolved and stirred in 4 M HCl in dioxane at 0 ˚C for 1 h, and the solvent was removed under reduced pressure. To this crude, excess Et_2O was added to precipitate H-AbzProAbz-Onpe•HCl, which was then dried *in vacuo*.

Yield: 74% overall
IR: ṽ = 3197 (w), 3067 (w), 2955 (w), 2857 (s), 2590 (s), 1717 (s), 1595 (s), 1516 (vs), 1436 (s), 1344 (vs), 1372 (s), 1176 (s), 1082 (s), 856 (w), 749 (vs), 698 (w) cm^{-1}
^{1}H NMR (500 MHz, DMSO-d_6) δ: 10.49 (s, 1H), 8.35 (s, 1H), 8.18 (d, *J* = 8.6 Hz, 2H), 7.83 (d, *J* = 7.9 Hz, 1H), 7.64 (d, *J* = 8.6 Hz, 2H), 7.58 (d, *J* = 7.6 Hz, 1H), 7.51 (d, *J* = 14.9 Hz, 3H), 7.44 (q, *J* = 7.1 Hz, 2H), 4.63 (dd, *J* = 8.2, 5.3 Hz, 1H), 4.53 (t, *J* = 6.3 Hz, 2H), 3.68 – 3.57 (m, 2H), 3.52 (dt, *J* = 10.1, 6.2 Hz, 1H), 3.37 (q, *J* = 7.0 Hz, 1H), 3.19 (t, *J* = 6.5 Hz, 2H), 2.32 (h, *J* = 8.4 Hz, 1H), 1.93 (dddt, *J* = 30.0, 17.8, 11.2, 6.4 Hz, 3H), 1.08 (t, *J* = 7.0 Hz, 1H).
^{13}C NMR (126 MHz, DMSO-d_6) δ: 170.71, 167.14, 165.48, 146.65, 146.32, 139.54, 137.67, 130.40, 130.04, 129.75, 129.23, 125.37, 123.91, 123.81, 123.73, 123.52, 121.19, 119.67, 66.37, 64.94, 64.60, 60.77, 49.99, 34.19, 29.74, 25.03, 15.18.
HRMS (ESI): calculated for $C_{27}H_{27}N_4O_6^+$: m/z = 503.1925 $[M+H]^+$; found: m/z = 503.1919 $[M+H]^+$

Compound 22

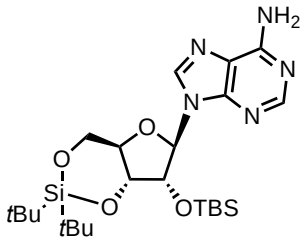


Phenyl chloroformate (4 ml, 31.9 mmol, 1 eq.) was dissolved in dry CH_2Cl_2 under nitrogen and cooled to 0 ˚C. Then *N*-methylimidazole (2.54 ml, 31.9 mmol, 1 eq.) was added dropwise. The mixture was allowed to stir at room temperature for 2 h. Afterwards the reaction mixture was filtered, the precipitate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

Yield: 95%
IR: ṽ = 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 literature.[314]

Compound 24

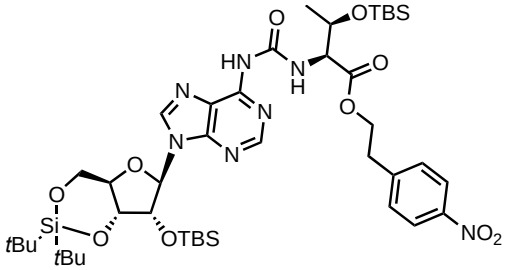


Adenosine **23** (1 g, 3.74 mmol, 1 eq.) was suspended in DMF and di-tert-butylsilyl ditriflate
(1.46 ml, 4.49 mmol, 1.2 eq.) was added dropwise under stirring at 0 ˚C. The resulting solution was stirred at 0 ˚C for 45 min. Then imidazole (1.27 g, 18.7 mmol, 5 eq.) was added and the reaction was warmed to r.t. 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: ṽ = 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 $[M+H]^+$.

General procedure of amino acid coupling to protected adenosine

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 r.t. for 2 h (the solution in time becomes clear). Afterwards the npe-protected amino acid salt (2 eq.) was added together with NEt_3 (2 eq.) as a solution in CH_2Cl_2 and the resulting solution was stirred overnight at r.t. 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.

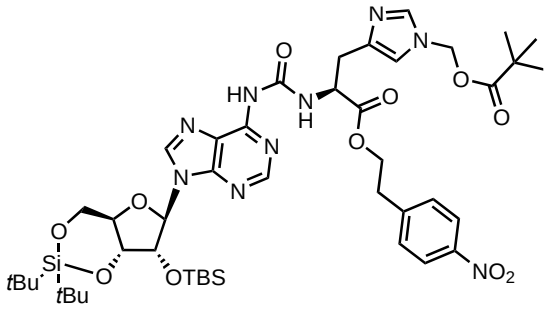
Compound 25T



Eluent: Hex/EtOAc (4/3, v/v).
Yield: 91%
IR: ṽ = 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

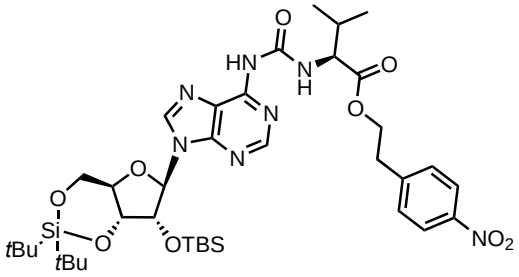
HRMS (ESI): calculated for $C_{43}H_{72}N_7O_{10}Si_3^+$: m/z = 930.4643 $[M+H]^+$; found: m/z = 930.4640 $[M+H]^+$.

Compound 25H



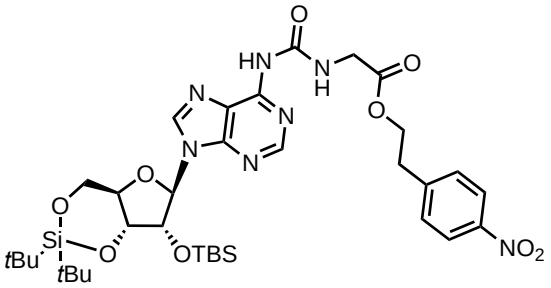
Eluent: 10% CH_2Cl_2 in EtOAc to pure EtOAc.
Yield: 86%
IR: ṽ = 3854 (w), 3745 (w), 3650 (w), 2932 (w), 2858 (w), 2361 (w), 2341 (w), 1735 (s), 1670 (s), 1654 (w), 1610 (w), 1587 (w), 1521 (vs), 1472 (s), 1395 (w), 1345 (vs), 1252 (w), 1166 (w), 1118 (vs), 1055 (vs), 999 (w), 894 (w), 826 (vs), 781 (s), 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 25V



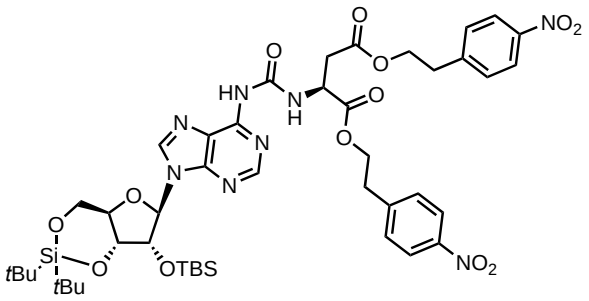
Eluent: Hex/EtOAc (4/3, v/v).
Yield: 78%
IR: ṽ = 3230 (w), 2960 (w), 2960 (w), 2858 (w), 1741 (s), 1702 (vs), 1611 (s), 1520 (vs), 1466 (s), 1345 (vs), 1250 (s), 1139 (vs), 1057 (vs), 999 (s), 894 (s), 810 (vs), 781 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.01 (d, *J* = 8.4 Hz, 1H), 8.50 (s, 1H), 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 $[M+H]^+$.

Compound 25G



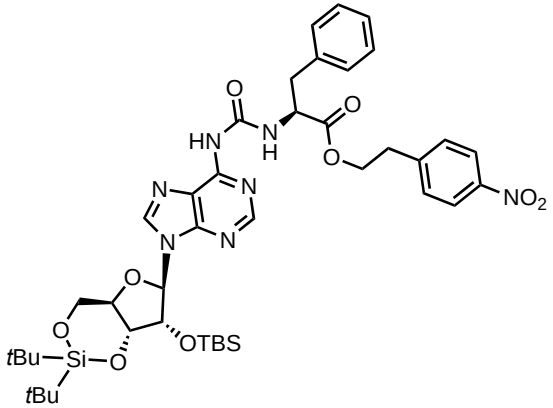
Eluent: Hex/EtOAc (4/3, v/v).
Yield: 72%
IR: ṽ = 3239 (w), 2932 (s), 2858 (s), 1749 (s), 1703 (vs), 1611 (s), 1520 (vs), 1468 (s), 1345 (s), 1252 (s), 1141 (w), 1055 (s), 990 (w), 894 (w), 826 (s), 750 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 9.95 (br. s, 1H), 8.50 (s, 1H), 8.21 – 8.03 (m, 3H), 7.38 (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 25D



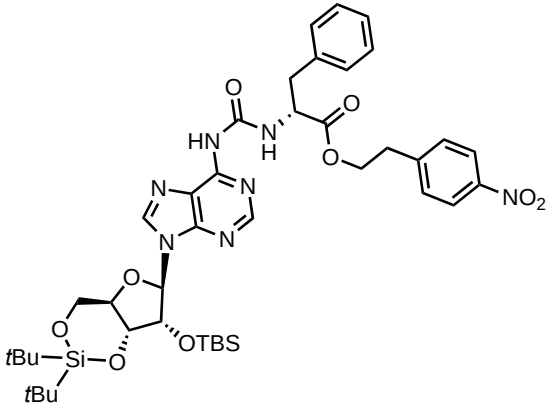
Eluent: Hex/EtOAc (2/1, v/v).
Yield: 87%
IR: ṽ = 3229 (w), 2953 (w), 2929 (w), 2857 (w), 1735 (s), 1693 (s), 1607 (s), 1589 (s), 1517 (vs), 1469 (s), 1391 (w), 1310 (vs), 1292 (w), 1251 (w), 1210 (w), 835 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.28 (s, 1H), 8.42 (s, 1H), 8.22 – 8.12 (m, 1H), 8.07 (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 $[M+H]^+$.

Compound 25F



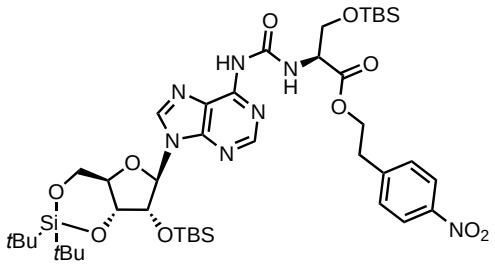
Eluent: Hex/EtOAc (4/1. v/v)
Yield: 68%
IR: ṽ = 3190 (w), 2933 (w), 2858 (w), 1742 (w), 1702 (vs), 1612 (s), 1587 (w), 1521 (vs), 1469 (vs), 1345 (vs), 1253 (s), 1057 (s), 1000 (w), 828 (vs) cm^{-1}
^{1}H NMR (400. MHz, CD_2Cl_2) δ: 10.00 (d, *J* = 7.5 Hz, 1H), 8.94 (s, 1H), 8.37 (d, *J* = 1.2 Hz, 2H), 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 25f



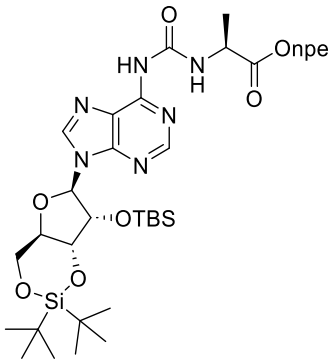
Eluent: Hex/EtOAc (4/1, v/v).
Yield: 71%
IR: ṽ = 3192 (w), 2933 (w), 2858 (w), 1740 (w), 1700 (vs), 1611 (s), 1520 (vs), 1466 (vs), 1345 (vs), 1252 (s), 1166 (w), 1139 (w), 1054 (vs), 998 (s), 893 (s), 826 (vs), 736 (vs) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 10.01 (d, *J* = 7.6 Hz, 1H), 8.96 (s, 1H), 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 25S



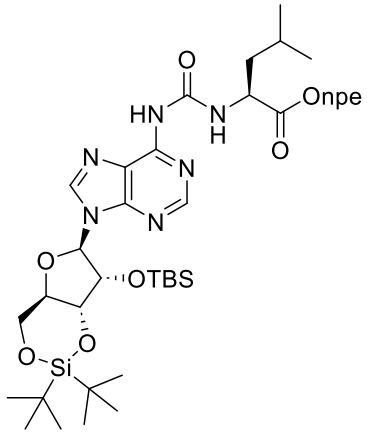
Eluent: Hex/EtOAc (1/1, v/v).
Yield: 85%
IR: ṽ = 3238 (w), 2931 (s), 2859 (s), 1737 (s), 1701 (vs), 1610 (s), 1520 (vs), 1470 (s), 1345 (s), 1251 (s), 1136 (w), 1057 (s), 998 (w), 899 (w), 840 (s), 778 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.18 (d, *J* = 8.3 Hz, 1H), 8.61 (s, 1H), 8.44 (s, 1H), 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 $[M+H]^+$.

Compound 25A



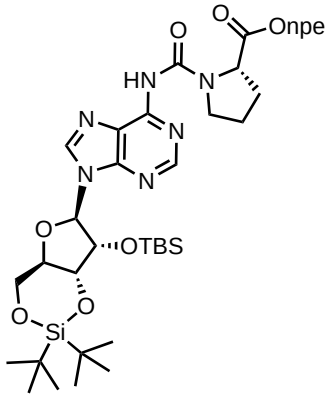
Yield: 70%
IR: ṽ = 3239 (w), 2929 (w), 2855 (w), 1744 (m), 1698 (s), 1610 (m), 1586 (m), 1519 (s), 1463 (m), 1344 (s), 1250 (s), 1138 (s), 1054 (s), 998 (m), 894 (s), 825 (s), 782 (s), 749 (s)
^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.93 – 9.86 (m, 1H), 8.50 (s, 1H), 8.20 (s, 1H), 8.11 – 8.03 (m, 3H), 7.41 – 7.35 (m, 2H), 5.97 (s, 1H), 4.67 – 4.57 (m, 2H), 4.55 – 4.38 (m, 4H), 4.24 (td, *J* = 10.1, 5.1 Hz, 1H), 4.05 (dd, *J* = 10.5, 9.3 Hz, 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 NMR (100 MHz, $CDCl_3$) *δ*: 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): calculated for $C_{36}H_{56}N_7O_9Si_2^+$: *m/z* = 786.3678 $[M+H]^+$, found: *m/z* = 786.3682 $[M+H]^+$.

Compound 25L



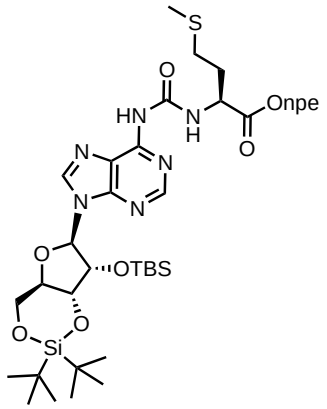
Yield: 82%
IR: ṽ = 3237 (w), 2168 (w), 1666 (s), 1572 (w), 1511 (s), 1429 (w), 1335 (m), 1271 (s), 1227 (m), 1178 (w), 1151 (w), 1119 (m), 1090 (s), 1019 (w), 908 (m), 843 (s), 781 (s) cm^{-1}
^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.80 (d, *J* = 7.7 Hz, 1H), 8.49 (s, 1H), 8.13 (s, 1H), 8.09 – 8.03 (m, 3H), 7.41 – 7.35 (m, 2H), 5.96 (s, 1H), 4.64 – 4.56 (m, 2H), 4.50 (ddd, *J* = 11.7, 9.4, 4.9 Hz, 2H), 4.43 (ddt, *J* = 11.1, 6.6, 4.5 Hz, 2H), 4.24 (td, *J* = 10.1, 5.1 Hz, 1H), 4.05 (dd, *J* = 10.5, 9.3 Hz, 1H), 3.08 (t, *J* = 6.5 Hz, 2H), 1.78 – 1.63 (m, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.94 (s, 15H), 0.18 (s, 3H), 0.16 (s, 3H)
^{13}C NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{39}H_{62}N_7O_9Si_2^+$: *m/z* = 828.4148 $[M+H]^+$, found: *m/z* = 828.4149 $[M+H]^+$.

Compound 25P



Yield: 71%
IR: ṽ = 2933 (w), 1741 (w), 1649 (w), 1519 (m), 1401 (m), 1344 (s), 1166 (m), 1140 (m), 1057 (s), 750 (m) cm^{-1}
For major rotamer:
^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$, 298 K) δ: 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): calculated for $C_{38}H_{58}N_7O_9Si_2^+$: m/z = 812.3829 $[M+H]^+$; found: m/z = 812.3835 $[M+H]^+$.

Compound 25M



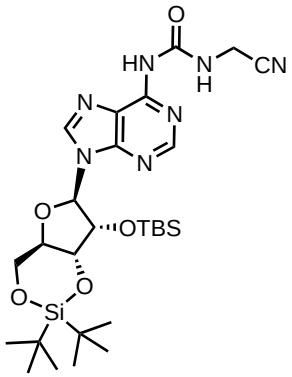
Yield: 98%

^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{38}H_{60}N_7O_9SSi_2^+$: m/z = 846.3706 $[M+H]^+$; found: m/z = 846.3704 $[M+H]^+$.

Compound 25Gcn



Yield: 77%

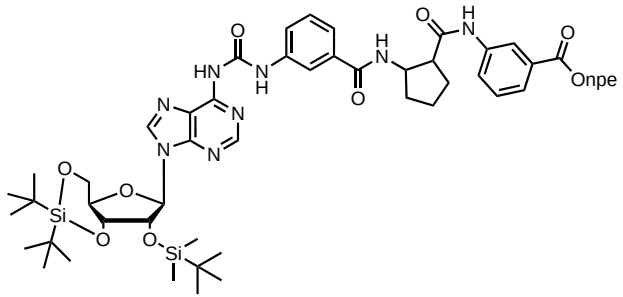
IR: ṽ = 3119 (w), 2930 (m), 2857 (m), 2168 (w), 1706 (s), 1658 (m), 1612 (m), 1525 (m), 1466 (m), 1394 (w), 1353 (m), 1249 (s), 1141 (s), 1058 (s), 997 (m), 892 (m), 825 (s), 783 (s) cm^{-1}

^{1}H NMR (600 MHz, $CDCl_3$) δ: 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 NMR (150 MHz, $CDCl_3$) δ: 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): calculated for $C_{27}H_{46}N_7O_5Si_2^+$: m/z = 604.3093 $[M+H]^+$; found: m/z = 604.3094 $[M+H]^+$.

Compound 25AbPAb

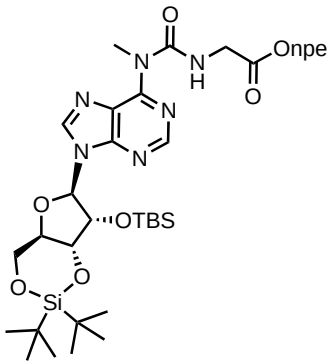


Yield: 70%
IR: ṽ = 3256 (w), 3081 (w), 2953 (w), 2927 (w), 2857 (w), 1700 (s), 1589 (vs), 1559 (vs), 1516 (vs), 1345 (vs), 1286 (s), 1240 (vs), 1107 (s), 858 (w), 749 (vs) cm^{-1}
^{1}H NMR (500 MHz, CD_2Cl_2) δ: 12.03 (s, 1H), 9.72 (s, 1H), 8.74 (s, 1H), 8.63 (s, 1H), 8.21 (d, *J* = 4.3 Hz, 2H), 8.16 (d, *J* = 8.6 Hz, 2H), 7.96 (s, 1H), 7.72 (s, 2H), 7.63 (d, *J* = 7.8 Hz, 1H), 7.49 (d, *J* = 8.6 Hz, 2H), 7.44 (t, *J* = 7.9 Hz, 1H), 7.31 (t, *J* = 8.0 Hz, 2H), 6.02 (s, 1H), 4.93 (dd, *J* = 7.5, 4.7 Hz, 1H), 4.64 (d, *J* = 4.6 Hz, 1H), 4.51 (dt, *J* = 8.5, 6.9 Hz, 4H), 4.26 (td, *J* = 10.0, 5.1 Hz, 1H), 4.12 – 4.06 (m, 1H), 3.70 – 3.57 (m, 2H), 3.18 (t, *J* = 6.5 Hz, 2H), 2.50 (tt, *J* = 8.9, 4.9 Hz, 1H), 2.10 (dt, *J* = 19.8, 6.6 Hz, 2H), 1.93 – 1.87 (m, 1H), 1.08 (s, 9H), 1.06 (s, 9H), 0.96 (s, 9H), 0.19 (s, 3H), 0.17 (s, 3H)
^{13}C NMR (126 MHz, CD_2Cl_2) δ: 171.51, 169.82, 166.26, 151.79, 151.35, 150.34, 147.29, 146.57, 141.87, 139.30, 138.84, 137.32, 130.98, 130.44, 129.75, 129.47, 129.29, 124.98, 124.43, 124.08, 122.77, 122.22, 122.10, 121.48, 120.77, 119.17, 92.88, 76.31, 76.05, 75.25, 68.14, 64.92, 61.54, 51.15, 41.23, 35.41, 27.65, 27.23, 26.07, 22.97, 20.63, 18.61, -4.14, -4.86
HRMS (ESI): calculated for $C_{52}H_{68}N_9O_{11}Si_2^+$: m/z = 1050.4571 $[M+H]^+$; found: m/z = 1050.4567 $[M+H]^+$

General procedure for methylation of *aa*^{6}A derivatives to m^6*aa*^{6}A derivatives

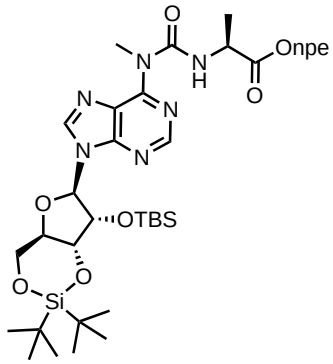
The amino acid-modified adenosine derivative **25*aa*** (1.0 eq.) was dissolved in DMF and cooled to 0 ˚C. To the solution were added K_2CO_3 (3.0 eq.) together with MeI (2.0 eq.) and the reaction was stirred at r.t. for 2 h. The reaction mixture was diluted with H_2O and extracted three times with EtOAc. 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 **25*aa*-m** as a white foam.

Compound 25G-m



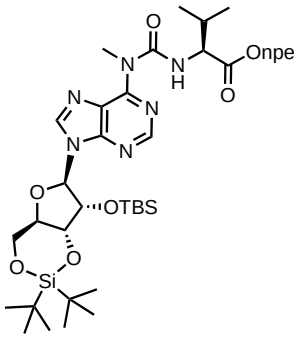
Yield: 76%;
IR: ṽ = 3235 (w), 2932 (w), 2858 (w), 1749 (w), 1686 (m), 1568 (w), 1521 (s), 1470 (m), 1347 (s), 1264 (s), 1167 (w), 1135 (m), 1055 (m), 1000 (w), 894 (w), 827 (m), 732 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.97 (t, *J* = 5.4 Hz, 1H), 8.50 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.97 (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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{36}H_{56}N_7O_9Si_2^+$: m/z = 786.3673 $[M+H]^+$; found: m/z = 786.3674 $[M+H]^+$.

Compound 25A-m



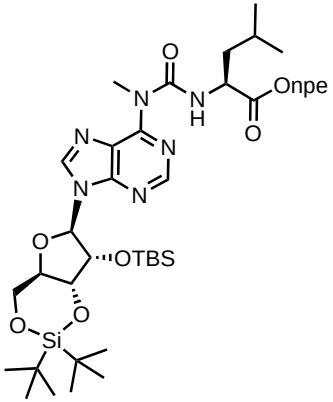
Yield: 85%
IR: $\tilde{\nu}$ = 3190 (w), 2934 (w), 2858 (w), 1744 (m), 1686 (m), 1567 (m), 1519 (s), 1469 (s), 1344 (s), 1250 (m), 1166 (m), 1133 (m), 1057 (s), 1000 (m), 895 (m), 825 (s), 777 (m), 749 (m) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$,) δ: 8.51 (s, 1H), 8.10 (d, *J* = 8.7 Hz, 2H), 8.01 (s, 1H), 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{37}H_{58}N_7O_9Si_2^+$: m/z = 800.3829 $[M+H]^+$; found: m/z = 800.3836 $[M+H]^+$.

Compound 25V-m



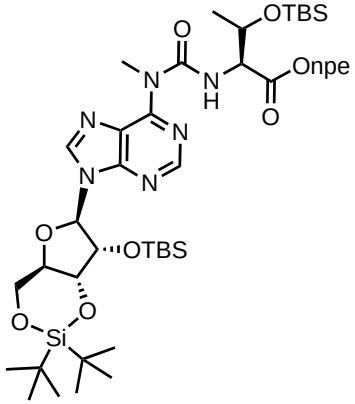
Yield: 90%;
IR: $\tilde{\nu}$ = 3246 (m), 2977 (w), 1732 (m), 1686 (s), 1524 (s), 1469 (w), 1372 (m), 1254 (s), 1177 (m), 1147 (w), 1107 (s), 1050 (s), 1020 (m), 926 (m), 853 (w), 790 (m), 744 (w) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 11.07 (d, *J* = 7.6 Hz, 1H), 8.48 (s, 1H), 8.09 (d, *J* = 8.7 Hz, 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{39}H_{62}N_7O_9Si_2^+$: m/z = 828.4142 $[M+H]^+$; found: m/z = 828.4143 $[M+H]^+$.

Compound 25L-m



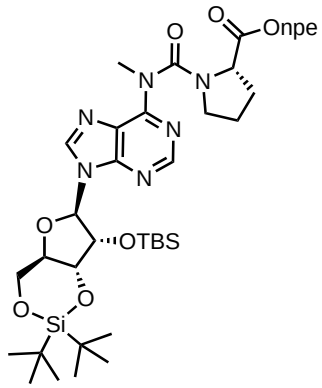
Yield: 76%
IR: ṽ = 3230 (w), 2933 (w), 1740 (s), 1690 (s), 1580 (s), 1520 (s), 1469 (s), 1345 (s), 1259 (s), 1134 (s), 1057 (s), 1013 (s), 900 (w), 826 (s), 780 (s), 750 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.92 (d, *J* = 6.8 Hz, 1H), 8.47 (s, 1H), 8.09 (d, *J* = 8.7 Hz, 2H), 7.97 (s, 1H), 7.38 (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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{40}H_{64}N_7O_9Si_2^+$: m/z = 842.4299 $[M+H]^+$; found: m/z = 842.4296 $[M+H]^+$.

Compound 25T-m



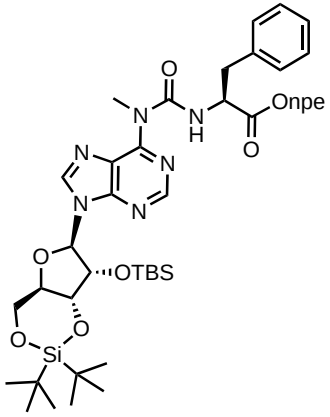
Yield: 73%
IR: ṽ = 3237 (w), 2931 (s), 2857 (s), 1737 (s), 1701 (s), 1610 (s), 1520 (s), 1465 (s), 1345 (s), 1250 (s), 1136 (w), 1057 (s), 998 (w), 894 (w), 840 (s), 777 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 11.00 (d, *J* = 8.7 Hz, 1H), 8.43 (s, 1H), 8.01-7.96 (m, 3H), 7.32 (d, *J* = 8.7 Hz, 2H), 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{44}H_{74}N_7O_{10}Si_3^+$: m/z = 944.4799 $[M+H]^+$; found: m/z = 944.4793 $[M+H]^+$.

Compound 25P-m



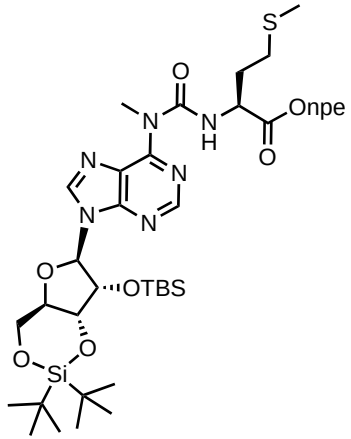
Yield: 73%
IR: ṽ = 2933 (w), 1744 (w), 1683 (m), 1583 (m), 1392 (m), 1345 (s), 1166 (m), 1056 (s), 1002 (m), 783 (m) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.46 (s, 1H), 8.17- 8.15 (m, 2H), 7.84 (s, 1H), 7.39-7.38 (m, 2H), 5.93 (s, 1H), 4.58-4.57 (m, 1H), 4.54 (br s, 2H), 4.49 (dd, *J* = 9.8, 5.1 Hz, 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{39}H_{60}N_7O_9Si_2^+$: m/z = 826.3985 $[M+H]^+$; found: m/z = 826.3991 $[M+H]^+$.

Compound 25F-m



Yield: 85%;
IR: ṽ = 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) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 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 NMR (100 MHz, CD_2Cl_2) δ: 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): calculated for $C_{43}H_{62}N_7O_9Si_2^+$: m/z = 876.4142 $[M+H]^+$; found: m/z = 876.4148 $[M+H]^+$.

Compound 25M-m



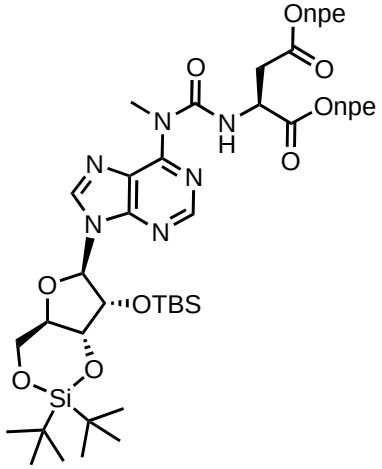
Yield: 78%

^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{39}H_{62}N_7O_9SSi_2^+$: m/z = 860.3863 $[M+H]^+$; found: m/z = 860.3858 $[M+H]^+$.

Compound 25D-m



Yield: 77%

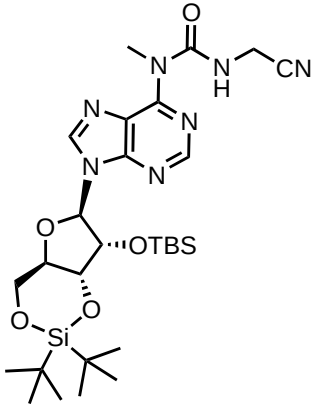
IR: ṽ = 2933 (w), 1737 (m), 1683 (m), 1569 (m), 1518 (s), 1344 (s), 1166 (m), 1057 (m), 1000 (m), 780 (m) cm^{-1}

^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{46}H_{65}O_{13}N_8Si_2^+$: m/z = 993.4203 $[M+H]^+$; found: m/z = 993.4215 $[M+H]^+$.

Compound 25Gcn-m

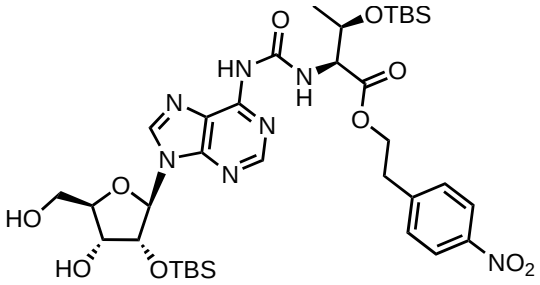


Yield: 74%
IR: ṽ = 3121 (w), 2933 (m), 2896 (w), 2857 (m), 2168 (w), 1692 (s), 1570 (s), 1525 (s), 1469 (s), 1422 (w), 1360 (m), 1328 (m), 1308 (w), 1299 (w), 1278 (m), 1249 (m), 1218 (w), 1198 (w), 1165 (s), 1141 (s), 1111 (m), 1062 (s), 1024 (s), 1001 (s), 968 (w), 889 (m), 825 (s), 784 (s) cm^{-1}
^{1}H NMR (600 MHz, $CDCl_3$) δ: 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 NMR (150 MHz, $CDCl_3$) δ: 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): calculated for $C_{28}H_{48}N_7O_5Si_2^+$: m/z = 618.3250 $[M+H]^+$; found: m/z = 618.3256 $[M+H]^+$.

General procedure for 3', 5' silyl-deprotection of (m^6)*aa*^{6}A derivatives

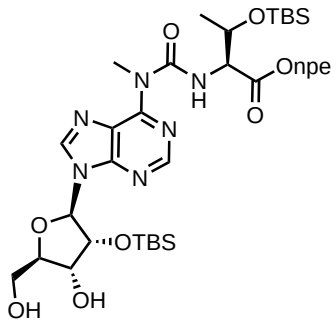
The modified adenosine (0.86 mmol) was dissolved in CH_2Cl_2 under N_2 atmosphere and transferred into a plastic flask. Pyridine (1 mL) was added and the solution was cooled in an icebath. Then Py*(HF)n (140 µL) was added and the mixture was stirred at 0 ˚C for 2 h. The reaction was quenched with sat. $NaHCO_3$ 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 26T



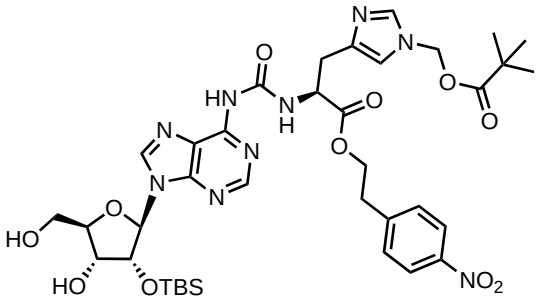
Yield: 95%
IR: ṽ = 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (vs), 1469 (s), 1345 (vs), 1313 (w), 1250 (vs), 1129 (w), 1093 (s), 835 (vs), 760 (vs) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 9.90 (d, *J* = 9.1 Hz, 1H), 8.50 (s, 1H), 8.46 (s, 1H), 8.10 – 8.08 (m, 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 26T-m



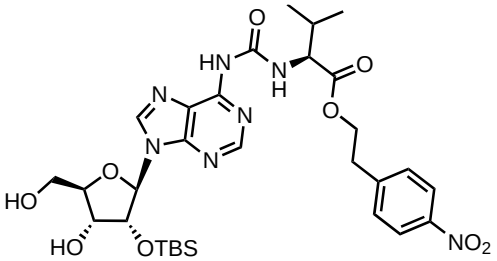
Yield: 94%;
IR: ṽ = 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (s), 1469 (s), 1345 (s), 1313 (w), 1250 (s), 1129 (w), 1093 (s), 835 (s), 760 (s) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 10.90 (d, *J* = 8.6 Hz, 1H), 8.48 (s, 1H), 8.10 (d, *J* = 8.6 Hz, 2H), 7.96 (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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{36}H_{58}N_7O_{10}Si_2^+$: m/z = 804.3778 $[M+H]^+$; found: m/z = 804.3768 $[M+H]^+$.

Compound 26H



Yield: 75%
IR: ṽ = 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 26V



Yield: 95%

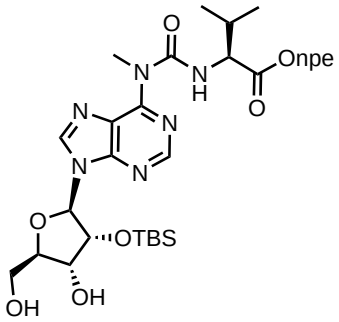
IR: ṽ = 3245 (w), 2930 (w), 2857 (w), 1743 (s), 1695 (vs), 1590 (s), 1518 (vs), 1471 (s), 1344 (vs), 1252 (s), 1212 (w), 1188 (s), 1130 (w), 1085 (s), 836 (vs) cm^{-1}

^{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 26V-m



Yield: 95%

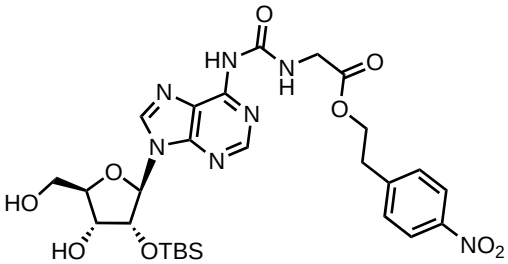
IR: ṽ = 3244 (w), 2952 (w), 2929 (w), 2359 (w), 1736 (w), 1681 (m), 1571 (m), 1518 (s), 1469 (m), 1422 (w), 1345 (s), 1255 (m), 1187 (m), 1145 (m), 1089 (m), 1046 (w), 1016 (m), 907 (m), 857 (m), 837 (s), 780 (s), 746 (m) cm^{-1}

^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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

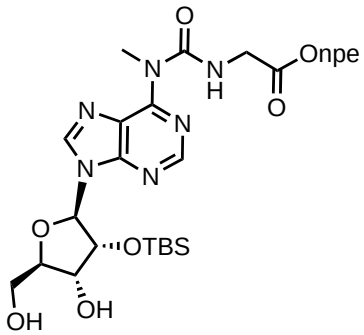
HRMS (ESI): calculated for $C_{31}H_{46}N_7O_9Si^+$: m/z = 688.3121 $[M+H]^+$; found: m/z = 688.3120 $[M+H]^+$.

Compound 26G



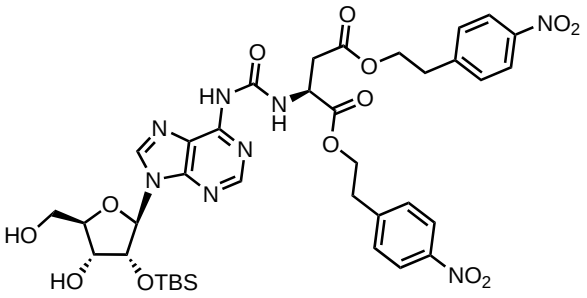
Yield: 97%
IR: $\tilde{\nu}$ = 3250 (w), 2932 (w), 2859 (w), 1740 (s), 1680 (vs), 1595 (s), 1520 (vs), 1470 (s), 1340 (vs), 1254 (s), 1213 (w), 1168 (s), 1130 (w), 1085 (s), 836 (vs) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 9.92 (t, *J* = 5.7 Hz, 1H), 8.92 (br. s, 1H), 8.54 (s, 1H), 8.25 (s, 1H), 8.13 (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 $[M + H]^+$.

Compound 26G-m



Yield: 97%
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, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{28}H_{40}N_7O_9Si^+$: m/z = 646.2651 $[M+H]^+$; found: m/z = 646.2645 $[M+H]^+$.

Compound 26D



Yield: 98%

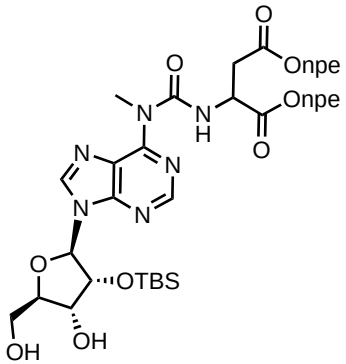
IR: ṽ = 3229 (w), 2953 (w), 2929 (w), 2857 (w), 1735 (s), 1693 (s), 1607 (s), 1589 (s), 1517 (vs), 1469 (s), 1391 (w), 1310 (vs), 1292 (w), 1251 (w), 1210 (w), 835 (s) cm^{-1}

^{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 26D-m



Yield: 89%

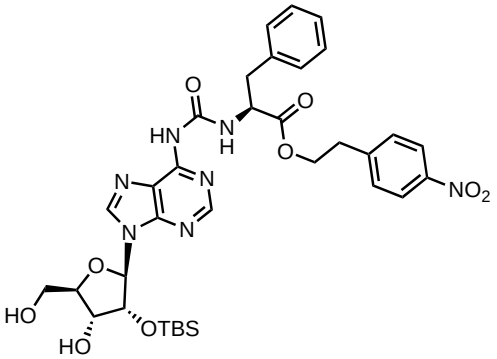
IR: ṽ = 2930 (w), 1735 (m), 1682 (m), 1570 (m), 1516 (s), 1468 (m), 1261 (m), 1018 (m), 837 (m), 781 (m) cm^{-1}

^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{38}H_{49}O_{13}N_8Si^+$: m/z = 853.3182 $[M+H]^+$; found: m/z = 853.3187 $[M+H]^+$.

Compound 26F



Yield: 80%

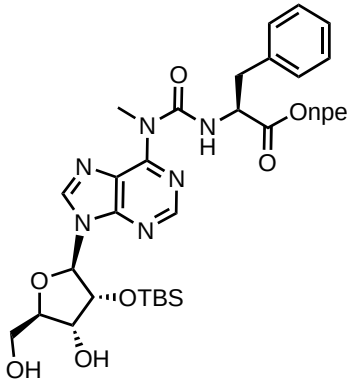
IR: ṽ = 3240 (w), 2929 (w), 2857 (w), 1742 (w), 1699 (vs), 1613 (s), 1520 (vs), 1471 (s), 1345 (vs), 1255 (w), 839 (s) cm^{-1}

^{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 26F-m



Yield: 90%

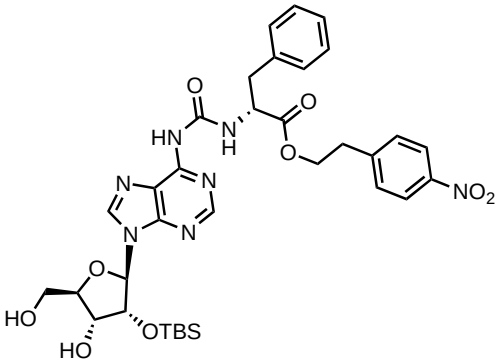
IR: ṽ = 3391 (w), 3194 (w), 2951 (w), 2855 (w), 1738 (s), 1681 (s), 1568 (s), 1516 (s), 1469 (s), 1344 (s), 1261 (s), 1171 (s), 1128 (s), 1091 (s), 1016 (s), 836 (s), 779 (s) cm^{-1}

^{1}H NMR (400 MHz, CD_2Cl_2) δ: 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 NMR (100 MHz, CD_2Cl_2) δ: 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): calculated for $C_{35}H_{46}N_7O_9Si^+$: m/z = 736.3121 $[M+H]^+$; found: m/z = 736.3118 $[M+H]^+$.

Compound 26f



Yield: 78%

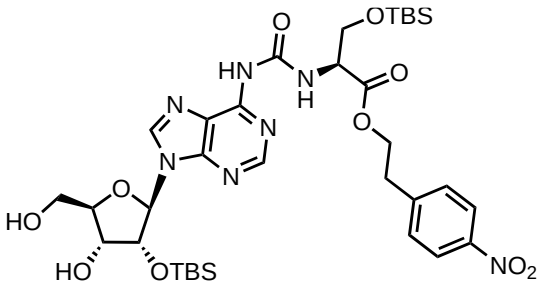
IR: ṽ = 3229 (w), 2930 (w), 2858 (w), 1740 w), 1697 (vs), 1612 (s), 1519 (vs), 1470 (s), 1345 (vs), 1254 (s), 1216 (w), 838 (vs), 781 (s), 736 (vs), 700 (s) cm^{-1}

^{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 26S



Yield: 96%

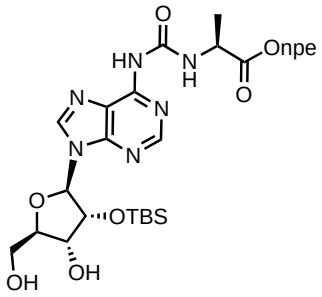
IR: ṽ = 3240 (w), 2950 (w), 2927 (w), 2856 (w), 1737 (w), 1695 (s), 1611 (s), 1589 (s), 1520 (vs), 1469 (s), 1346 (vs), 1313 (w), 1250 (vs), 1130 (w), 1093 (s), 835 (vs), 780 (vs) cm^{-1}

^{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 $[M+Na]^+$.

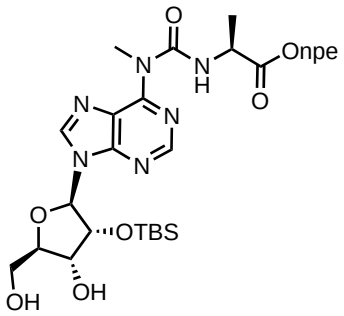
Compound 26A



Yield: 74%
IR: $\tilde{\nu}$ (cm^{-1}) = 2858 (w), 1743 (m), 1698 (s), 1613 (s), 1518 (m), 1470 (m), 1345 (s), 1253 (m), 1212 (w) , 1149 (w), 1064 (w), 837 (m), 780 (s), 696 (m), 494 (s) cm^{-1}

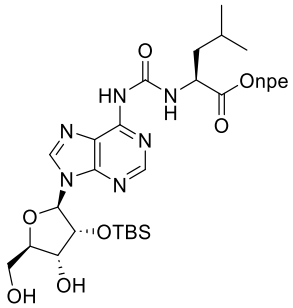
^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.83 (d, *J* = 7.1 Hz, 1H), 8.56 (s, 1H), 8.35 (s, 1H), 8.21 – 8.12 (m, 2H), 8.08 (s, 1H), 7.46 – 7.34 (m, 2H), 5.88 – 5.79 (m, 2H), 5.09 (dd, *J* = 7.3, 4.8 Hz, 1H), 4.62 (p, J = 7.2 Hz, 1H), 4.45 (td, *J* = 6.7, 1.0 Hz, 2H), 4.41 – 4.34 (m, 2H), 4.02 – 3.90 (m, 1H), 3.82 – 3.72 (m, 1H), 3.11 (t, *J* = 6.7 Hz, 2H), 2.85 (s, 1H), 1.51 (d, *J* = 7.2 Hz, 3H), 0.80 (s, 9H), -0.18 (s, 3H), -0.38 (s, 3H)
^{13}C NMR (126 MHz, $CDCl_3$) δ: 173.0, 153.0, 151.1, 150.9, 149.4, 147.1, 145.5, 143.2, 129.9, 123.9, 122.2, 91.4, 87.7, 74.7, 72.9, 64.9, 63.4, 49.3, 35.0, 25.6, 18.5, 18.0, -5.2, -5.3
HRMS (ESI): Calcd. for $C_{28}H_{39}N_7O_9SiNa^+$: *m/z* = 668.2476 $[M+Na]^+$, found: *m/z* = 668.2463 $[M+Na]^+$.

Compound 26A-m



Yield: 95%
IR: $\tilde{\nu}$ = 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) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{29}H_{42}N_7O_9Si^+$: m/z = 660.2808 $[M+H]^+$; found: m/z = 660.2807 $[M+H]^+$.

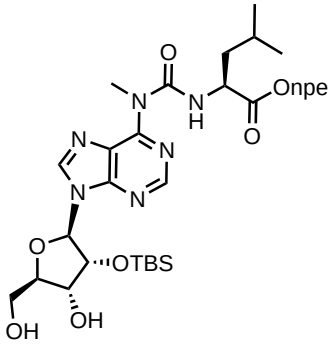
Compound 26L



Yield: 82%
IR: $\tilde{\nu}$ = 2955 (w), 2930 (w), 2858 (w), 1742 (w), 1696 (s), 1614 (m), 1589 (m), 1519 (s), 1470 (m), 1345 (s), 1253 (m), 1148 (m), 1089 (m), 1052 (m), 1005 (w), 956 (w), 915 (w), 856 (m), 836 (s), 780 (s), 746 (w), 697 (w), 647 (m), 627 (w) cm^{-1}

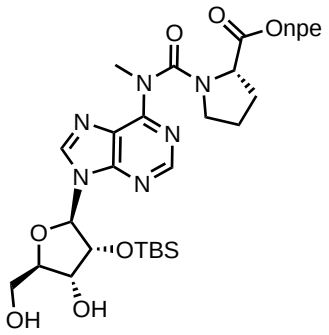
^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.75 (d, *J* = 7.5 Hz, 1H), 8.55 (s, 1H), 8.31 (s, 1H), 8.18 – 8.11 (m, 2H), 8.05 (s, 1H), 7.45 – 7.37 (m, 2H), 5.88 (dd, *J* = 11.9, 2.1 Hz, 1H), 5.83 (d, *J* = 7.3 Hz, 1H), 5.09 (dd, *J* = 7.3, 4.7 Hz, 1H), 4.60 (td, *J* = 7.9, 5.8 Hz, 1H), 4.49 – 4.37 (m, 2H), 4.39 – 4.34 (m, 2H), 3.95 (dt, *J* = 12.8, 2.0 Hz, 1H), 3.82 – 3.72 (m, 1H), 3.18 – 3.04 (m, 2H), 2.86 (s, 1H), 1.77 – 1.61 (m, 3H), 0.99 – 0.90 (m, 6H), 0.80 (s, 9H), -0.17 (s, 3H), -0.38 (s, 3H)
^{13}C NMR (126 MHz, $CDCl_3$) δ: 173.1, 153.3, 151.0, 151.0, 150.0, 149.3, 147.1, 145.5, 143.1, 130.0, 123.9, 122.1, 91.4, 87.7, 74.7, 72.9, 64.8, 63.4, 52.2, 41.3, 35.0, 26.8, 25.6, 25.2, 23.0, 22.1, 18.0, -5.2, -5.3
HRMS (ESI): Calcd. for $C_{31}H_{46}N_7O_9Si^+$: *m/z* = 688.3126 $[M+H]^+$, found: *m/z*: 688.3113 $[M+H]^+$.

Compound 26L-m



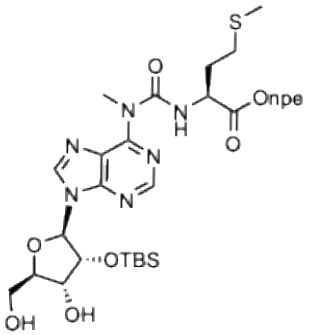
Yield: 98%
IR: $\tilde{\nu}$ = 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (s), 1469 (s), 1345 (s), 1313 (w), 1250 (s), 1129 (w), 1093 (s), 835 (s), 760 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.86 (d, *J* = 7.1 Hz, 1H), 8.63 (s, 1H), 8.58 (s, 1H), 8.13 (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 NMR (100 MHz, Acetone-d_6,) δ: 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): calculated for $C_{32}H_{48}N_7O_9Si^+$: m/z = 702.3277 $[M+H]^+$; found: m/z = 702.3279 $[M+H]^+$.

Compound 26P-m



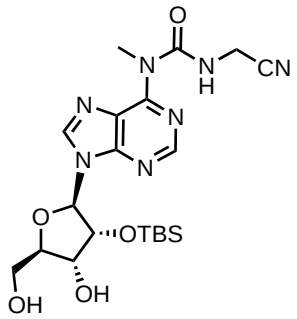
Yield: 82%;
IR: $\tilde{\nu}$ = 2929 (w), 1743 (w), 1679 (m), 1585 (s), 1519 (m), 1391 (m), 1344 (s), 1090 (m), 1046 (m), 780 (m) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.47 (s, 1H), 8.17- 8.14 (m, 2H), 7.81 (s, 1H), 7.39 (br s, 2H), 6.35 (d, *J* = 12.1 Hz, 1H), 5.76 (d, *J* = 7.4 Hz, 1H), 5.15 (dd, *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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{31}H_{44}N_7O_9Si^+$: m/z = 686.2964 $[M+H]^+$; found: m/z = 686.2963 $[M+H]^+$.

Compound 26M-m



Yield: 96%
^{1}H NMR (400 MHz, Acetone-d_6) δ: 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 NMR (100 MHz, acetone-d_6) δ: 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): calculated for $C_{31}H_{46}N_7O_9SSi^+$: m/z = 720.2841 $[M+H]^+$; found: m/z = 720.2833 $[M+H]^+$.

Compound 26Gcn-m



Yield: 92%

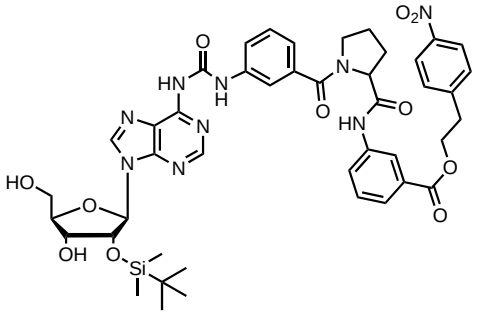
IR: ṽ = 3347 (w), 2929 (m), 2857 (m), 1731 (w), 1681 (s), 1570 (s), 1515 (s), 1462 (s), 1422 (m), 1360 (w), 1329 (w), 1262 (s), 1217 (m), 1126 (s), 1035 (s), 994 (m), 901 (m), 866 (m), 835 (s), 779 (s) cm^{-1}

^{1}H NMR (600 MHz, $CDCl_3$) δ: 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 NMR (150 MHz, $CDCl_3$) δ: 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): calculated for $C_{20}H_{32}N_7O_5Si^+$: m/z = 478.2229 $[M+H]^+$; found: m/z = 478.2231 $[M+H]^+$.

Compound 26AbAb



Yield: 68%

IR: ṽ = 3253 (w), 3080 (w), 2952 (w), 2930 (w), 2857 (w), 1700 (s), 1588 (vs), 1558 (vs), 1518 (vs), 1437 (s), 1345 (vs), 1286 (s), 1240 (vs), 1180 (w), 1108 (w), 1000 (w), 858 (w), 749 (vs), 698 (s) cm^{-1}

^{1}H NMR (500 MHz, CD_2Cl_2) δ: 11.90 (s, 1H), 9.71 (s, 1H), 8.95 (s, 1H), 8.65 (s, 1H), 8.26 (s, 1H), 8.22 (s, 1H), 8.19 – 8.14 (m, 2H), 8.00 (s, 1H), 7.72 (d, *J* = 8.2 Hz, 1H), 7.62 (d, *J* = 7.7 Hz, 1H), 7.49 (d, *J* = 8.6 Hz, 2H), 7.44 (t, *J* = 7.9 Hz, 1H), 5.91 (d, *J* = 7.1 Hz, 1H), 5.54 (d, *J* = 11.3 Hz, 1H), 5.08 (dd, *J* = 7.1, 4.7 Hz, 1H), 4.92 (dd, *J* = 7.5, 4.7 Hz, 1H), 4.52 (t, *J* = 6.5 Hz, 2H), 4.37 (d, *J* = 4.8 Hz, 1H), 4.35 (s, 1H), 3.94 (dd, *J* = 12.8, 1.9 Hz, 1H), 3.76 (t, *J* = 11.5 Hz, 1H), 3.69 – 3.58 (m, 2H), 3.18 (t, *J* = 6.6 Hz, 2H), 2.48 (q, *J* = 5.5 Hz, 1H), 2.12 (dq, *J* = 13.2, 6.9 Hz, 2H), 1.95 – 1.87 (m, 1H), 0.80 (s, 9H), -0.16 (s, 3H), -0.34 (s, 3H)

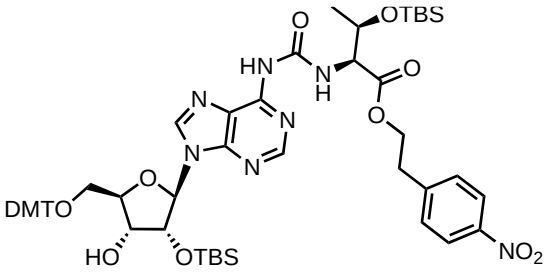
^{13}C NMR (126 MHz, $CDCl_3$) δ: 171.41, 169.89, 166.26, 151.59, 151.10, 150.97, 150.15, 149.99, 147.28, 146.58, 143.87, 139.30, 138.67, 137.36, 136.24, 130.97, 130.44, 129.46, 129.28, 124.97, 124.43, 124.08, 122.91, 122.42, 120.76, 119.27, 91.43, 88.00, 75.23, 73.06, 64.92, 63.48, 61.55, 51.13, 35.40, 27.82, 25.80, 25.67, 18.12, -5.21, -5.30

HRMS (ESI): calculated for $C_{44}H_{52}N_9O_{11}Si^+$: m/z = 910.3550 $[M+H]^+$; found: m/z = 910.3549 $[M+H]^+$.

DMT protection of 5'OH of (m^6)aa^6A derivatives

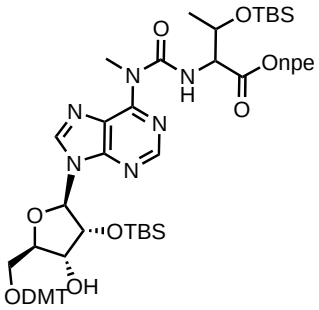
The 3'-5'-deprotected adenosine derivative (1 eq.) was dissolved in pyridine under N_2
atmosphere. DMT chloride (1.2 eq.) was added in two portions and the mixture was stirred at
room temperature overnight. Then the volatiles were evaporated, and crude product was purified by flash chromatography eluting 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 a white to pale yellow foam.

Compound 27T



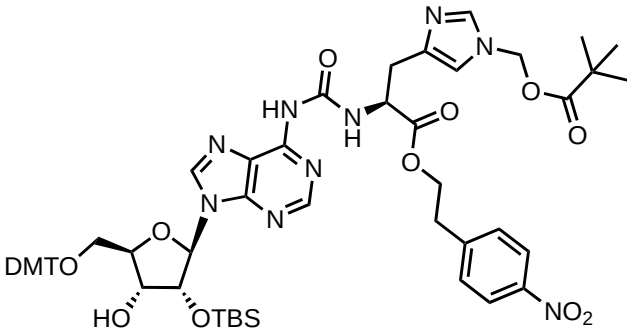
Yield: 90%
IR: $\tilde{\nu}$ = 3350 (w), 2930 (w), 2856 (w), 1729 (w), 1684 (s), 1608 (s), 1521 (vs), 1464 (s), 1345 (vs), 1248 (vs), 1174 (w), 1094 (w), 10033 (s), 827 (vs), 777 (vs) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.00 (d, *J* = 9.0 Hz, 1H), 8.95 (s, 1H), 8.55 (s, 1H), 8.40 (s, 1H), 7.98 (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 27T-m



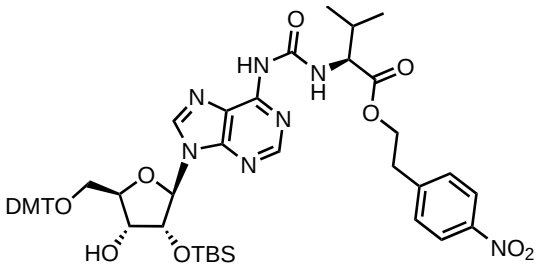
Yield: 68%
IR: $\tilde{\nu}$ = 2908 (w), 1757 (w), 1718 (w), 1670 (w), 1608 (w), 1507 (s), 1441 (w), 1294 (w), 1248 (s), 1177 (s), 1090 (s), 1034 (s), 975 (s), 913 (s), 869 (s), 776 (s), 703 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.89 (d, *J* = 8.7 Hz, 1H), 8.50 (s, 1H), 8.42 (s, 1H), 8.03 (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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{57}H_{76}N_7O_{12}Si_2^+$: m/z = 1106.5085 $[M+H]^+$; found: m/z = 1106.5103 $[M+H]^+$.

Compound 27H



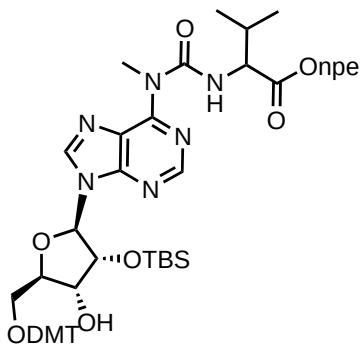
Eluent: 10% CH_2Cl_2 in EtOAc to pure EtOAc containing 0.1% of pyridine.
Yield: 85%
IR: $\tilde{\nu}$ = 3854 (w), 3746 (w), 3650 (w), 3630 (w), 2931 (w), 2361 (w), 2341 (w), 1740 (s), 1700 (vs), 1654 (w), 1609 (s), 1587 (w), 1559 (w), 1508 (vs), 1472 (s), 1396 (w), 1345 (s), 1300 (w), 1250 (vs), 1176 (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 27V



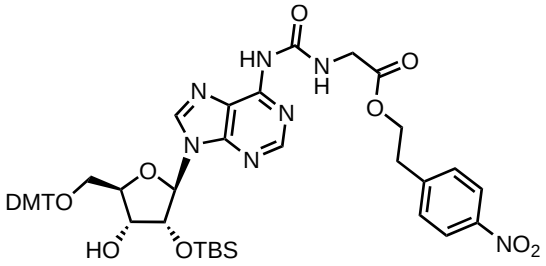
Yield: 70%
IR: $\tilde{\nu}$ = 2929 (w), 1735 (w), 1684 (w), 1569 (s), 1508 (vs), 1464 (s), 1344 (s), 1249 (vs), 1176 (w), 1015 (s), 800 (vs) 749 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.02 (d, *J* = 9.0 Hz, 1H), 8.55 (s, 1H), 8.51 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 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 27V-m



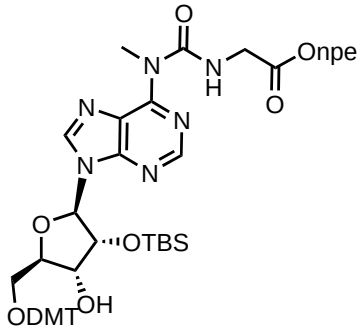
Yield: 75%;
IR: ṽ = 2950 (w), 2850 (w), 1730 (w), 1670 (w), 1607 (m), 1577 (s), 1508 (s), 1464 (w), 1347 (s), 1250 (s), 1177 (s), 1150 (w), 1090 (s), 1035 (m), 981 (w), 913 (s), 866 (s), 839 (s), 701 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 11.03 (d, *J* = 7.7 Hz, 1H), 8.52 (s, 1H), 8.49 (s, 1H), 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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{52}H_{64}N_7O_{11}Si^+$: m/z = 990.4428 $[M+H]^+$; found: m/z = 990.4430 $[M+H]^+$.

Compound 27G



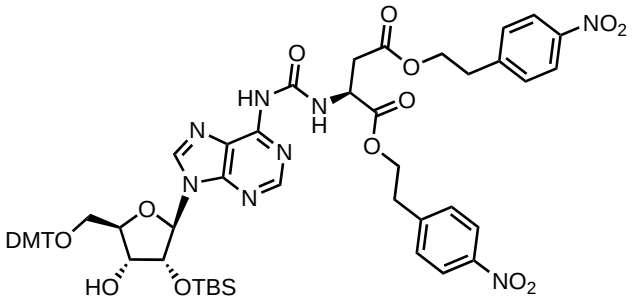
Yield: 85%
IR: ṽ = 2931 (w), 1748 (w), 1703 (s), 1609 (s), 1588 (w), 1509 (s), 1469 (s), 1345 (vs), 1250 (vs), 1177 (vs), 1035 (w), 835 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 9.93 (t, *J* = 5.5 Hz, 1H), 9.05 (br. s, 1H), 8.58 (s, 1H), 8.46 (s, 1H), 8.11 (d, *J* = 8.7 Hz, 2H), 7.58 (d, *J* = 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 C48H55N7O11Si+: m/z = 934.3802 $[M+H]^+$; found: m/z = 934.3812 $[M+H]^+$.

Compound 27G-m



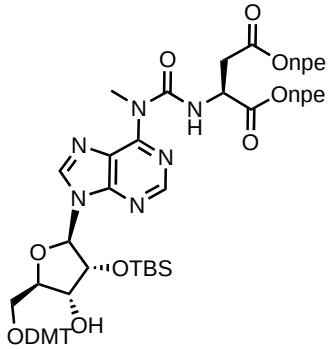
Yield: 86%
IR: ṽ = 3320 (w), 2929 (w), 2853 (w), 1749 (w), 1681 (m), 1606 (w), 1568 (m), 1510 (s), 1466 (m), 1345 (s), 1300 (w), 1250 (s), 1213 (m), 1176 (s), 1066 (w), 1034 (s), 1005 (w), 916 (w), 856 (m), 834 (s), 782 (m), 699 (m) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.85 (t, *J* = 5.6 Hz, 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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{49}H_{58}N_7O_{11}Si^+$: m/z = 948.3958 $[M+H]^+$; found: m/z = 948.3949 $[M+H]^+$.

Compound 27D



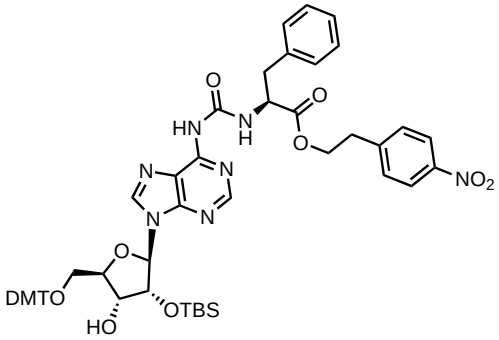
Yield: 95%
IR: ṽ = 2930 (w), 1734 (s), 1698 (s), 1607 (s), 1509 (vs), 1466 (s), 1370 (vs), 1248 (s), 1174 (s), 1032 (s), 829 (vs), 699 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.29 (d, *J* = 7.9 Hz, 1H), 9.18 (s, 1H), 8.60 (s, 1H), 8.42 (s, 1H), 8.10 – 8.04 (m, 4H), 7.58 – 7.51 (m, 4H), 7.39 – 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 27D-m



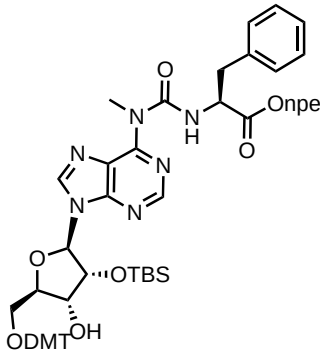
Yield: 80%
IR: ṽ =2932 (w), 1737 (m), 1682 (m), 1571 (m), 1518 (s), 1464 (s), 1251 (s), 1176 (s), 1018 (m), 836 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 11.21 (d, *J* = 7.5 Hz, 1H), 8.49 (s, 1H), 8.39 (s, 1H), 8.09-8.04 (m, 4H), 7.55-7.49 (m, 6H), 7.39-7.35 (m, 4H), 7.29 (dd, *J* = 7.5, 7.5 Hz, 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 NMR (101 MHz, Acetone-d_6) δ: 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): calculated for $C_{59}H_{67}O_{15}N_8Si^+$: m/z = 1155.4489 $[M+H]^+$; found: m/z = 1155.4504 $[M+H]^+$.

Compound 27F



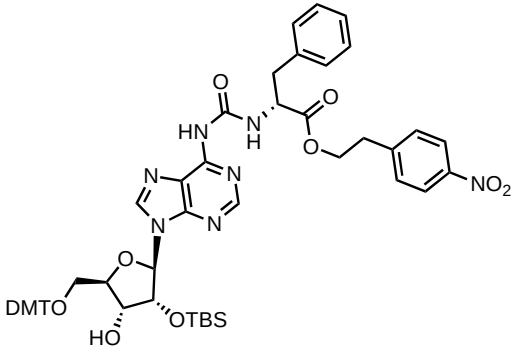
Eluent: CH_2Cl_2/MeOH (20/1, v/v).
Yield: 82%
IR: ṽ = 3245 (w), 2952 (w), 2931 (w), 1741 (w), 1699 (s), 1608 (s), 1587 (w), 1519 (s), 1509 (s), 1469 (w), 1345 (s), 1249 (s), 1176 (s), 1034 (w), 834 (s), 783 (w), 735 (vs), 701 (vs) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 9.83 (d, *J* = 7.5 Hz, 1H), 8.29 (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 27F-m



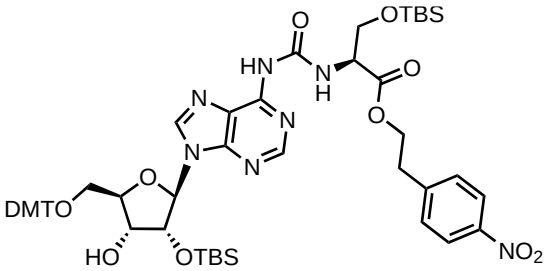
Yield: 90%
IR: ṽ = 3538 (w), 2953 (w), 2855 (w), 1738 (w), 1681 (w), 1568 (s), 1508 (s), 1463 (s), 1344 (s), 1249 (s), 1174 (s), 1031 (s), 1016 (s), 834 (s), 781 (s) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 10.89 (d, *J* = 6.8 Hz, 1H), 8.22 (s, 1H), 8.15 (s, 1H), 8.07 (d, *J* = 8.7 Hz, 2H), 7.50-7.45 (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 NMR (100 MHz, CD_2Cl_2) δ: 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): calculated for $C_{56}H_{64}N_7O_{11}Si^+$: m/z = 1038.4428 $[M+H]^+$; found: m/z = 1038.4447 $[M+H]^+$.

Compound 27f



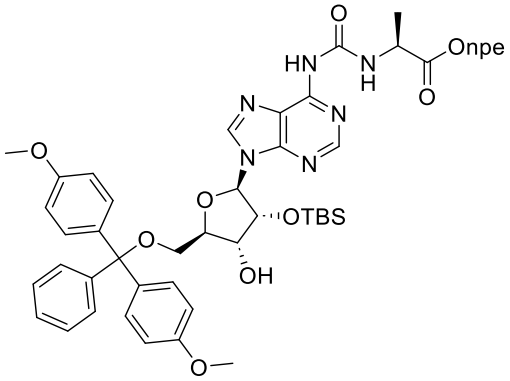
Eluent: CH_2Cl_2/MeOH (20/1, v/v).
Yield: 85%
IR: ṽ = 3231 (w), 2930 (w), 2858 (w), 1741 (w), 1697 (s), 1612 (s), 1587 (w), 1518 (vs), 1470 (s), 1344 (vs), 1253 (w), 1180 (w), 1134 (w), 1085 (w), 836 (s), 780 (s), 734 (vs), 698 (vs) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 9.83 (d, *J* = 7.5 Hz, 1H), 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 27S



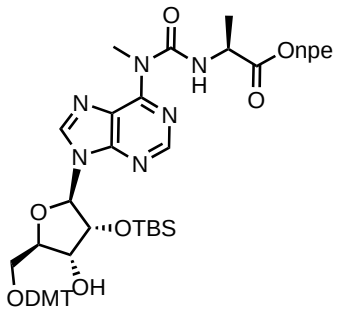
Yield: 90%
IR: $\tilde{\nu}$ = 2930 (w), 1740 (s), 1694 (s), 1607 (w), 1582 (w), 1519 (vs), 1465 (w), 1345 (s), 1249 (s), 1176 (w), 1110 (w), 1034 (w), 820 (vs), 778 (s), 699 (w) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.13 (d, *J* = 8.2 Hz, 1H), 8.99 (s, 1H), 8.58 (s, 1H), 8.44 (s, 1H), 8.09 (d, *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 $[M+Na]^+$.

Compound 27A



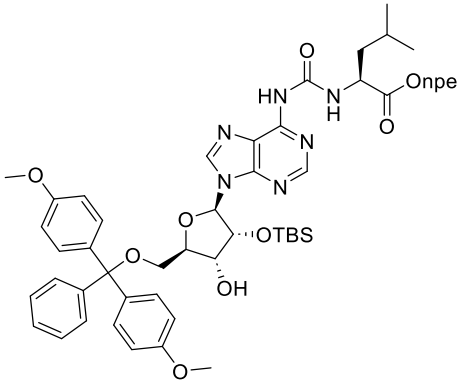
Yield: 88%
IR: $\tilde{\nu}$ = 2856 (w), 1743 (w), 1700 (m), 1608 (m), 1586 (m), 1508 (s), 1468 (m), 1345 (s), 1301 (m), 1248 (s), 1174 (s), 1148 (m), 1112 (m), 1064 (m), 1033 (s), 912 (w), 781 (s), 748 (m), 727 (w), 701 (s), 646 (w) 582 (s), 478 (w), 449 (w) cm^{-1}
^{1}H NMR (500 MHz, CD_2Cl_2) δ: 9.79 (d, J = 7.2 Hz, 1H), 8.44 (s, 1H), 8.17 (s, 1H), 8.13 – 8.06 (m, 2H), 7.91 (s, 1H), 7.49 – 7.39 (m, 4H), 7.37 – 7.31 (m, 4H), 7.32 – 7.24 (m, 2H), 7.26 – 7.19 (m, 1H), 6.86 – 6.78 (m, 4H), 6.04 (d, *J* = 5.0 Hz, 1H), 5.01 (t, *J* = 5.0 Hz, 1H), 4.55 (p, *J* = 7.2 Hz, 1H), 4.49 – 4.34 (m, 3H), 4.24 (q, *J* = 3.8 Hz, 1H), 3.77 (s, 6H), 3.48 (dd, *J* = 10.6, 3.2 Hz, 1H), 3.39 (dd, *J* = 10.7, 4.3 Hz, 1H), 3.09 (t, *J* = 6.5 Hz, 2H), 2.66 (d, *J* = 4.7 Hz, 1H), 1.47 (d, *J* = 7.2 Hz, 3H), 0.85 (s, 9H), 0.01 (s, 3H), -0.12 (s, 3H)
^{13}C NMR (126 MHz, CD_2Cl_2) δ: 173.1, 159.1, 153.4, 151.5, 150.7, 150.5, 150.1, 147.3, 146.2, 145.3, 142.2, 136.3, 136.1, 130.5, 130.5, 130.3, 128.5, 128.3, 127.3, 123.9, 121.3, 113.5, 89.2, 87.0, 84.6, 75.9, 71.8, 65.0, 63.8, 55.6, 49.6, 35.2, 25.8, 18.4, 18.2, -4.8, -5.0.
HRMS (ESI): Calcd. for $C_{49}H_{57}N_7O_{11}SiNa^+$: *m/z* = 970.3783 $[M+Na]^+$, found: *m/z* = 970.3773 $[M+Na]^+$.

Compound 27A-m



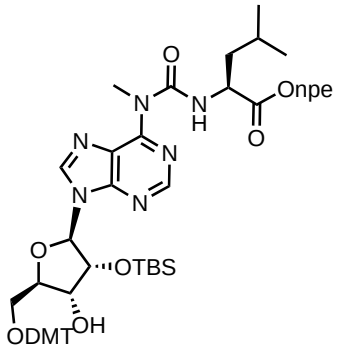
Yield: 85%
IR: $\tilde{\nu}$ = 2928 (w), 1741 (w), 1681 (w), 1610 (m), 1568 (m), 1508 (s), 1463 (m), 1344 (m), 1251 (s), 1174 (m), 1018 (m), 835 (s), 781 (m) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 10.91 (d, *J* = 6.5 Hz, 1H), 8.44 (s, 1H), 8.17 (s, 1H), 8.10 (d, *J* = 8.7 Hz, 2H), 7.49-7.44 (m, 2H), 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 NMR (100 MHz, CD_2Cl_2) δ: 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): calculated for $C_{50}H_{60}N_7O_{11}Si^+$: m/z = 962.4115 $[M+H]^+$; found: m/z = 962.4128 $[M+H]^+$.

Compound 27L



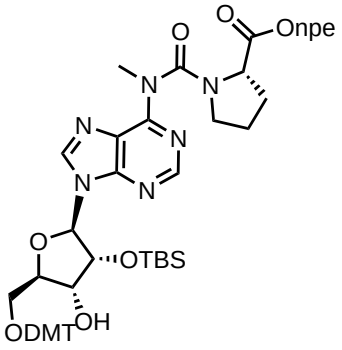
Yield: 75%
IR: $\tilde{\nu}$ = 2955 (w), 2930 (w), 2857 (w), 1742 (w), 1700 (s), 1608 (m), 1587 (w), 1508 (s), 1470 (m), 1445 (w), 1345 (s), 1301 (w), 1248 (s), 1176 (m), 1150 (w), 1112 (w), 1065 (w), 1033 (m), 913 (w), 834 (s), 782 (m), 750 (w), 700 (m), 647 (w), 582 (m) cm^{-1}
^{1}H NMR (500 MHz, CD_2Cl_2) δ: 9.76 (d, *J* = 7.5 Hz, 1H), 8.43 (s, 1H), 8.20 (s, 1H), 8.11 (s, 1H), 8.11 – 8.05 (m, 2H), 7.50 – 7.38 (m, 4H), 7.35 (dq, *J* = 8.8, 1.9 Hz, 4H), 7.32 – 7.23 (m, 2H), 7.26 – 7.18 (m, 1H), 6.85 – 6.78 (m, 4H), 6.04 (d, *J* = 4.9 Hz, 1H), 5.03 (t, *J* = 5.0 Hz, 1H), 4.54 (ddd, *J* = 8.6, 7.5, 5.8 Hz, 1H), 4.48 – 4.34 (m, 3H), 4.25 (q, *J* = 3.8 Hz, 1H), 3.77 (d, *J* = 0.8 Hz, 6H), 3.48 (dd, *J* = 10.7, 3.2 Hz, 1H), 3.40 (dd, *J* = 10.7, 4.3 Hz, 1H), 3.09 (t, *J* = 6.5 Hz, 2H), 2.71 (d, *J* = 4.8 Hz, 1H), 1.79 – 1.60 (m, 3H), 0.95 (s, 3H), 0.93 (s, 3H), 0.85 (s, 9H), 0.01 (s, 3H), -0.10 (s, 3H)
^{13}C NMR (126 MHz, CD_2Cl_2) δ: 173.1, 159.1, 153.7, 151.4, 150.7, 150.5, 150.2, 147.2, 146.3, 145.2, 142.3, 136.2, 136.0, 136.0, 130.5, 130.3, 128.5, 128.3, 127.2, 123.9, 113.5, 89.2, 87.0, 84.5, 75.8, 71.7, 64.9, 63.8, 55.6, 52.4, 41.4, 35.2, 29.5, 25.7, 25.4, 23.0, 22.0, 18.2, -4.8, -5.0
HRMS (ESI): Calcd. for $C_{52}H_{64}N_7O_{11}Si^+$: *m/z* = 990.4433 $[M+H]^+$, found: *m/z* = 990.4417 $[M+H]^+$.

Compound 27L-m



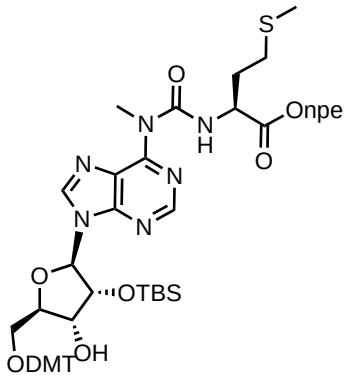
Yield: 72%
IR: ṽ = 2950 (w), 2852 (w), 1729 (w), 1670 (w), 1607 (s), 1577 (s), 1508 (s), 1464 (w), 1347 (s), 1250 (s), 1177 (s), 1152 (w), 1091 (s), 1035 (s), 981 (w), 913 (s), 866 (s), 839 (s), 699 (s) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.89 (d, *J* = 7.1 Hz, 1H), 8.50 (s, 1H), 8.49 (s, 1H), 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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{53}H_{66}N_7O_{11}Si^+$: m/z = 1004.4584 $[M+H]^+$; found: m/z = 1004.4579 $[M+H]^+$.

Compound 27P-m



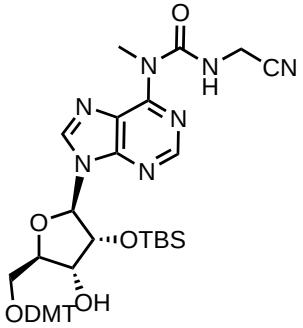
Yield: 80%
IR: ṽ = 2930 (w), 1743 (w), 1680 (m), 1582 (s), 1509 (m), 1391 (m), 1345 (s), 1249 (s), 1174 (s), 782 (m) cm^{-1}
^{1}H NMR (400 MHz, Acetone-d_6) δ: 8.37 (s, 1H), 8.33 (s, 1H), 8.18-8.16 (m, 2H), 7.56 (br s, 2H), 7.49 (d, *J* = 7.3 Hz, 2H), 7.37-7.35 (m, 4H), 7.28 (dd, *J* = 7.3, 7.3 Hz, 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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{52}H_{62}N_7O_{11}Si^+$: m/z = 988.4270 $[M+H]^+$; found: m/z = 988.4280 $[M+H]^+$.

Compound 27M-m



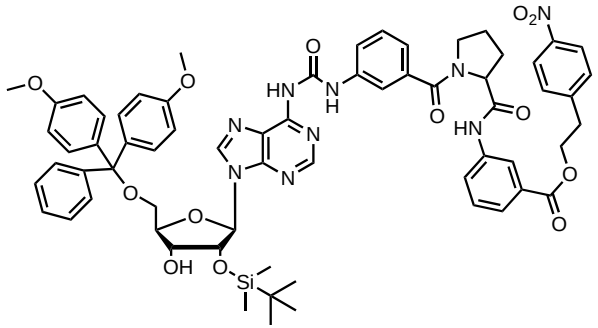
Yield: 92%
^{1}H NMR (400 MHz, Acetone-d_6) δ: 10.98 (d, *J* = 7.1 Hz, 1H), 8.50 (s, 1H), 8.49 (s, 1H), 8.14-8.07 (d, *J* = 8.9 Hz, 2H), 7.59 (d, *J* = 8.6 Hz, 2H), 7.50 (d, *J* = 7.2 Hz, 2H), 7.37 (dd, *J* = 9.0, 2.7 Hz, 4H), 7.29 (t, *J* = 7.4 Hz, 2H), 7.22 (t, *J* = 7.2 Hz, 1H), 6.86 (dd, *J* = 8.9, 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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{52}H_{64}N_7O_{11}SSi^+$: m/z = 1022.4148 $[M+H]^+$; found: m/z = 1022.4137 $[M+H]^+$.

Compound 27Gcn-m



Yield: 86%
IR: ṽ = 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) cm^{-1}
^{1}H NMR (600 MHz, $CDCl_3$) δ: 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 NMR (150 MHz, $CDCl_3$) δ: 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): calculated for $C_{41}H_{50}N_7O_7Si^+$: m/z = 780.3535 $[M+H]^+$; found: m/z = 780.3538 $[M+H]^+$.

Compound 27AbPAb

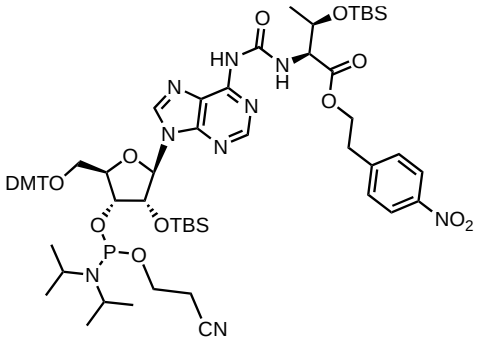


Yield: 75%
IR: $\tilde{\nu}$ = 3199 (w), 3078 (w), 2952 (m), 2930 (m), 2857 (w), 1702 (s), 1587 (vs), 1560 (vs), 1437 (m), 1345 (s), 1245 (vs), 1176 (s), 1109 (w), 1000 (w), 905 (w), 835 (s), 750 (vs), 700 (s) cm^{-1}
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 11.97 (s, 1H), 9.73 (s, 1H), 8.55 (s, 1H), 8.44 (s, 1H), 8.26 (s, 1H), 8.22 (s, 1H), 8.17 (d, *J* = 8.7 Hz, 2H), 7.96 (s, 1H), 7.75 – 7.66 (m, 2H), 7.64 (d, *J* = 8.2 Hz, 1H), 7.52 – 7.41 (m, 5H), 7.35 (d, *J* = 8.8 Hz, 4H), 7.33 – 7.29 (m, 2H), 7.27 (d, *J* = 7.7 Hz, 2H), 7.22 (t, *J* = 7.2 Hz, 1H), 6.84 – 6.79 (m, 4H), 6.06 (d, *J* = 5.0 Hz, 1H), 5.03 (t, *J* = 5.0 Hz, 1H), 4.94 (dd, *J* = 7.5, 4.7 Hz, 1H), 4.52 (t, *J* = 6.5 Hz, 2H), 4.39 (d, *J* = 3.9 Hz, 1H), 4.27 (q, *J* = 3.8 Hz, 1H), 3.77 (s, 6H), 3.70 – 3.55 (m, *J* = 6.9 Hz, 2H), 3.50 (dd, *J* = 10.7, 3.2 Hz, 1H), 3.41 (dd, *J* = 10.7, 4.3 Hz, 1H), 3.18 (t, *J* = 6.5 Hz, 2H), 2.72 (d, *J* = 4.5 Hz, 1H), 2.54 (q, *J* = 8.0 Hz, 1H), 2.12 (tt, *J* = 12.1, 7.2 Hz, 2H), 1.91 (t, *J* = 6.1 Hz, 1H), 0.86 (s, 9H), 0.02 (s, 3H), -0.10 (s, 3H).
^{13}C NMR (101 MHz, CD_2Cl_2) δ: 171.59, 169.71, 166.24, 159.11, 151.61, 151.32, 150.89, 150.29, 147.28, 146.57, 145.22, 142.47, 139.28, 138.85, 137.28, 136.02, 136.01, 131.00, 130.50, 130.46, 130.44, 129.46, 129.29, 128.48, 128.27, 124.99, 124.43, 124.08, 122.74, 122.16, 121.37, 120.76, 119.11, 113.52, 89.21, 87.02, 84.65, 75.98, 71.78, 64.91, 63.80, 61.51, 60.63, 55.61, 51.15, 35.40, 27.51, 25.82, 25.74, 21.18, 18.20, 14.39, -4.81, -4.98.
HRMS (ESI): calculated for $C_{65}H_{70}N_9O_{13}Si^+$: m/z = 1212.4875 $[M+H]^+$; found: m/z = 1212.4833 $[M+H]^+$

General synthesis of (m^6)aa^6A nucleoside phosphoramidites

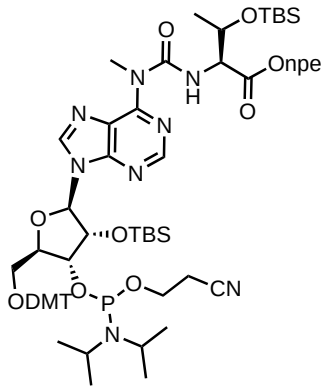
In an oven-dried flask under argon atmosphere, 5'-DMT protected compound (1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 ˚C. Hünig's base (4 eq.) was added dropwise followed by the addition of 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (2.5 eq.). The solution was stirred at room temperature 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 HPLC grade Hex/EtOAc (1/1, v/v) solvent containing 0.1% of pyridine, unless otherwise specified. The phosphoramidite was obtained as a mixture of P-diastereomers as white foam.

Compound 28T



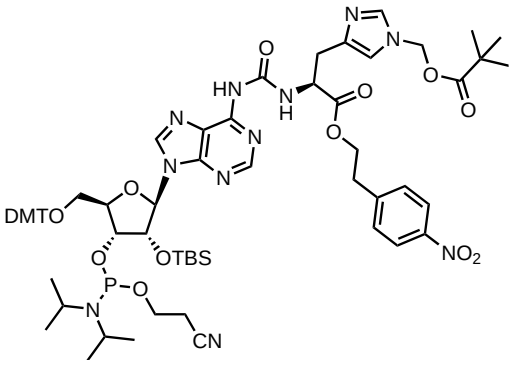
Yield: 85%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.16, 148.45
HRMS (ESI): calculated for $C_{65}H_{91}N_9O_{13}PSi_2^+$: m/z = 1292.6007 $[M+H]^+$; found: m/z = 1292.6033 $[M+H]^+$.

Compound 28T-m



Yield: 62%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.2, 148.5
HRMS (ESI): calculated for $C_{66}H_{93}N_9O_{13}PSi_2^+$: m/z = 1306.6164 $[M+H]^+$; found: m/z = 1306.6189 $[M+H]^+$.

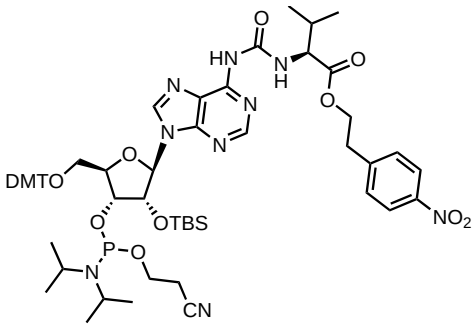
Compound 28H



Eluent: 30% CH2Cl2 in EtOAc containing 0.1% of pyridine.
Yield: 66%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.26, 148.61

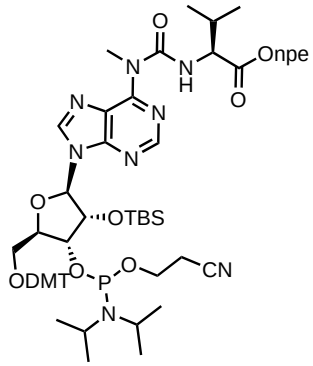
HRMS (ESI): calculated for $C_{67}H_{87}N_{11}O_{14}PSi^+$: m/z = 1328.5935 $[M+H]^+$; found: m/z = 1328.5944 $[M+H]^+$.

Compound 28V



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 $[M+H]^+$.

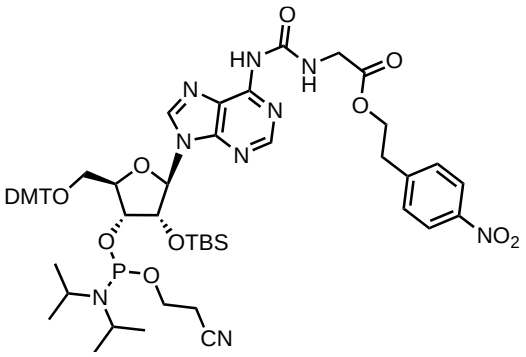
Compound 28V-m



Yield: 77%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.1, 148.7

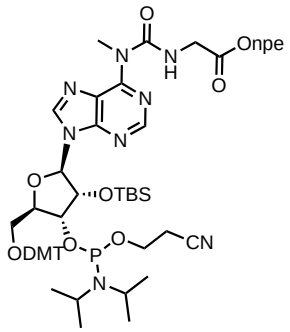
HRMS (ESI): calculated for $C_{61}H_{81}N_9O_{12}PSi^+$: m/z = 1190.5506 $[M+H]^+$; found: m/z = 1190.5492 $[M+H]^+$.

Compound 28G



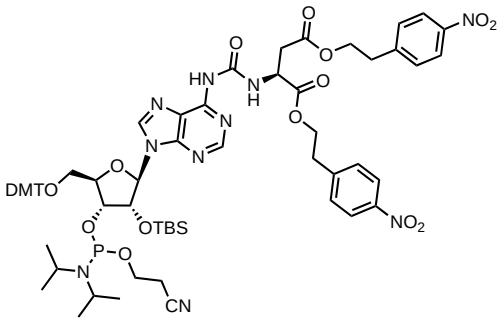
Yield: 86%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.32, 148.60
HRMS (ESI): calculated for $C_{57}H_{72}N_9O_{12}PSi^+$: m/z = 1134.4880 $[M+H]^+$; found: m/z = 1134.4894 $[M+H]^+$.

Compound 28G-m



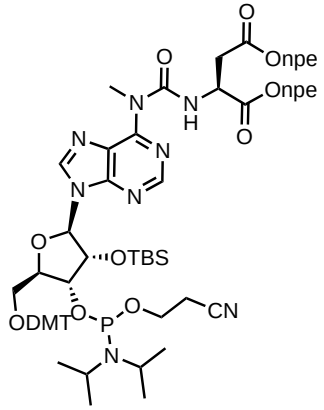
Yield: 85%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.1, 148.7
HRMS (ESI): calculated for $C_{58}H_{75}N_9O_{12}PSi^+$: m/z = 1148.5037 $[M+H]^+$; found: m/z = 1148.5052 $[M+H]^+$.

Compound 28D



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 $[M+H]^+$.

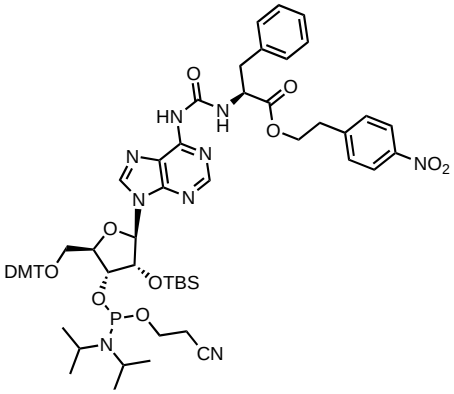
Compound 28D-m



Yield: 65%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.1, 148.7

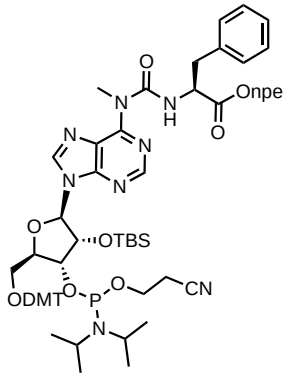
HRMS (ESI): calculated for $C_{68}H_{84}O_{16}N_{10}PSi^+$: m/z = 1355.5567 $[M+H]^+$; found: m/z = 1355.5590 $[M+H]^+$.

Compound 28F



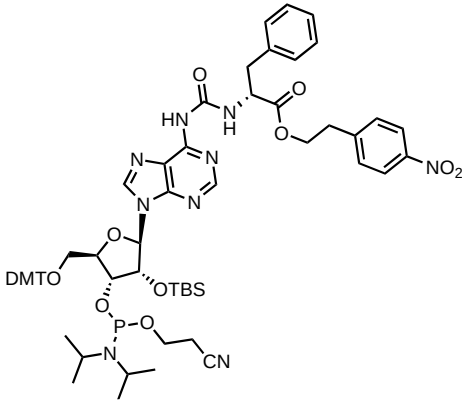
Yield: 67%
^{31}P NMR (162 MHz, CD_2Cl_2) δ: 150.7, 149.1
HRMS (ESI): calculated for $C_{64}H_{79}N_9O_{12}PSi^+$: m/z = 1224.5350 $[M+H]^+$; found: m/z = 1224.5374 $[M+H]^+$.

Compound 28F-m



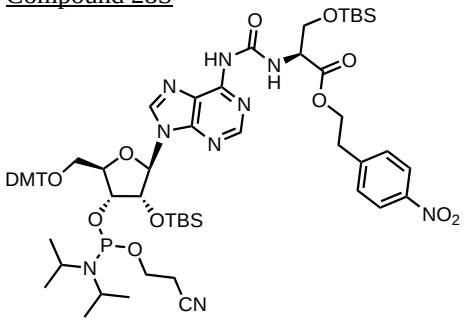
Yield: 90%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.7, 149.1;
HRMS (ESI): calculated for $C_{65}H_{81}N_9O_{12}PSi^+$: m/z = 1238.5506 $[M+H]^+$; found: m/z = 1238.5530 $[M+H]^+$.

Compound 28f



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 $[M+H]^+$.

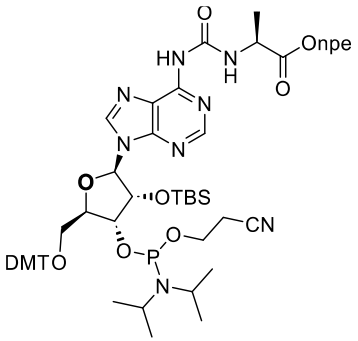
Compound 28S



Yield: 70%

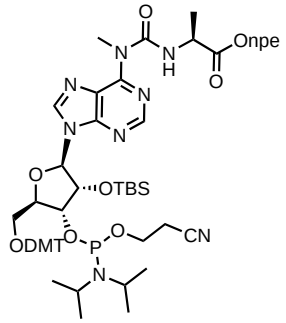
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.34, 148.45
HRMS (ESI): calculated for $C_{64}H_{89}N_9O_{13}PSi_2^+$: m/z = 1278.5856 $[M+H]^+$; found: m/z = 1278.5877 $[M+H]^+$.

Compound 28A



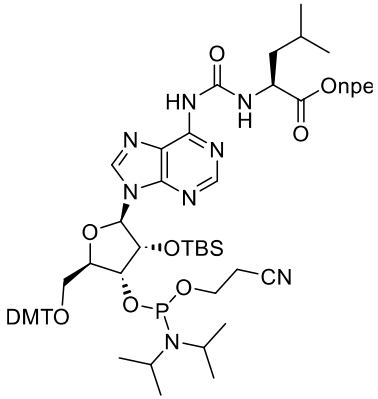
Yield: 74%
^{31}P NMR (202 MHz, CD_2Cl_2) δ: 150.77, 149.17
HRMS (ESI): calcualted for $C_{58}H_{74}N_9O_{12}PSiNa^+$: *m/z* = 1170.4862 $[M+Na]^+$, found: *m/z* = 1170.4836 $[M+Na]^+$.

Compound 28A-m



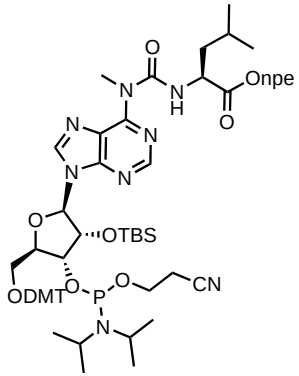
Yield: 85%
^{31}P NMR (162 MHz, CD_2Cl_2) δ: 150.6, 149.2
HRMS (ESI): calculated for $C_{59}H_{77}N_9O_{12}PSi^+$: m/z = 1162.5193 $[M+H]^+$; found: m/z = 1162.5221 $[M+H]^+$.

Compound 28L



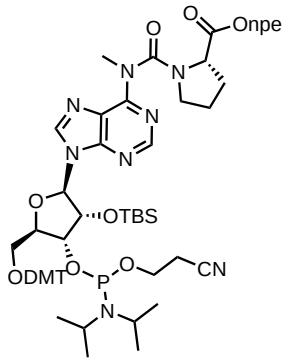
Yield: 85%
^{31}P NMR (202 MHz, CD_2Cl_2) δ: 150.70, 149.15
HRMS (ESI): calculated for $C_{61}H_{81}N_9O_{12}PSi^+$: *m/z* = 1190.5512 $[M+H]^+$, found: *m/z* = 1190.5508 $[M+H]^+$.

Compound28L-m



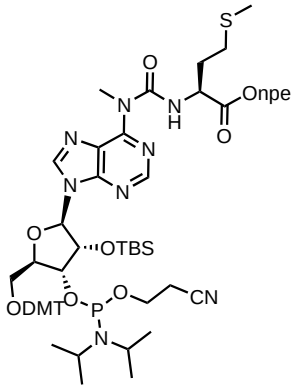
Yield: 75%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.1, 148.7
HRMS (ESI): calculated for $C_{62}H_{83}N_9O_{12}PSi^+$: m/z = 1204.5663 $[M+H]^+$; found: m/z = 1204.5682 $[M+H]^+$.

Compound 28P-m



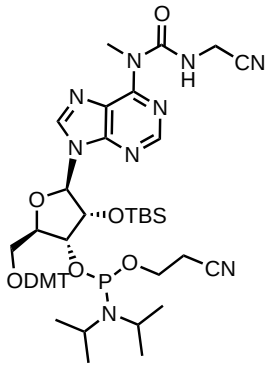
Yield: 80%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.2, 148.6
HRMS (ESI): calculated for $C_{61}H_{79}N_9O_{12}PSi^+$: m/z = 1188.5350 $[M+H]^+$; found: m/z 1188.5388 $[M+H]^+$.

Compound 28M-m



Yield: 89%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.1, 148.6
HRMS (ESI): calculated for $C_{61}H_{81}N_9O_{12}PSSi^+$: m/z = 1222.5227 $[M+H]^+$; found: m/z = 1222.5215 $[M+H]^+$.

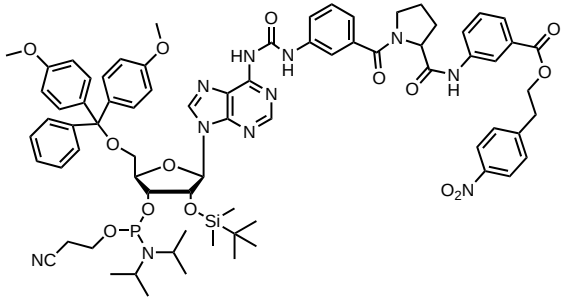
Compound 28Gcn-m



Yield: 89%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.3, 148.6

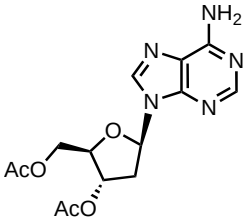
HRMS (ESI): calculated for $C_{50}H_{67}N_9O_8PSi^+$: m/z = 980.4614 $[M+H]^+$; found: m/z = 980.4611 $[M+H]^+$.

Compound 28AbPAb



Yield: 83%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.8, 149.2
HRMS (ESI): calculated for $C_{74}H_{85}N_{11}O_{14}PSi^-$: m/z = 1410.5790 $[M-H]^-$; found: m/z = 1410.5816 $[M-H]^-$.

Compound 30



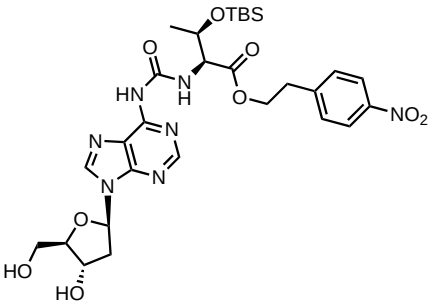
Acetic anhydride (1.1 ml, 11.5 mmol, 6.2 eq.) was added to a mixture of deoxyadenosine monohydrate 53 (0.5 g, 1.85 mmol, 1 eq.), pyridine (7 ml) and 4-(dimethylamino)pyridine (25 mg, 0.4 mmol, 0.1 eq.). The reaction mixture was stirred at room temperature for 4 h. Subsequently, iced water was added, 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.[315]

Compound 31T



The acetyl-protected deoxyadenosine derivative **30** (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 r.t. for 2 h (the solution in time becomes clear). Afterwards the protected threonine derivative **7** (1.1 g, 3 mmol, 2 eq.) was added together with NEt_3 (415 µL, 3 mmol, 2 eq.) as a solution in CH_2Cl_2 and the resulting solution was stirred overnight at r.t. 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 h at r.t., the reaction mixture was evaporated. The residue was purified via flash chromatography eluting with CH_2Cl_2/MeOH (10/1, v/v).

Yield: 65%

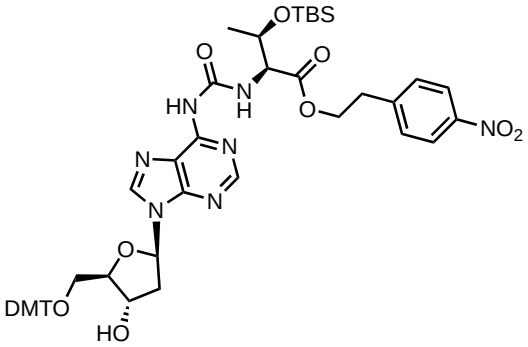
IR: $\tilde{\nu}$ = 3227 (w), 2929 (w), 2855 (w), 1734 (w), 1686 (vs), 1610 (s), 1587 (s), 1518 (vs), 1465 (vs), 1344 (vs), 1312 (w), 1248 (s), 1213 (w), 1094 (vs), 939 (s), 827 (vs) cm^{-1}

^{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 32T



The 3'-5'-deprotected adenosine derivative **31T** (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 r.t. 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%

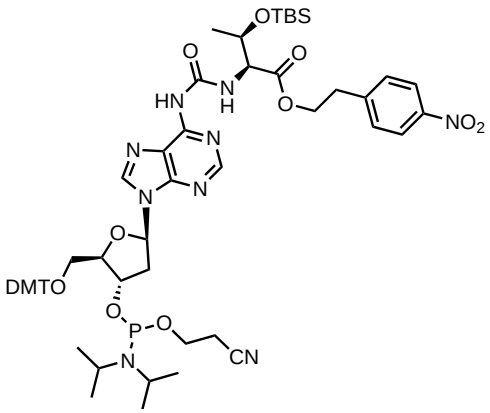
IR: $\tilde{\nu}$ = 2960 (w), 2930 (w), 1734 (w), 1696 (s), 1609 (s), 1586 (w), 1520 (vs), 1509 (vs), 1467 (s), 1345 (vs), 1304 (w), 1250 (vs), 1175 (s), 1095 (w), 1034 (s), 940 (w), 828 (vs), 777 (s), 699 (w) cm^{-1}

^{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 $[M+H]^+$.

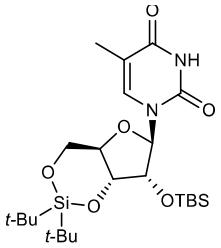
Compound 33T



In an oven-dried flask under argon atmosphere, 5'-DMT protected compound **32T** (0.1 g, 0.1 mmol, 1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 ^{o}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 r.t. 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 $[M+Na]^+$.

Compound 35



5-Methyluridine **34** (3 g, 11.62 mmol, 1 eq.) was suspended in DMF and di-tert-butylsilyl ditriflate (4.68 ml, 13.94 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 (3.96 g, 58.1 mmol, 5 eq.) was added and the reaction was warmed to r.t. over a period of 30 min. Then TBSCl (2.10 g, 13.9 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, 2/1, *v/v*).

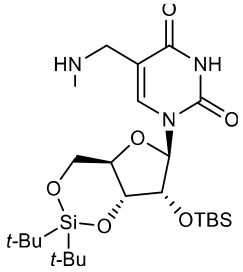
Yield: 97%
^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.78 (s, 1H), 7.01 (q, *J* = 1.2 Hz, 1H), 5.65 (s, 1H), 4.49 (dd, *J* = 9.2, 5.1 Hz, 1H), 4.29 (d, *J* = 4.8 Hz, 1H), 4.12 (td, *J* = 10.1, 5.1 Hz, 1H), 3.98 (dd, *J* = 10.5, 9.3 Hz, 1H), 3.92 (dd, *J* = 9.6, 4.9 Hz, 1H), 1.93 (d, *J* = 1.2 Hz, 3H), 1.06 (s, 9H), 1.02 (s, 9H), 0.92 (s, 9H), 0.16 (s, 3H), 0.13 (s, 3H)
HRMS (ESI): calculated for $C_{24}H_{45}N_2O_6Si_2^+$: *m/z* = 513.2816 $[M+H]^+$, found: *m/z* = 518.2813 $[M+H]^+$.

The analytical data is in agreement with the literature.[219]

General procedure for the synthesis of 36a and 36b

A solution of **35** (1.0 equiv.) in dry $CHCl_3$ was heated at 60°C. N-bromosuccinimide (NBS) (1.2 eq., previously purified by recrystallization) and azobisisobutyronitrile (AIBN) (0.12 eq.) 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 **36a**or NH_3 (0.5 M in 1,4-dioxane, 5.0 equiv.) for **36b** 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 (MgSO4), filtered and concentrated. The crude was purified by silica gel column chromatography to furnish **36a,b** as a yellow foam.

Compound 36a



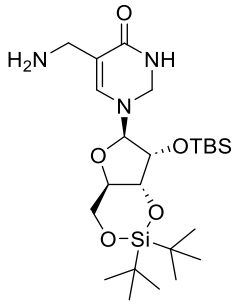
Eluent: 9:1 CH_2Cl_2/MeOH
Yield: 48%

IR: ṽ = 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) cm^{-1}

^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ:** 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): calculated for $C_{25}H_{48}N_3O_6Si_2^+$: m/z = 542.3076 $[M+H]^+$; found: m/z = 542.3076 $[M+H]^+$.

Compound 36b



Yield: 27%

IR: ṽ = 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) cm^{-1}

^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.33 (s, 1H), 5.69 (s, 1H), 4.48 (dd, *J* = 9.2, 5.0 Hz, 1H), 4.29 (d, *J* = 4.8 Hz, 1H), 4.13 (td, *J* = 10.0, 5.1 Hz, 1H), 4.02 (t, *J* = 9.9 Hz, 1H), 3.93 (dd, *J* = 9.6, 4.8 Hz, 1H), 3.62 (s, 2H), 1.05 (s, 9H), 1.02 (s, 9H), 0.92 (s, 9H), 0.16 (s, 3H), 0.13 (s, 3H)

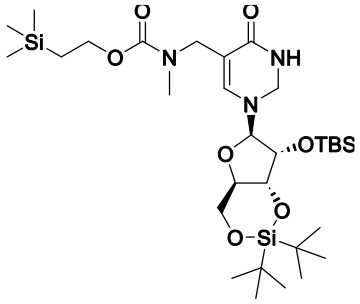
^{13}C NMR (100 MHz, $CDCl_3$) δ: 163.2, 149.6, 136.3, 94.1, 76.2, 75.4, 74.6, 67.7, 39.2, 27.6, 27.1, 26.0, 22.9, 20.5, 18.4, -4.2, -4.9

HRMS (ESI): calculated for $C_{24}H_{46}N_3O_6Si_2^+$: *m*/*z* = 528.2925 $[M+H]^+$, found: *m*/*z*: 528.29185 $[M+H]^+$.

General procedure for synthesis of 37a,b

To a solution of **36a,b** (1.0 equiv.) in 1,4-dioxane and H_2O (1:1 v/v) were added teoc-OSu (1.1 equiv.) and NEt_3 (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 **37a,b** as a white solid.

Compound 37a



Yield: 79%

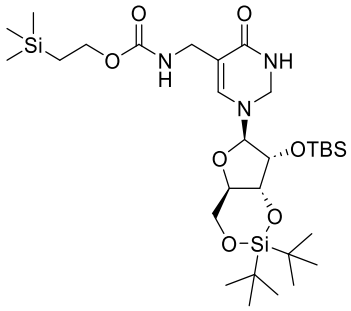
IR: ṽ = 3054 (w), 2956 (w), 2359 (w), 1692 (m), 1463 (w), 1422 (w), 1264 (s), 1214 (w), 1167 (w), 1146 (w), 1059 (w), 1000 (w), 938 (w), 895 (m), 838 (m), 730 (s), 702 (s) cm^{-1}

^{1}H NMR (500 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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

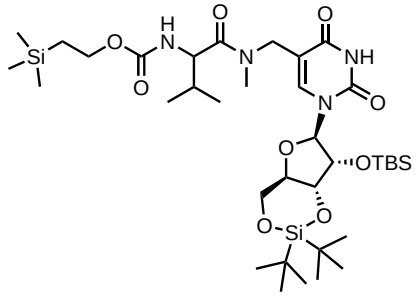
HRMS (ESI): calculated for $C_{31}H_{60}N_3O_8Si_3^+$: *m*/*z* = 686.3683 $[M+H]^+$, found: *m*/*z* = 686.3683 $[M+H]^+$.

Compound 37b



Yield: 74%
IR: ṽ = 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) cm^{-1}
^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.82 (s, 1H), 7.47 (s, 1H), 5.64 (s, 1H), 5.32 (t, *J* = 6.3 Hz, 1H), 4.50 (dd, *J* = 9.1, 5.0 Hz, 1H), 4.28 (d, *J* = 4.6 Hz, 1H), 4.22 – 4.03 (m, 4H), 3.98 (d, *J* = 6.4 Hz, 2H), 3.91 (dd, *J* = 9.6, 4.6 Hz, 1H), 1.06 (s, 9H), 1.02 (s, 9H), 0.99 – 0.92 (m, 11H), 0.18 (s, 3H), 0.14 (s, 3H), 0.03 (s, 9H)
^{13}C NMR (100 MHz, $CDCl_3$) δ: 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): Calcd. for $C_{30}H_{57}N_3NaO_8Si_3^+$: *m/z* = 694.3351 $[M+Na]^+$, found: *m/z* = 694.3341 $[M+Na]^+$.

Compound 37c



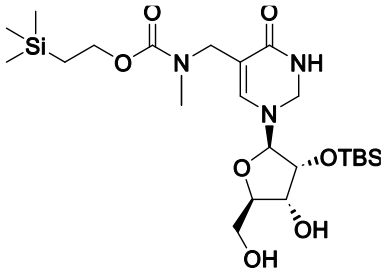
Teoc-protected valine was synthesized following a previously reported procedure in literature.[221] Teoc-Val-OH (1.2 equiv.) was dissolved in dry DCM and DMF (99:1 v/v). To the solution, 1-hydroxybenzotriazole hydrate (HOBt•H_2O) (1.2 eq.), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (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 **36a** (1.0 eq.) in CH_2Cl_2 was added and the reaction was stirred for 24 h. The reaction mixture was extracted three times with CH_2Cl_2. The combined organic layers were dried ($MgSO_4$), filtered and concentrated. Purification by silica gel column chromatography furnished the amino acid conjugate **37c** as a white foam.
Yield: 87%
IR: ṽ = 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) cm^{-1}
For the major rotamer:
^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{36}H_{69}N_4O_9Si_3^+$: m/z = 785.4367 $[M+H]^+$; found: m/z = 785.4363 $[M+H]^+$.

General procedure for 3', 5' silyl-deprotection of (m)nm^{5}U derivatives

The modified 5-methyluridine **37a-c** (1.0 equiv.) was dissolved in CH_2Cl_2/pyridine (9:1 v/v) and cooled to 0 ˚C in a plastic reaction vessel. Subsequently, a solution of 70% HF-pyridine (5.0 eq.) 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 CH_2Cl_2. The combined organic layers were washed with water, dried ($MgSO_4$), filtered and

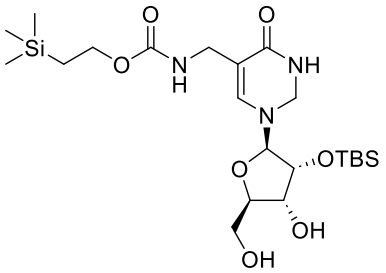
concentrated. The crude product was purified by silica gel column chromatography to afford the diol compound **38a-c** as a white foam.

Compound 38a



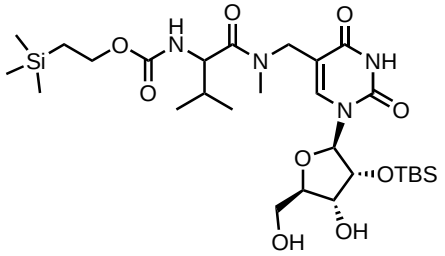
Yield: 80%;
IR: ṽ = 3417 (w), 3060 (w), 2949 (w), 2856 (w), 2359 (w), 1673 (s), 1462 (m), 1401 (w), 1362 (w), 1250 (m), 1214 (w), 1144 (m), 1088 (m), 1060 (m), 1005 (w), 938 (w), 833 (s), 777 (s), 693 (w) cm^{-1}
^{1}H NMR (400 MHz, $CDCl_3$) δ: 9.47 (s, 1H), 8.19 (s, 1H), 5.87 (d, J = 5.2 Hz, 1H), 4.48 (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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{23}H_{44}N_3O_8Si_2^+$: m/z = 546.2661 $[M+H]^+$; found: m/z = 546.2666 $[M+H]^+$.

Compound 38b



Yield: 90%
IR: ṽ = 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) cm^{-1}
^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.96 (s, 1H), 8.10 (s, 1H), 5.81 (d, *J* = 4.8 Hz, 1H), 5.46 (t, *J* = 6.4 Hz, 1H), 4.48 (t, *J* = 5.0 Hz, 1H), 4.26 (q, *J* = 4.4 Hz, 1H), 4.15 – 4.07 (m, 3H), 4.01 – 3.93 (m, 3H), 3.81 (d, *J* = 12.6 Hz, 1H), 3.66 (s, 1H), 2.71 (d, *J* = 4.6 Hz, 1H), 0.95 (dd, *J* = 10.1, 7.0 Hz, 2H), 0.90 (s, 9H), 0.09 (s, 6H), 0.02 (s, 9H)
^{13}C NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{22}H_{41}N_3NaO_8Si_2^+$: *m*/*z* = 554.2330 $[M+Na]^+$, found: *m*/*z* = 554.2323 $[M+Na]^+$.

Compound 38c



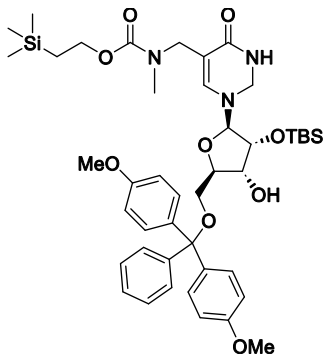
Yield: 91%

IR: ṽ = 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) cm^{-1}
For the major rotamer:
^{1}H NMR (400 MHz, $CDCl_3$) δ: 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 NMR (100 MHz, $CDCl_3$) δ: 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): calculated for $C_{28}H_{53}N_4O_9Si_2^+$: m/z = 645.3346 $[M+H]^+$; found: m/z = 645.3349 $[M+H]^+$.

DMT protection of 5'OH of (m)nm^{5}U derivatives

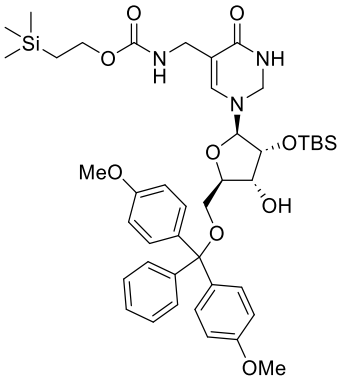
To a solution of the 3',5'-deprotected 5-methyluridine derivative **38a-c** (1.0 equiv.) in pyridine was added 4,4'- dimethoxytrityl chloride (DMTCl) (1.5 equiv.). After stirring at r.t. for 16 h, the reaction mixture was concentrated and purified by silica gel column chromatography with an addition of 0.1% of pyridine to the eluent to afford the DMT-protected compound **39a-c** as a white foam.

Compound 39a



Yield: 83%
IR: ṽ = 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) cm^{-1}
For the major rotamer:
^{1}H NMR (400 MHz, Acetone-d_6) δ: 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 NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{44}H_{60}N_3O_{10}Si_2^-$: m/z = 846.3823 $[M-H]^-$; found: m/z = 846.3825 $[M-H]^-$.

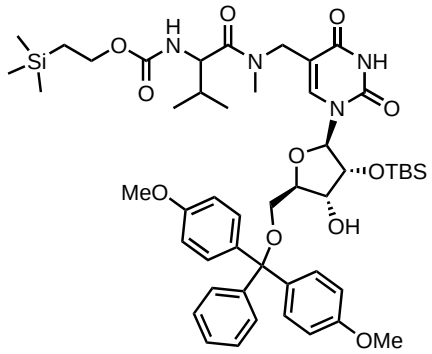
Compound 39b



Yield: 99%

IR: ṽ = 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) cm^{-1}
^{1}H NMR (500 MHz, CD_2Cl_2) δ: 8.67 (s, 1H), 7.70 (s, 1H), 7.49 – 7.42 (m, 2H), 7.38 – 7.27 (m, 6H), 7.26 – 7.22 (m, 1H), 6.90 – 6.83 (m, 4H), 5.91 (d, *J* = 4.8 Hz, 1H), 5.09 (t, *J* = 6.0 Hz, 1H), 4.42 (t, *J* = 5.0 Hz, 1H), 4.22 (q, *J* = 4.8 Hz, 1H), 4.18 – 4.01 (m, 3H), 3.78 (s, 6H), 3.56 (dd, *J* = 14.7, 5.8 Hz, 1H), 3.43 (dd, *J* = 15.1, 7.0 Hz, 1H), 3.39 (d, *J* = 2.9 Hz, 2H), 2.62 (d, *J* = 5.0 Hz, 1H), 0.92 (s, 11H), 0.14 (m, 6H), 0.02 (s, 9H).
^{13}C NMR (100 MHz, Acetone-d_6) δ: 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): calculated for $C_{43}H_{59}N_3NaO_{10}Si_2^+$: *m*/*z* = 856.3637 $[M+Na]^+$, found: *m*/*z* = 856.3626 $[M+Na]^+$.

Compound 39c

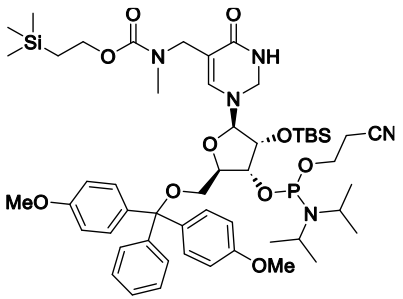


Yield: 89%
IR: ṽ = 3054 (w), 2954 (w), 2930 (w), 2857 (w), 2359 (w), 1687 (s), 1644 (w), 1608 (w), 1508 (m), 1463 (m), 1389 (w), 1263 (m), 1249 (s), 1175 (m), 1115 (w), 1083 (w), 1061 (w), 1035 (m), 967 (w), 935 (w), 914 (w), 858 (m), 833 (s), 780 (w), 733 (s), 700 (s) cm^{-1}
For the major rotamer:
^{1}H NMR (400 MHz, $CDCl_3$) *δ*: 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 NMR (100 MHz, $CDCl_3$) *δ*: 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): calculated for $_{49}H_{69}N_4O_{11}Si_2^-$: m/z = 945.4507 $[M-H]^-$; found: m/z = 945.4508 $[M-H]^-$.

General synthesis of (m)nm^{5}U nucleoside phosphoramidites

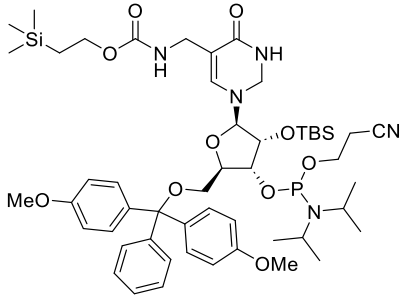
A solution of 5'-DMT-protected compound **39a-c** (1.0 equiv.) and DIPEA (4.0 equiv.) in dry CH_2Cl_2 was cooled to 0 °C. To this solution was slowly added 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (CED-Cl) (2.5 equiv.) and the reaction mixture was stirred at r.t. for 5 h. The reaction was quenched by addition of aq. sat. NaHCO3 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 **40a-c** was obtained as a mixture of diastereoisomers and rotamers as a white foam.

Compound 40a



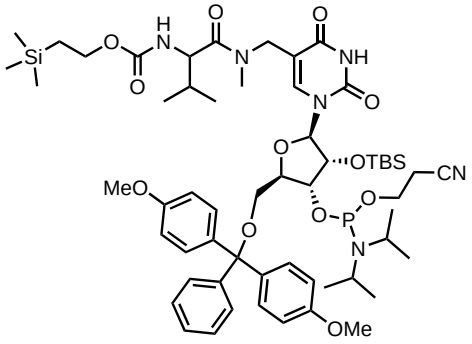
Yield: 78%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.3, 150.2, 148.8, 148.3
HRMS (ESI): calculated for $C_{53}H_{77}N_5O_{11}PSi_2^-$: m/z = 1046.4901 [M-H]$^-$; found: m/z = 1046.4896 [M-H]$^-$.

Compound 40b



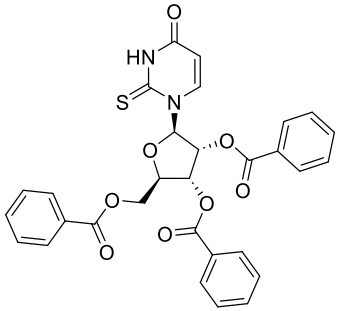
Yield: 92%
^{31}P NMR (202 MHz, CD_2Cl_2) δ: 150.31, 149.20
HRMS (ESI): calculated for $C_{52}H_{77}N_5O_{11}PSi_2^+$: *m*/*z* = 1034.4900 [M+H]$^+$, found: *m*/*z* = 1034.4887 [M+H]$^+$.

Compound 40c



Yield: 89%
^{31}P NMR (162 MHz, Acetone-d_6) δ: 150.0, 149.9, 149.4, 149.3
HRMS (ESI): calculated for $C_{58}H_{86}N_6O_{12}PSi_2^-$: m/z = 1145.5585 [M-H]$^-$; found 1145.5595 [M-H]$^-$.

Compound 42



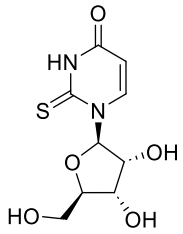
2-Thiouracil **41** (1.0 equiv.) was suspended in HMDS and stirred at r.t.. TMS-Cl (2.0 equiv.) was added dropwise. The reaction was refluxed overnight and subsequently a solution of 1-*O*-Acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (0.7 equiv.) in 1,2-dichloroethane was added, so that a 1:1 ratio of HMDS and 1,2-dichloroethane was formed. This mixture was cooled to 0 ˚C and $SnCl_4$ (1.4 equiv.

from 1 M solution in DCM) was added dropwise. After addition was completed, the cooling was removed, and the reaction was stirred at r.t. overnight. The reaction was quenched by addition of aqueous saturated $NaHCO_3$ solution and stirred for 1 h. After that, H_2O was added. The organic phase was separated, and the aqueous phase was extracted with CH_2Cl_2 (three times). The combined organic layers were dried over Na_2SO_4 and the solvents were removed *in vacuo*. Purification of the crude product via flash column chromatography (0-1.5% MeOH in DCM) yielded target compound **42** as a white foam.

Yield: 55%
^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.90 (s, 1H), 8.10 – 8.06 (m, 2H), 8.02 – 7.98 (m, 2H), 7.95 – 7.90 (m, 2H), 7.70 (d, *J* = 8.3 Hz, 1H), 7.68 – 7.62 (m, 1H), 7.61 – 7.54 (m, 2H), 7.52 (t, *J* = 7.8 Hz, 2H), 7.42 (t, *J* = 7.9 Hz, 2H), 7.38 (t, *J* = 7.8 Hz, 2H), 7.26 (d, *J* = 4.8 Hz, 1H), 5.84 – 5.76 (m, 3H), 4.88 (dd, *J* = 12.6, 2.6 Hz, 1H), 4.78 (dt, *J* = 5.3, 2.8 Hz, 1H), 4.69 (dd, *J* = 12.7, 3.0 Hz, 1H)
HRMS (ESI): calculated for $C_{30}H_{23}N_2O_8S^-$: *m*/*z* = 571.1175 $[M-H]^-$; found: *m*/*z* = 571.1183 $[M-H]^-$.

The analytical data is in agreement with the literature.[231]

Compound 43

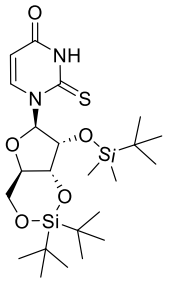


Protected thiouridine **42** (1.0 equiv.) was dissolved in MeOH under N_2 atmosphere and the solution was cooled to 0°C. Afterwards $NaOCH_3$ (6.0 equiv.) was added in several portions and the reaction was stirred for 4 h at r.t.. Next, freshly protonated ion-exchange resin (Dowex 50WX8) was added until the solution showed a neutral pH. The resin was then filtered and washed with MeOH. The combined solvents were vaporized *in vacuo*. The crude was suspended in H_2O and the aqueous phase was washed with a small amount of EtOAc (two times). The aqueous phase was lyophilized to afford target compound **43** as a white powder.

Yield: 61%
^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.44 (s, 1H), 8.17 (d, *J* = 8.1 Hz, 1H), 6.54 (d, *J* = 3.5 Hz, 1H), 5.94 (d, *J* = 8.1 Hz, 1H), 4.05 – 4.00 (m, 1H), 3.97 (t, *J* = 5.4 Hz, 1H), 3.89 (dt, *J* = 5.6, 2.6 Hz, 1H), 3.69 (dd, *J* = 12.4, 2.7 Hz, 1H), 3.60 (dd, *J* = 12.3, 2.7 Hz, 1H)
HRMS (ESI): calculated for $C_9H_{11}N_2O_5S^+$: *m*/*z* = 259.0389 $[M-H]^-$; found: *m*/*z* = 259.0326 $[M-H]^-$.

The analytical data is in agreement with the literature.[231]

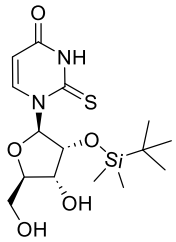
Compound 44



Yield: 80%
^{1}H NMR (500 MHz, $CDCl_3$) δ: 10.09 (s, 1H), 7.35 (d, *J* = 8.2 Hz, 1H), 6.46 (s, 1H), 6.03 (dd, *J* = 8.2, 1.8 Hz, 1H), 4.55 (dd, *J* = 9.4, 5.3 Hz, 1H), 4.34 (d, *J* = 4.4 Hz, 1H), 4.25 (td, *J* = 10.2, 5.3 Hz, 1H), 4.01 (dd, *J* = 10.6, 9.5 Hz, 1H), 3.78 (dd, *J* = 9.8, 4.5 Hz, 1H), 1.05 (d, *J* = 1.3 Hz, 9H), 1.02 (s, 9H), 0.95 (s, 9H), 0.23 (s, 3H), 0.16 (s, 3H)
HRMS (ESI): calculated for $C_{23}H_{43}N_2O_5SSi_2^+$: *m*/*z* = 515.2431 $[M+H]^+$; found: *m*/*z* = 515.2429 $[M+H]^+$.

The analytical data is in agreement with the literature.[231]

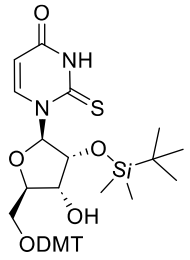
Compound 45



Yield: 75%
^{1}H NMR (500 MHz, $CDCl_3$) δ: 10.39 (s, 1H), 8.22 (d, *J* = 8.2 Hz, 1H), 6.67 (d, *J* = 3.3 Hz, 1H), 6.00 (d, *J* = 8.2 Hz, 1H), 4.31 (dd, *J* = 4.7, 3.3 Hz, 1H), 4.18 (q, *J* = 4.8 Hz, 1H), 4.14 (dt, *J* = 5.7, 2.0 Hz, 1H), 4.07 (dd, *J* = 12.0, 2.1 Hz, 1H), 3.93 (dd, *J* = 12.0, 1.9 Hz, 1H), 2.69 – 2.63 (m, 1H), 2.51 (s, 1H), 0.92 (s, 9H), 0.20 (s, 3H), 0.13 (s, 3H)
HRMS (ESI): calculated for $C_{15}H_{25}N_2O_5SSi^-$: *m*/*z* = 373.1253 $[M-H]^-$, found: *m*/*z* = 373.1257 $[M-H]^-$.

The analytical data is in agreement with the literature.[231]

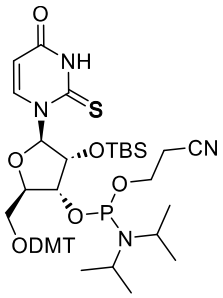
Compound 46



Yield: 72%
^{1}H NMR (500 MHz, CD_2Cl_2) δ: 9.93 (s, 1H), 8.14 (d, *J* = 8.2 Hz, 1H), 7.42 – 7.36 (m, 2H), 7.36 – 7.19 (m, 8H), 6.90 – 6.83 (m, 4H), 6.69 (d, *J* = 3.6 Hz, 1H), 5.44 (d, *J* = 8.2 Hz, 1H), 4.44 (dd, *J* = 4.7, 3.7 Hz, 1H), 4.39 (t, *J* = 5.0 Hz, 1H), 4.15 (dt, *J* = 4.9, 2.2 Hz, 1H), 3.79 (d, *J* = 0.7 Hz, 6H), 3.50 (qd, *J* = 11.2, 2.3 Hz, 2H), 0.94 (s, 9H), 0.22 (s, 3H), 0.18 (s, 3H)
HRMS (ESI): calculated for $C_{36}H_{43}N_2O_7SSi^-$: *m*/*z* = 675.2560 $[M-H]^-$, found: *m*/*z* = 675.2564 $[M-H]^-$.

The analytical data is in agreement with the literature.[231]

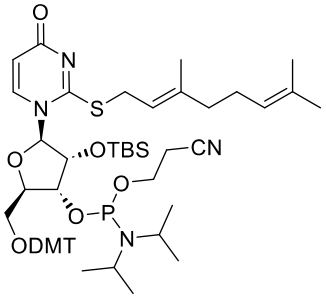
Compound 47



Yield: 82%
^{31}P NMR (202 MHz, CD_2Cl_2) δ: 149.95, 149.87
HRMS (ESI): calculated for $C_{45}H_{61}N_4O_8PSSi^+$: *m*/*z* = 877.3795 $[M+H]^+$, found: *m*/*z* = 877.3804 $[M+H]^+$.

The analytical data is in agreement with the literature.[231]

Compound 48

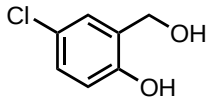


47 (1.0 equiv.) was dissolved in anhydrous methanol to give a 10 mM concentration, followed by the addition of DIPEA (3.0 equiv.) and geranyl bromide (1.5 equiv.). The reaction was stirred at r.t. under inert atmosphere for 3 h. The reaction mixture was diluted with DCM and washed with brine 3 times. The crude was then purified by flash-column chromatography to afford the geranylated 2-thiouridine phosphoramidite **48** as a white solid.

Yield: 90%
^{31}P NMR (202 MHz, CD_2Cl_2) δ: 151.47, 148.98.

The analytical data is in agreement with the literature.[232]

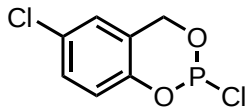
Compound 50



Yield: 75%
^{1}H NMR (500 MHz, DMSO-d_6) δ: 9.68 (s, 1H), 7.26 (d, *J* = 2.7 Hz, 1H), 7.06 (dd, *J* = 8.5, 2.8 Hz, 1H), 6.76 (d, *J* = 8.5 Hz, 1H), 5.13 (t, *J* = 5.7 Hz, 1H), 4.44 (d, *J* = 5.6 Hz, 2H).

The analytical data is in agreement with the literature.[316]

Compound 51

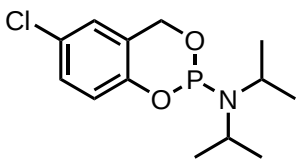


Yield: 42%
^{1}H NMR (400 MHz, CD_3CN) δ: 7.31 (dd, *J* = 8.8, 2.5 Hz, 1H), 7.15 (d, *J* = 2.5 Hz, 1H), 7.00 (d, *J* = 8.8 Hz, 1H), 5.26 (d, *J* = 6.2 Hz, 2H).

^{31}P NMR (162 MHz, CD_3CN) δ: 139.2

The analytical data is in agreement with the literature.[316]

Compound 52



Yield: 82%
^{1}H NMR (400 MHz, CD_2Cl_2) δ: 7.14 (dd, *J* = 8.7, 2.6 Hz, 1H), 6.98 (d, *J* = 2.5 Hz, 1H), 6.83 (d, *J* = 8.6 Hz, 1H), 5.07 (dd, *J* = 14.5, 5.0 Hz, 1H), 4.81 (dd, *J* = 19.7, 14.5 Hz, 1H), 3.60 (dp, *J* = 10.7, 6.8 Hz, 2H), 1.23 (d, *J* = 6.8 Hz, 6H), 1.21 (d, *J* = 6.8 Hz, 6H)
^{31}P NMR (162 MHz, CD_2Cl_2) δ: 136.6

The analytical data is in agreement with the literature.[316]

5. List of references

1 Aristotle. *History of Animal Book V*. (c. 343 BC).

2 Lennox, J. G. *Cambridge studies in philosophy and biology: Aristotle's philosophy of biology: Studies in the origins of life science*. 229-258 (Cambridge University Press, 2000).

3 Balme, D. M. Development of Biology in Aristotle and Theophrastus: Theory of Spontaneous Generation. *Phronesis* **7**, 91-104 (1962).

4 Dobell, C. *Antony van Leewenhoek and his " Little animals"*. (Dover Publications, 1960).

5 Farley, J. *Spontaneous generation controversy from Descartes to Oparin*. (Johns Hopkins University Press, 1977).

6 Oparin, A. I. *Translation (by Ann Synge) of The Origin of Life on the Earth by A. I. Oparin*. (Academic Press, 1957).

7 Berzelius, J. J. *Föreläsningar i Djurkemien*. (Tryckte hos Carl Delén, 1806).

8 Pasteur, L. *Mémoire sur les corpuscules organisés qui existent dans l'atmosphère : examen de la doctrine des générations spontanées*. (Mallet-Bachelier, 1862).

9 Wöhler, F. Ueber künstliche Bildung des Harnstoffs. *Annalen der Physik* **88**, 253-256 (1828).

10 Tsipis, C. A. & Karipidis, P. A. Mechanism of a Chemical Classic: Quantum Chemical Investigation of the Autocatalyzed Reaction of the Serendipitous Wöhler Synthesis of Urea. *J. Am. Chem. Soc.* **125**, 2307-2318 (2003).

11 Darwin, C. *On the Origin of Species by Means of Natural Selection*. (John Murray, 1859).

12 Bronn, H. G. [Review of] Ch. Darwin: on the Origin of Species by means of Natural Selection, or the preservation of favoured races in the struggle for life (502 pp. 8˚, London 1859). *Neues Jahrbuch für Mineralogie, Geognosie, Geologie und Petrefaktenkunde*, 112-116 (1860).

13 Hull, D. L. *Darwin and His Critics: The Reception of Darwin's Theory of Evolution by the Scientific Community*. 120-124 (Harvard University Press, 1973).

14 Spencer, H. Ch. I. Organic Matter, 3-24 (Williams and Norgate, 1864).

15 Dyer, W. T. T. On Spontaneous Generation and Evolution. *J. Cell Sci.* **s2-10**, 333-354 (1870).

16 Darwin Correspondence Project. *Letter no. 7471, Charles Darwin to J. D. Hooker*. accessed on 17 May 2024. <https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-7471.xml>.

17 Haldane, J. B. S. The Origin of Life. *The Rationalist Annual*, 3-10 (1929).

18 Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **117**, 528-529 (1953).

19 Johnson, A. P. *et al*. The Miller Volcanic Spark Discharge Experiment. *Science* **322**, 404-404 (2008).

20 Cobb, A. K., Pudritz, R. E. & Pearce, B. K. D. Nature's starships. I. Observed abundances and relative frequencies of amino acids in meteorites. *Astrophys. J.* **783** (2014).

21 Chyba, C. & Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**, 125-132 (1992).
22 Cobb, A. K., Pudritz, R. E. & Pearce, B. K. D. Nature's Starships. II. Simulating the Synthesis of Amino Acids in Meteorite Parent Bodies. *Astrophys. J.* **809**, 6 (2015).
23 Kvenvolden, K. *et al*. Evidence for Extraterrestrial Amino-acids and Hydrocarbons in the Murchison Meteorite. *Nature* **228**, 923-926 (1970).
24 Lawless, J. G. & Peterson, E. Amino acids in carbonaceous chondrites. *Orig. Life* **6**, 3-8 (1975).
25 Glavin, D. P. *et al*. Extraterrestrial amino acids and L-enantiomeric excesses in the CM2 carbonaceous chondrites Aguas Zarcas and Murchison. *Meteorit. Planet. Sci*. **56**, 148-173 (2021).
26 Glavin, D. P., Callahan, M. P., Dworkin, J. P. & Elsila, J. E. The effects of parent body processes on amino acids in carbonaceous chondrites. *Meteorit. Planet. Sci*. **45**, 1948-1972 (2010).
27 Potiszil, C. *et al*. Insights into the formation and evolution of extraterrestrial amino acids from the asteroid Ryugu. *Nat. Commun*. **14**, 1482 (2023).
28 Martins, Z. *et al*. Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett*. **270**, 130-136 (2008).
29 Callahan, M. P. *et al*. Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci*. **108**, 13995-13998 (2011).
30 Oba, Y. *et al*. Identifying the wide diversity of extraterrestrial purine and pyrimidine nucleobases in carbonaceous meteorites. *Nat. Commun*. **13**, 2008 (2022).
31 Oba, Y. *et al*. Uracil in the carbonaceous asteroid (162173) Ryugu. *Nat. Commun*. **14**, 1292 (2023).
32 Balsiger, H. *et al*. Rosina – Rosetta Orbiter Spectrometer for Ion and Neutral Analysis. *Space Sci. Rev*. **128**, 745-801 (2007).
33 Altwegg, K. *et al*. Prebiotic chemicals—amino acid and phosphorus—in the coma of comet 67P/Churyumov-Gerasimenko. *Sci. Adv*. **2**, e1600285 (2016).
34 Le Roy, L. *et al*. Inventory of the volatiles on comet 67P/Churyumov-Gerasimenko from Rosetta/ROSINA. *A&A* **583**, A1 (2015).
35 Turner, B. E. & Bally, J. Detection of Interstellar PN: The First Identified Phosphorus Compound in the Interstellar Medium. *Astrophys. J.* **321**, L75 (1987).
36 Tenenbaum, E. D., Woolf, N. J. & Ziurys, L. M. Identification of Phosphorus Monoxide ($X^2\Pi_r$) in VY Canis Majoris: Detection of the First PO Bond in Space. *Astrophys. J.* **666**, L29-L32 (2007).
37 Agúndez, M., Cernicharo, J., Decin, L., Encrenaz, P. & Teyssier, D. Confirmation of Circumstellar Phosphine. *Astrophys. J.* **790**, L27 (2014).
38 Agúndez, M., Cernicharo, J. & Guélin, M. Discovery of Phosphaethyne (HCP) in Space: Phosphorus Chemistry in Circumstellar Envelopes. *Astrophys. J.* **662**, L91-L94 (2007).
39 Liszt, H. & Lucas, R. Comparative chemistry of diffuse clouds. *A&A* **370**, 576-585 (2001).
40 Irvine, W. M. & Schloerb, F. P. Cyanide and isocyanide abundances in the cold, dark cloud TMC-1. *Astrophys. J.* **282**, 516-521 (1984).
41 Hily-Blant, P., Walmsley, M., Pineau des Forêts, G. & Flower, D. Nitrogen chemistry and depletion in starless cores*. *A&A* **513**, A41 (2010).

42 Jørgensen, J. K., Schöier, F. L. & van Dishoeck, E. F. Molecular inventories and chemical evolution of low-mass protostellar envelopes*. *A&A* **416**, 603-622 (2004).
43 Garrod, R. T. A Three-phase Chemical Model of Hot Cores: The Formation of Glycine. *Astrophys. J.* **765**, 60 (2013).
44 Meierhenrich, U. J. *et al*. Precursors of Biological Cofactors from Ultraviolet Irradiation of Circumstellar/Interstellar Ice Analogues. *Chem. Eur. J.* **11**, 4895-4900 (2005).
45 Bossa, J.-B., Borget, F., Duvernay, F., Theulé, P. & Chiavassa, T. How a usual carbamate can become an unusual intermediate: a new chemical pathway to form glycinate in the interstellar medium. *J. Phys. Org. Chem.* **23**, 333-339 (2010).
46 Loison, J.-C., Wakelam, V. & Hickson, K. M. The interstellar gas-phase chemistry of HCN and HNC. *Mon. Not. R. Astron. Soc.* **443**, 398-410 (2014).
47 Oró, J. Comets and the Formation of Biochemical Compounds on the Primitive Earth. *Nature* **190**, 389-390 (1961).
48 Brent Dalrymple, G., Lewis, C. L. E. & Knell, S. J. in *The Age of the Earth: From 4004 BC to AD 2002*. 205-221 (Geological Society of London, 2001).
49 Cloud, P. A working model of the primitive Earth. *Am. J. Sci.* **272**, 537-548 (1972).
50 Bell, E. A., Boehnke, P., Harrison, T. M. & Mao, W. L. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc. Natl. Acad. Sci.* **112**, 14518-14521 (2015).
51 Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466-469 (2005).
52 Genda, H., Brasser, R. & Mojzsis, S. J. The terrestrial late veneer from core disruption of a lunar-sized impactor. *Earth Planet. Sci. Lett.* **480**, 25-32 (2017).
53 Walker, R. J. Highly siderophile elements in the Earth, Moon and Mars: Update and implications for planetary accretion and differentiation. *Geochemistry* **69**, 101-125 (2009).
54 Zellner, N. E. B. Cataclysm No More: New Views on the Timing and Delivery of Lunar Impactors. *Orig. Life Evol. Biosph.* **47**, 261-280 (2017).
55 Benner, S. A. *et al*. When Did Life Likely Emerge on Earth in an RNA-First Process? *ChemSystemsChem* **2**, e1900035 (2020).
56 Sasselov, D. D., Grotzinger, J. P. & Sutherland, J. D. The origin of life as a planetary phenomenon. *Science Advances* **6**, eaax3419 (2020).
57 Kroll, J. H. *et al*. Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. *Nature Chemistry* **3**, 133-139 (2011).
58 Eschenmoser, A. & Volkan Kisakürek, M. Chemistry and the Origin of Life. *Helv. Chim. Acta* **79**, 1249-1259 (1996).
59 Oró, J. Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.* **2**, 407-412 (1960).
60 Oró, J. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **191**, 1193-1194 (1961).
61 Wakamatsu, H., Yamada, Y., Saito, T., Kumashiro, I. & Takenishi, T. Synthesis of Adenine by Oligomerization of Hydrogen Cyanide. *J. Org. Chem.* **31**, 2035-2036 (1966).

62 Ferris, J. P. & Orgel, L. E. Studies in Prebiotic Synthesis. I. Aminomalononitrile and 4-Amino-5-cyanoimidazole1,2. *J. Am. Chem. Soc.* **88**, 3829-3831 (1966).
63 Sanchez, R. A., Ferbis, J. P. & Orgel, L. E. Studies in Prebiodc Synthesis: II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **30**, 223-253 (1967).
64 Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis: IV. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. *J. Mol. Biol.* **38**, 121-128 (1968).
65 Miyakawa, S., James Cleaves, H. & Miller, S. L. The Cold Origin of Life: A. Implications Based On The Hydrolytic Stabilities Of Hydrogen Cyanide And Formamide. *Orig. Life Evol. Biosph.* **32**, 195-208 (2002).
66 Ruiz-Bermejo, M., Zorzano, M.-P. & Osuna-Esteban, S. Simple Organics and Biomonomers Identified in HCN Polymers: An Overview. *Life* **3**, 421-448 (2013).
67 Matthews, C. N. Dark matter in the solar system: Hydrogen cyanide polymers. *Orig. Life Evol. Biosph.* **21**, 421-434 (1991).
68 Matthews, C. N. Hydrogen cyanide polymers: from laboratory to space. *Planet. Space Sci.* **43**, 1365-1370 (1995).
69 Ruiz-Bermejo, M., de la Fuente, J. L., Pérez-Fernández, C. & Mateo-Martí, E. A Comprehensive Review of HCN-Derived Polymers. *Processes* **9**, 597 (2021).
70 Lowe, C. U., Rees, M. W. & Markham, R. Synthesis of Complex Organic Compounds from Simple Precursors: Formation of Amino-Acids, Amino-Acid Polymers, Fatty Acids and Purines from Ammonium Cyanide. *Nature* **199**, 219-222 (1963).
71 Ferris, J. P., Joshi, P. C., Edelson, E. H. & Lawless, J. G. HCN: A plausible source of purines, pyrimidines and amino acids on the primitive earth. *J. Mol. Evol.* **11**, 293-311 (1978).
72 Sanchez, R., Ferris, J. & Orgel, L. E. Conditions for Purine Synthesis: Did Prebiotic Synthesis Occur at Low Temperatures? *Science* **153**, 72-73 (1966).
73 Bada, J. L., Bigham, C. & Miller, S. L. Impact melting of frozen oceans on the early Earth: Implications for the origin of life. *Proc. Natl. Acad. Sci.* **91**, 1248-1250 (1994).
74 Schwartz, A. W., Joosten, H. & Voet, A. B. Prebiotic adenine synthesis via HCN oligomerization in ice. *Biosystems* **15**, 191-193 (1982).
75 Miyakawa, S., Cleaves, H. J. & Miller, S. L. The Cold Origin of Life: B. Implications Based on Pyrimidines and Purines Produced From Frozen Ammonium Cyanide Solutions. *Orig. Life Evol. Biosph.* **32**, 209-218 (2002).
76 Levy, M., Miller, S. L., Brinton, K. & Bada, J. L. Prebiotic Synthesis of Adenine and Amino Acids Under Europa-like Conditions. *Icarus* **145**, 609-613 (2000).
77 Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Cyanoacetylene in Prebiotic Synthesis. *Science* **154**, 784-785 (1966).
78 Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: III. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol.* **33**, 693-704 (1968).
79 Ferris, J. P., Zamek, O. S., Altbuch, A. M. & Freiman, H. Chemial evolution. *J. Mol. Evol.* **3**, 301-309 (1974).
80 Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **375**, 772-774 (1995).

81 Nelson, K. E., Robertson, M. P., Levy, M. & Miller, S. L. Concentration by Evaporation and the Prebiotic Synthesis of Cytosine. *Orig. Life Evol. Biosph.* **31**, 221-229 (2001).
82 Robertson, M. P., Levy, M. & Miller, S. L. Prebiotic synthesis of diaminopyrimidine and thiocytosine. *J. Mol. Evol.* **43**, 543-550 (1996).
83 Shapiro, R. Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life. *Proc. Natl. Acad. Sci.* **96**, 4396-4401 (1999).
84 Okamura, H. *et al.* A one-pot, water compatible synthesis of pyrimidine nucleobases under plausible prebiotic conditions. *Chem. Commun.* **55**, 1939-1942 (2019).
85 Bada, J. L., Chalmers, J. H. & Cleaves, H. J. Is formamide a geochemically plausible prebiotic solvent? *Phys. Chem. Chem. Phys.* **18**, 20085-20090 (2016).
86 Niether, D., Afanasenkau, D., Dhont, J. K. G. & Wiegand, S. Accumulation of formamide in hydrothermal pores to form prebiotic nucleobases. *Proc. Natl. Acad. Sci.* **113**, 4272-4277 (2016).
87 Yamada, H. & Okamoto, T. A one-step synthesis of purin ring from formamide. *Chem. Pharm. Bull.* **20**, 623-624 (1972).
88 Saladino, R., Crestini, C., Costanzo, G., Negri, R. & Di Mauro, E. A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3H)-pyrimidinone from formamide: implications for the origin of life. *Bioorg. Med. Chem.* **9**, 1249-1253 (2001).
89 Saladino, R. *et al.* One-Pot TiO_2-Catalyzed Synthesis of Nucleic Bases and Acyclonucleosides from Formamide: Implications for the Origin of Life. *ChemBioChem* **4**, 514-521 (2003).
90 Bernal, J. D. *The Physical Basis of Life*. (Routledge and Paul, 1951).
91 Saladino, R., Crestini, C., Ciciriello, F., Costanzo, G. & Di Mauro, E. Formamide Chemistry and the Origin of Informational Polymers. *Chem. Biodiversity* **4**, 694-720 (2007).
92 Saladino, R. *et al.* Synthesis and Degradation of Nucleobases and Nucleic Acids by Formamide in the Presence of Montmorillonites. *ChemBioChem* **5**, 1558-1566 (2004).
93 Saladino, R., Crestini, C., Cossetti, C., Di Mauro, E. & Deamer, D. Catalytic effects of Murchison Material: Prebiotic Synthesis and Degradation of RNA Precursors. *Orig. Life Evol. Biosph.* **41**, 437-451 (2011).
94 Saladino, R. *et al.* Origin of Informational Polymers: The Concurrent Roles of Formamide and Phosphates. *ChemBioChem* **7**, 1707-1714 (2006).
95 Saladino, R. *et al.* Synthesis and Degradation of Nucleic Acid Components by Formamide and Iron Sulfur Minerals. *J. Am. Chem. Soc.* **130**, 15512-15518 (2008).
96 Saladino, R. *et al.* The Role of the Formamide/Zirconia System in the Synthesis of Nucleobases and Biogenic Carboxylic Acid Derivatives. *J. Mol. Evol.* **71**, 100-110 (2010).
97 Saladino, R., Barontini, M., Cossetti, C., Di Mauro, E. & Crestini, C. The Effects of Borate Minerals on the Synthesis of Nucleic Acid Bases, Amino Acids and Biogenic Carboxylic Acids from Formamide. *Orig. Life Evol. Biosph.* **41**, 317-330 (2011).

98 Shanker, U., Bhushan, B., Bhattacharjee, G. & Kamaluddin. Formation of Nucleobases from Formamide in the Presence of Iron Oxides: Implication in Chemical Evolution and Origin of Life. *Astrobiology* **11**, 225-233 (2011).
99 Saladino, R. *et al*. Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *Proc. Natl. Acad. Sci.* **112**, E2746-E2755 (2015).
100 Barks , H. L. *et al*. Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation. *ChemBioChem* **11**, 1240-1243 (2010).
101 Hudson, J. S. *et al*. A Unified Mechanism for Abiotic Adenine and Purine Synthesis in Formamide. *Angew. Chem. Int. Ed.* **51**, 5134-5137 (2012).
102 Fox, S. W. & Harada, K. Synthesis of Uracil under Conditions of a Thermal Model of Prebiological Chemistry. *Science* **133**, 1923-1924 (1961).
103 Lagoja, I. M. & Herdewijn, P. One-Step Synthesis of Hypoxanthine from Glycinamide and Diformylurea. *Chem. Biodiversity* **1**, 106-111 (2004).
104 Lagoja, I. M. & Herdewijn, P. A Potential Prebiotic Route to Adenine from Hypoxanthine. *Chem. Biodiversity* **2**, 923-927 (2005).
105 Chittenden, G. J. F. & Schwartz, A. W. Possible pathway for prebiotic uracil synthesis by photodehydrogenation. *Nature* **263**, 350-351 (1976).
106 Schwartz, A. W. & Chittenden, G. J. F. Synthesis of uracil and thymine under simulated prebiotic conditions. *Biosystems* **9**, 87-92 (1977).
107 Butlerow, A. Formation Synthétique D'une Substance Sucreé. *CR Acad. Sci.* **53**, 145-147 (1861).
108 Butlerow, A. Bildung einer zuckerartigen Substanz durch Synthese. *Annal. Chem.* **120**, 295-298 (1861).
109 Breslow, R. On the mechanism of the formose reaction. *Tetrahedron Lett.* **1**, 22-26 (1959).
110 Zubay, G. Studies on the Lead-Catalyzed Synthesis of Aldopentoses. *Orig. Life Evol. Biosph.* **28**, 13-26 (1998).
111 Zubay, G. & Mui, T. Prebiotic Synthesis of Nucleotides. *Orig. Life Evol. Biosph.* **31**, 87-102 (2001).
112 Prieur, B. E. Étude de l'activité prébiotique potentielle de l'acide borique. *Comptes Rendus de l'Académie des Sciences - Series IIC - Chemistry* **4**, 667-670 (2001).
113 Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate Minerals Stabilize Ribose. *Science* **303**, 196-196 (2004).
114 Lambert, J. B., Lu, G., Singer, S. R. & Kolb, V. M. Silicate Complexes of Sugars in Aqueous Solution. *J. Am. Chem. Soc.* **126**, 9611-9625 (2004).
115 Krishnamurthy, R., Arrhenius, G. & Eschenmoser, A. Formation of Glycolaldehyde Phosphate from Glycolaldehyde in Aqueous Solution. *Orig. Life Evol. Biosph.* **29**, 333-354 (1999).
116 Müller, D. *et al*. Chemie von a-Aminonitrilen. Aldomerisierung von Glycolaldehyd-phosphat zu racemischen Hexose-2,4,6-triphosphaten und (in Gegenwart von Formaldehyd) racemischen Pentose-2,4-diphosphaten: rac-Allose-2,4,6-triphosphat und rac-Ribose-2,4-diphosphat sind die Reaktionshauptprodukte. *Helv. Chim. Acta* **73**, 1410-1468 (1990).

117 Sagi, V. N., Punna, V., Hu, F., Meher, G. & Krishnamurthy, R. Exploratory Experiments on the Chemistry of the “Glyoxylate Scenario”: Formation of Ketosugars from Dihydroxyfumarate. *J. Am. Chem. Soc.* **134**, 3577-3589 (2012).
118 Breslow, R. & Appayee, C. Transketolase reaction under credible prebiotic conditions. *Proc. Natl. Acad. Sci.* **110**, 4184-4187 (2013).
119 Kofoed, J., Reymond, J.-L. & Darbre, T. Prebiotic carbohydrate synthesis: zinc–proline catalyzes direct aqueous aldol reactions of α-hydroxy aldehydes and ketones. *Org. Biomol. Chem.* **3**, 1850-1855 (2005).
120 Tsujino, Y., Wakai, C., Matubayashi, N. & Nakahara, M. Noncatalytic Cannizzaro-type Reaction of Formaldehyde in Hot Water. *Chem. Lett.* **28**, 287-288 (2003).
121 Peng, X. D. & Barteau, M. A. Adsorption of formaldehyde on model magnesia surfaces: evidence for the Cannizzaro reaction. *Langmuir* **5**, 1051-1056 (1989).
122 Tambawala, H. & Weiss, A. H. Homogeneously catalyzed formaldehyde condensation to carbohydrates: II. Instabilities and Cannizzaro effects. *J. Catal.* **26**, 388-400 (1972).
123 Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67**, 25-33 (1972).
124 Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. VII. Solid state synthesis of purine nucleosides. *J. Mol. Evol.* **1**, 249-257 (1972).
125 Ingar, A.-A., Luke, R. W. A., Hayter, B. R. & Sutherland, J. D. Synthesis of Cytidine Ribonucleotides by Stepwise Assembly of the Heterocycle on a Sugar Phosphate. *ChemBioChem* **4**, 504-507 (2003).
126 Kim, H.-J. & Benner, S. A. Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. *Proc. Natl. Acad. Sci.* **114**, 11315-11320 (2017).
127 Kim, H.-J. & Kim, J. A Prebiotic Synthesis of Canonical Pyrimidine and Purine Ribonucleotides. *Astrobiology* **19**, 669-674 (2019).
128 Becker, S. *et al*. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833-836 (2016).
129 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).
130 Becker, S. *et al*. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76-82 (2019).
131 Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis: V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol.* **47**, 531-543 (1970).
132 Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239-242 (2009).
133 Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct Assembly of Nucleoside Precursors from Two- and Three-Carbon Units. *Angew. Chem. Int. Ed.* **45**, 6176-6179 (2006).
134 Anastasi, C., Crowe, M. A. & Sutherland, J. D. Two-Step Potentially Prebiotic Synthesis of α-d-Cytidine-5‘-phosphate from d-Glyceraldehyde-3-phosphate. *J. Am. Chem. Soc.* **129**, 24-25 (2007).
135 Fernández-García, C., Grefenstette, N. M. & Powner, M. W. Selective aqueous acetylation controls the photoanomerization of α-cytidine-5'-phosphate. *Chem. Commun.* **54**, 4850-4853 (2018).

136 Stairs, S. *et al*. Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat*. *Commun*. **8**, 15270 (2017).
137 Su, M. *et al*. Triplet-Encoded Prebiotic RNA Aminoacylation. *J*. *Am*. *Chem*. *Soc*. **145**, 15971-15980 (2023).
138 Węgrzyn, E. *et al*. RNA-Templated Peptide Bond Formation Promotes L-Homochirality. *Angew*. *Chem*. *Int*. *Ed*. **63**, e202319235 (2024).
139 Feldmann, J. *et al*. A Unifying Concept for the Prebiotic Formation of RNA Pyrimidine Nucleosides. *Chem*. *Eur*. **1**, e202300013 (2023).
140 Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738 (1953).
141 Cobb, M. & Comfort, N. What Rosalind Franklin truly contributed to the discovery of DNA's structure. *Nature* **616**, 657-660 (2023).
142 CSHL Archives Repository. *On Protein Synthesis*. accessed on 29 May 2024. <https://libgallery.cshl.edu/items/show/52220>.
143 Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563 (1970).
144 Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465-523 (1971).
145 Kruger, K. *et al*. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147-157 (1982).
146 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
147 Rich, A. in *Horizons In Biochemistry*. *On the problems of evolution and biochemical information transfer* (eds M. Kasha & B. Pullman) 103-126 (Academic Press, 1962).
148 Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618-618 (1986).
149 Dworkin, J. P., Lazcano, A. & Miller, S. L. The roads to and from the RNA world. *J*. *Theor*. *Biol*. **222**, 127-134 (2003).
150 Wimberly, B. T. *et al*. Structure of the 30S ribosomal subunit. *Nature* **407**, 327-339 (2000).
151 Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **289**, 905-920 (2000).
152 Harms, J. *et al*. High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. *Cell* **107**, 679-688 (2001).
153 Yonath, A. E. *et al*. Crystallization of the large ribosomal subunits from Bacillus stearothermophilus. *Biochem*. *Int*. **1**, 428-435 (1980).
154 Noller, H. F., Hoffarth, V. & Zimniak, L. Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures. *Science* **256**, 1416-1419 (1992).
155 Xu, D. & Wang, Y. Protein-free ribosomal RNA scaffolds can assemble poly-lysine oligos from charged tRNA fragments. *Biochem*. *Biophys*. *Res*. *Commun*. **544**, 81-85 (2021).
156 Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc*. *Natl*. *Acad*. *Sci*. **74**, 5088-5090 (1977).
157 Coenye, T. & Vandamme, P. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol*. *Lett*. **228**, 45-49 (2003).

158 Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87**, 4576-4579 (1990).
159 Tuerk, C. & Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **249**, 505-510 (1990).
160 Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
161 Beaudry, A. A. & Joyce, G. F. Directed Evolution of an RNA Enzyme. *Science* **257**, 635-641 (1992).
162 Pan, T. & Uhlenbeck, O. C. In vitro selection of RNAs that undergo autolytic cleavage with lead(2+). *Biochemistry* **31**, 3887-3895 (1992).
163 Pan, T. & Uhlenbeck, O. C. A small metalloribozyme with a two-step mechanism. *Nature* **358**, 560-563 (1992).
164 Ellington, A. D., Chen, X., Robertson, M. & Syrett, A. Evolutionary origins and directed evolution of RNA. *Int. J. Biochem. Cell Biol.* **41**, 254-265 (2009).
165 Joyce, G. F. RNA evolution and the origins of life. *Nature* **338**, 217-224 (1989).
166 Been, M. D. & Cech, T. R. RNA as an RNA Polymerase: Net Elongation of an RNA Primer Catalyzed by the Tetrahymena Ribozyme. *Science* **239**, 1412-1416 (1988).
167 Doudna, J. A. & Szostak, J. W. RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**, 519-522 (1989).
168 Bartel, D. P. & Szostak, J. W. Isolation of New Ribozymes from a Large Pool of Random Sequences. *Science* **261**, 1411-1418 (1993).
169 Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. *Science* **269**, 364-370 (1995).
170 Ekland, E. H. & Bartel, D. P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238 (1995).
171 Ekland, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373-376 (1996).
172 Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. *Science* **292**, 1319-1325 (2001).
173 Attwater, J., Wochner, A. & Holliger, P. In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* **5**, 1011-1018 (2013).
174 Tjhung, K. F., Shokhirev, M. N., Horning, D. P. & Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci.*, 201914282 (2020).
175 Papastavrou, N., Horning, D. P. & Joyce, G. F. RNA-catalyzed evolution of catalytic RNA. *Proc. Natl. Acad. Sci.* **121**, e2321592121 (2024).
176 Brown, R. S., Dewan, J. C. & Klug, A. Crystallographic and biochemical investigation of the lead(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry* **24**, 4785-4801 (1985).
177 Breslow, R. & Huang, D. L. Effects of metal ions, including Mg^{2+} and lanthanides, on the cleavage of ribonucleotides and RNA model compounds. *Proc. Natl. Acad. Sci.* **88**, 4080-4083 (1991).
178 Pyle, A. M. Ribozymes: A Distinct Class of Metalloenzymes. *Science* **261**, 709-714 (1993).

179 Kazakov, S. & Altman, S. A trinucleotide can promote metal ion-dependent specific cleavage of RNA. *Proc. Natl. Acad. Sci.* **89**, 7939-7943 (1992).
180 Wachowius, F. & Holliger, P. Non-Enzymatic Assembly of a Minimized RNA Polymerase Ribozyme. *ChemSystemsChem* **1**, 1-4 (2019).
181 Le Vay, K. & Mutschler, H. The difficult case of an RNA-only origin of life. *Emerg. Top Life Sci.* **3**, 469-475 (2019).
182 Ferreira, I., Jolley, E. A., Znosko, B. M. & Weber, G. Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters. *Chem. Phys.* **521**, 69-76 (2019).
183 Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E. & Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *Elife* **7**, e35255 (2018).
184 Kristoffersen, E. L., Burman, M., Noy, A. & Holliger, P. Rolling circle RNA synthesis catalyzed by RNA. *Elife* **11**, e75186 (2022).
185 Attwater, J. *et al*. Trinucleotide building blocks enable exponential ribozyme-catalysed RNA replication and open-ended growth of diverse RNA sequence pools. *bioRxiv*, 2023.2003.2017.533225 (2023).
186 Lohrmann, R. Formation of nucleoside 5'-phosphoramidates under potentially prebiological conditions. *J. Mol. Evol.* **10**, 137-154 (1977).
187 Mariani, A., Russell, D. A., Javelle, T. & Sutherland, J. D. A Light-Releasable Potentially Prebiotic Nucleotide Activating Agent. *J. Am. Chem. Soc.* **140**, 8657-8661 (2018).
188 Yi, R., Hongo, Y. & Fahrenbach, A. C. Synthesis of imidazole-activated ribonucleotides using cyanogen chloride. *Chem. Commun.* **54**, 511-514 (2018).
189 Leslie E, O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99-123 (2004).
190 Sawai, H. & Orgel, L. E. Oligonucleotide synthesis catalyzed by the zinc(2+) ion. *J. Am. Chem. Soc.* **97**, 3532-3533 (1975).
191 Sawai, H. Catalysis of internucleotide bond formation by divalent metal ions. *J. Am. Chem. Soc.* **98**, 7037-7039 (1976).
192 Sleeper, H. L. & Orgel, L. E. The catalysis of nucleotide polymerization by compounds of divalent lead. *J. Mol. Evol.* **12**, 357-364 (1979).
193 Monnard, P.-A., Kanavarioti, A. & Deamer, D. W. Eutectic Phase Polymerization of Activated Ribonucleotide Mixtures Yields Quasi-Equimolar Incorporation of Purine and Pyrimidine Nucleobases. *J. Am. Chem. Soc.* **125**, 13734-13740 (2003).
194 Kanavarioti, A., Monnard, P.-A. & Deamer, D. W. Eutectic Phases in Ice Facilitate Nonenzymatic Nucleic Acid Synthesis. *Astrobiology* **1**, 271-281 (2001).
195 Sawai, H., Kuroda, K. & Hojo, T. Uranyl Ion as a Highly Effective Catalyst for Internucleotide Bond Formation. *Bull. Chem. Soc. Jpn.* **62**, 2018-2023 (2006).
196 Sawai, H., Higa, K. & Kuroda, K. Synthesis of cyclic and acyclic oligocytidylates by uranyl ion catalyst in aqueous solution. *J. Chem. Soc. Perkin 1*, 505-508 (1992).
197 Huang, W. & Ferris, J. P. One-Step, Regioselective Synthesis of up to 50-mers of RNA Oligomers by Montmorillonite Catalysis. *J. Am. Chem. Soc.* **128**, 8914-8919 (2006).
198 Walton, T., Zhang, W., Li, L., Tam, C. P. & Szostak, J. W. The Mechanism of Nonenzymatic Template Copying with Imidazole-Activated Nucleotides. *Angew. Chem. Int. Ed.* **58**, 10812-10819 (2019).

199 O'Flaherty, D. K., Zhou, L. & Szostak, J. W. Nonenzymatic Template-Directed Synthesis of Mixed-Sequence 3'-NP-DNA up to 25 Nucleotides Long Inside Model Protocells. *Journal of the American Chemical Society* **141**, 10481-10488 (2019).
200 Laughrea, M. Speed-accuracy relationships during in vitro and in vivo protein biosynthesis. *Biochimie* **63**, 145-168 (1981).
201 Abraham, A. K. in *Progress in Nucleic Acid Research and Molecular Biology*. Vol. 28 (ed Waldo E. Cohn) 81-100 (Academic Press, 1983).
202 Ogle, J. M. & Ramakrishnan, V. STRUCTURAL INSIGHTS INTO TRANSLATIONAL FIDELITY. *Annu. Rev. Biochem.* **74**, 129-177 (2005).
203 Lee, J. W. *et al*. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443**, 50-55 (2006).
204 Beebe, K., Mock, M., Merriman, E. & Schimmel, P. Distinct domains of tRNA synthetase recognize the same base pair. *Nature* **451**, 90-93 (2008).
205 Chapeville, F. *et al*. On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci.* **48**, 1086-1092 (1962).
206 Hopfield, J. J. Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proc. Natl. Acad. Sci.* **75**, 4334-4338 (1978).
207 Black, R. A. & Blosser, M. C. A Self-Assembled Aggregate Composed of a Fatty Acid Membrane and the Building Blocks of Biological Polymers Provides a First Step in the Emergence of Protocells. *Life* **6**, 33 (2016).
208 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).
209 Müller, F. *et al*. A prebiotically plausible scenario of an RNA–peptide world. *Nature* **605**, 279-284 (2022).
210 Boudou, V. *et al*. Synthesis of the Anticodon Hairpin $^{t}RNA_{f}^{Met}$ Containing N-{[9-(β-D-Ribofuranosyl)-9H-purin-6-yl]carbamoyl}-L-threonine (=N^{6}-{{[(1S,2R)-1-Carboxy-2-hydroxypropyl]amino}carbonyl}adenosine, $t^{6}A$). *Helv. Chim. Acta* **83**, 152-161 (2000).
211 Mitsunobu, O. & Yamada, M. Preparation of Esters of Carboxylic and Phosphoric Acid via Quaternary Phosphonium Salts. *Bull. Chem. Soc. Jpn.* **40**, 2380-2382 (1967).
212 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).
213 Ogilvie, K. K. & Hakimelahi, G. H. A general method for selective silylation of primary hydroxyl groups in carbohydrates and related compounds. *Carbohydr. Res.* **115**, 234-239 (1983).
214 Sekine, M., Iimura, S. & Furusawa, K. Synthesis of a new class of 2'-phosphorylated oligoribonucleotides capable of conversion to oligoribonucleotides. *J. Org. Chem.* **58**, 3204-3208 (1993).
215 Schaller, H., Weimann, G., Lerch, B. & Khorana, H. G. Studies on Polynucleotides. XXIV.[1]The Stepwise Synthesis of Specific Deoxyribopolynucleotides (4).[2]Protected Derivatives of Deoxyribonucleosides and New Syntheses of Deoxyribonucleoside-3" Phosphates[3]. *J. Am. Chem. Soc.* **85**, 3821-3827 (1963).

216 Parthasarathy, R., Ohrt, J. M. & Chheda, G. B. Conformation and possible role of hypermodified nucleosides adjacent to 3'-end of anticodon in tRNA: N-(purin-6-ylcarbamoyl)-L-threonine riboside. *Biochem. Biophys. Res. Commun.* **60**, 211-218 (1974).
217 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).
218 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).
219 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).
220 Grover, R. K. *et al.* O-Glycoside Orientation Is an Essential Aspect of Base J Recognition by the Kinetoplastid DNA-Binding Protein JBP1. *Angew. Chem. Int. Ed.* **46**, 2839-2843 (2007).
221 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).
222 Gartner, Z. J., Kanan, M. W. & Liu, D. R. Multistep Small-Molecule Synthesis Programmed by DNA Templates. *J. Am. Chem. Soc.* **124**, 10304-10306 (2002).
223 Liu, Z. *et al.* Harnessing chemical energy for the activation and joining of prebiotic building blocks. *Nat. Chem.* **12**, 1-6 (2020).
224 Canavelli, P., Islam, S. & Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **571**, 546-549 (2019).
225 Singer, J. N. *et al.* Loading of Amino Acids onto RNA in a Putative RNA-Peptide World. *Angew. Chem. Int. Ed.* **62**, e202302360 (2023).
226 Rubio Gomez, M. A. & Ibba, M. Aminoacyl-tRNA synthetases. *RNA* **26**, 910-936 (2020).
227 Chan, C. Y., Singer, J. N. & Carell, T. Membrane bound geranylated RNAs establish a primitive peptide synthesis system. *bioRxiv*, 2024.2008.2002.606298 (DOI:606210.601101/602024.606208.606202.606298) (2024).
228 Leszczynska, G. *et al.* Reaction of S-geranyl-2-thiouracil modified oligonucleotides with alkyl amines leads to the N2-alkyl isocytosine derivatives. *Org. Biomol. Chem.* **15**, 5332-5336 (2017).
229 Sierant, M. *et al.* S-Geranyl-2-thiouridine wobble nucleosides of bacterial tRNAs; chemical and enzymatic synthesis of S-geranylated-RNAs and their physicochemical characterization. *Nucleic Acids Res.* **44**, 10986-10998 (2016).
230 Leszczynska, G., Sadowska, K., Bartos, P., Nawrot, B. & Sochacka, E. S-Geranylated 2-Thiouridines of Bacterial tRNAs: Chemical Synthesis and Physicochemical Properties. *European J. Org. Chem.* **2016**, 3482-3485 (2016).
231 Wang, R., Haruehanroengra, P. & Sheng, J. Synthesis of Geranyl-2-Thiouridine-Modified RNA. *Curr. Protoc. Nucleic Acid Chem.* **68**, 4.72.71-74.72.13 (2017).

232 Haruehanroengra, P. *et al*. Terpene Chain Length Affects the Base Pairing Discrimination of S-geranyl-2-thiouridine in RNA Duplex. *iScience* **23**, 101866 (2020).
233 Zhu, T. F., Budin, I. & Szostak, J. W. in *Methods Enzymol*. Vol. 533 (ed Jon Lorsch) 267-274 (Academic Press, 2013).
234 Sarac, I. & Meier, C. Solid-Phase Synthesis of DNA and RNA 5'-O-Triphosphates Using cycloSal Chemistry. *Curr. Protoc. Nucleic Acid Chem.* **64**, 4.67.61-64.67.13 (2016).
235 Sarac, I. & Meier, C. Efficient Automated Solid-Phase Synthesis of DNA and RNA 5'-Triphosphates. *Chemistry – A European Journal* **21**, 16421-16426 (2015).
236 Attwater, J. *et al*. Chemical fidelity of an RNA polymerase ribozyme. *Chem. Sci.* **4**, 2804-2814 (2013).
237 Shechner, D. M. *et al*. Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme. *Science* **326**, 1271-1275 (2009).
238 Joo, S. H. Cyclic peptides as therapeutic agents and biochemical tools. *Biomol. Ther. (Seoul)* **20**, 19-26 (2012).
239 Abdalla, M. A. & McGaw, L. J. Natural Cyclic Peptides as an Attractive Modality for Therapeutics: A Mini Review. *Molecules* **23**, 2080 (2018).
240 Sarojini, V., Cameron, A. J., Varnava, K. G., Denny, W. A. & Sanjayan, G. Cyclic Tetrapeptides from Nature and Design: A Review of Synthetic Methodologies, Structure, and Function. *Chem. Rev.* **119**, 10318-10359 (2019).
241 Kubik, S. & Goddard, R. Fine Tuning of the Cation Affinity of Artificial Receptors Based on Cyclic Peptides by Intramolecular Conformational Control. *European J. Org. Chem.* **2001**, 311-322 (2001).
242 Kubik, S. & Goddard, R. A New Cyclic Pseudopeptide Composed of (l)-Proline and 3-Aminobenzoic Acid Subunits as a Ditopic Receptor for the Simultaneous Complexation of Cations and Anions. *J. Org. Chem.* **64**, 9475-9486 (1999).
243 Wiesner, M., Revell, J. D. & Wennemers, H. Tripeptides as Efficient Asymmetric Catalysts for 1,4-Addition Reactions of Aldehydes to Nitroolefins–A Rational Approach. *Angew. Chem. Int. Ed.* **47**, 1871-1874 (2008).
244 Wiesner, M., Neuburger, M. & Wennemers, H. Tripeptides of the Type H-D-Pro-Pro-Xaa-NH2 as Catalysts for Asymmetric 1,4-Addition Reactions: Structural Requirements for High Catalytic Efficiency. *Chem. Eur. J.* **15**, 10103-10109 (2009).
245 Duschmalé, J., Kohrt, S. & Wennemers, H. Peptide catalysis in aqueous emulsions. *Chem. Commun.* **50**, 8109-8112 (2014).
246 Wiesner, M., Upert, G., Angelici, G. & Wennemers, H. Enamine Catalysis with Low Catalyst Loadings - High Efficiency via Kinetic Studies. *J. Am. Chem. Soc.* **132**, 6-7 (2010).
247 Rigling, C. *et al*. Conformational Properties of a Peptidic Catalyst: Insights from NMR Spectroscopic Studies. *J. Am. Chem. Soc.* **140**, 10829-10838 (2018).
248 Cappannini, A. *et al*. MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Res.* **52**, D239-D244 (2023).
249 Lei, H.-T. *et al*. tModBase: deciphering the landscape of tRNA modifications and their dynamic changes from epitranscriptome data. *Nucleic Acids Res.* **51**, D315-D327 (2023).
250 McCown, P. J. *et al*. Naturally occurring modified ribonucleosides. *WIREs RNA* **11**, e1595 (2020).

251 Machnicka, M. A. *et al*. MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res*. **41**, D262-D267 (2012).
252 Agris, P. F., Narendran, A., Sarachan, K., Väre, V. Y. P. & Eruysal, E. in *The Enzymes*. Vol. 41 (ed Guillaume F. Chanfreau) 1-50 (Academic Press, 2017).
253 Valadon, C. & Namy, O. The Importance of the Epi-Transcriptome in Translation Fidelity. *Non-Coding RNA* **7**, 51 (2021).
254 Yared, M.-J., Marcelot, A. & Barraud, P. Beyond the Anticodon: tRNA Core Modifications and Their Impact on Structure, Translation and Stress Adaptation. *Genes* **15**, 374 (2024).
255 Chan, C. T. Y. *et al*. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun*. **3**, 937 (2012).
256 Huang, H.-Y. & Hopper, A. K. Multiple Layers of Stress-Induced Regulation in tRNA Biology. *Life* **6**, 16 (2016).
257 Giegé, R. & Eriani, G. The tRNA identity landscape for aminoacylation and beyond. *Nucleic Acids Res*. **51**, 1528-1570 (2023).
258 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).
259 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).
260 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 threonin transfer ribonucleic acid. *Biochemistry* **11**, 3910-3915 (1972).
261 Bujnicki, J. M. *et al*. Identification of a bifunctional enzyme MnmC involved in the biosynthesis of a hypermodified uridine in the wobble position of tRNA. *RNA* **10**, 1236-1242 (2004).
262 Hagervall, T. G., Edmonds, C. G., McCloskey, J. A. & Björk, G. R. Transfer RNA(5-methylaminomethyl-2-thiouridine)-methyltransferase from Escherichia coli K-12 has two enzymatic activities. *J. Biol. Chem*. **262**, 8488-8495 (1987).
263 Kim, J. & Almo, S. C. Structural basis for hypermodification of the wobble uridine in tRNA by bifunctional enzyme MnmC. *BMC Struct. Biol*. **13**, 5 (2013).
264 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*. (2021).
265 Jash, B., Tremmel, P., Jovanovic, D. & Richert, C. Single nucleotide translation without ribosomes. *Nat. Chem*. **13**, 751-757 (2021).
266 Hentzen, D., Mandel, P. & Garel, J.-P. Relation between aminoacyl-tRNA stability and the fixed amino acid. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **281**, 228-232 (1972).
267 Peacock, J. R. *et al*. Amino acid–dependent stability of the acyl linkage in aminoacyl-tRNA. *RNA* **20**, 758-764 (2014).
268 Werner, E. A. *The Chemistry of Urea, the Theory of Its Constitution, and of the Origin and Mode of Its Formation in Living Organisms*. (Longmans, 1923).
269 Shaw, W. H. R. & Bordeaux, J. J. The Decomposition of Urea in Aqueous Media. *J. Am. Chem. Soc*. **77**, 4729-4733 (1955).

270 Robertson, M. P. & Miller, S. L. Prebiotic Synthesis of 5-Substituted Uracils: A Bridge Between the RNA World and the DNA-Protein World. *Science* **268**, 702-705 (1995).
271 Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859-1862 (1981).
272 Roy, S. & Caruthers, M. Synthesis of DNA/RNA and Their Analogs via Phosphoramidite and H-Phosphonate Chemistries. *Molecules* **18**, 14268-14284 (2013).
273 Cavalier-Smith, T. Obcells as Proto-Organisms: Membrane Heredity, Lithophosphorylation, and the Origins of the Genetic Code, the First Cells, and Photosynthesis. *J. Mol. Evol.* **53**, 555-595 (2001).
274 O'Connor, P. B. F. The Evolutionary Transition of the RNA World to Obcells to Cellular-Based Life. *J. Mol. Evol.* (2024).
275 Budker, V. G., Kazatchkov, Y. A. & Naumova, L. P. Polynucleotides adsorb on mitochondrial and model lipid membranes in the presence of bivalent cations. *FEBS Lett.* **95**, 143-146 (1978).
276 Michanek, A., Kristen, N., Höök, F., Nylander, T. & Sparr, E. RNA and DNA interactions with zwitterionic and charged lipid membranes — A DSC and QCM-D study. *BBA. Biomembranes* **1798**, 829-838 (2010).
277 Kamat, N. P., Tobé, S., Hill, I. T. & Szostak, J. W. Electrostatic Localization of RNA to Protocell Membranes by Cationic Hydrophobic Peptides. *Angew. Chem. Int. Ed.* **54**, 11735-11739 (2015).
278 Izgu, E. C. *et al.* N-Carboxyanhydride-Mediated Fatty Acylation of Amino Acids and Peptides for Functionalization of Protocell Membranes. *J. Am. Chem. Soc.* **138**, 16669-16676 (2016).
279 Bonfio, C., Russell, D. A., Green, N. J., Mariani, A. & Sutherland, J. D. Activation chemistry drives the emergence of functionalised protocells. *Chem. Sci.* **11**, 10688-10697 (2020).
280 Lu, D. & Rhodes, D. G. Binding of phosphorothioate oligonucleotides to zwitterionic liposomes. *BBA. Biomembranes* **1563**, 45-52 (2002).
281 Vlassov, A., Khvorova, A. & Yarus, M. Binding and disruption of phospholipid bilayers by supramolecular RNA complexes. *Proc. Natl. Acad. Sci.* **98**, 7706-7711 (2001).
282 Janas, T., Janas, T. & Yarus, M. Specific RNA binding to ordered phospholipid bilayers. *Nucleic Acids Res.* **34**, 2128-2136 (2006).
283 Dumelin, C. E., Chen, Y., Leconte, A. M., Chen, Y. G. & Liu, D. R. Discovery and biological characterization of geranylated RNA in bacteria. *Nat. Chem. Biol.* **8**, 913-919 (2012).
284 Xu, J. *et al.* A prebiotically plausible synthesis of pyrimidine β-ribonucleosides and their phosphate derivatives involving photoanomerization. *Nat. Chem.* **9**, 303-309 (2017).
285 Aleksandrova, M., Rahmatova, F., Russell, D. A. & Bonfio, C. Ring Opening of Glycerol Cyclic Phosphates Leads to a Diverse Array of Potentially Prebiotic Phospholipids. *J. Am. Chem. Soc.* (2023).
286 McRae, E. K. S. *et al.* Cryo-EM structure and functional landscape of an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci.* **121**, e2313332121 (2024).

287 Woese, C. R. On the evolution of the genetic code. *Proc. Natl. Acad. Sci.* **54**, 1546-1552 (1965).
288 Vetsigian, K., Woese, C. & Goldenfeld, N. Collective evolution and the genetic code. *Proc. Natl. Acad. Sci.* **103**, 10696-10701 (2006).
289 Lecompte, O., Ripp, R., Thierry, J. C., Moras, D. & Poch, O. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Research* **30**, 5382-5390 (2002).
290 Kurland, C. G. Ribosome Structure and Function Emergent. *Science* **169**, 1171-1177 (1970).
291 Fox, G. E. Origin and evolution of the ribosome. *Cold Spring Harbor perspectives in biology* **2**, a003483-a003483 (2010).
292 Barbieri, M. Evolution of the genetic code: The ambiguity-reduction theory. *Biosystems* **185**, 104024 (2019).
293 Burks, A. W. *Essays on Cellular Automata*. (University of Illinois Press, 1970).
294 Novozhilov, A. S., Wolf, Y. I. & Koonin, E. V. Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biology Direct* **2**, 24 (2007).
295 Gilis, D., Massar, S., Cerf, N. J. & Rooman, M. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* **2**, research0049.0041 (2001).
296 Itzkovitz, S. & Alon, U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res.* **17**, 405-412 (2007).
297 Drummond, D. A. & Wilke, C. O. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell* **134**, 341-352 (2008).
298 Foden, C. S. *et al.* Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865-869 (2020).
299 Wu, T. & Orgel, L. E. Nonenzymic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *Journal of the American Chemical Society* **114**, 5496-5501 (1992).
300 Wu, T. & Orgel, L. E. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. *Journal of the American Chemical Society* **114**, 7963-7969 (1992).
301 Maury, C. P. J. Amyloid and the origin of life: self-replicating catalytic amyloids as prebiotic informational and protometabolic entities. *Cellular and Molecular Life Sciences* **75**, 1499-1507 (2018).
302 Woese, C. R., Olsen, G. J., Ibba, M. & Söll, D. Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process. *Microbiology and Molecular Biology Reviews* **64**, 202-236 (2000).
303 Maizels, N. & Weiner, A. M. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl. Acad. Sci.* **91**, 6729-6734 (1994).
304 Petrov, A. S. *et al.* History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci.* **112**, 15396-15401 (2015).
305 Ninio, J. *Molecular approaches to evolution*. Vol. 762 (Princeton University Press, 2014).
306 Gamow, G. Possible Relation between Deoxyribonucleic Acid and Protein Structures. *Nature* **173**, 318-318 (1954).

307 Woese, C. R., Dugre, D. H., Saxinger, W. C. & Dugre, S. A. The molecular basis for the genetic code. *Proc. Natl. Acad. Sci.* **55**, 966-974 (1966).
308 Szathmáry, E. The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* **15**, 223-229 (1999).
309 Wolf, Y. I. & Koonin, E. V. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct* **2**, 14 (2007).
310 Riddle, D. L. & Carbon, J. Frameshift Suppression: a Nucleotide Addition in the Anticodon of a Glycine Transfer RNA. *Nature New Biology* **242**, 230-234 (1973).
311 O'Connor, M. Insertions in the anticodon loop of $tRNA_1^{Gln}$(sufG) and $tRNA^{Lys}$ promote quadruplet decoding of CAAA. *Nucleic Acids Research* **30**, 1985-1990 (2002).
312 Yourno, J. & Kohno, T. Externally Suppressible Proline Quadruplet CCC^U. *Science* **175**, 650-652 (1972).
313 de Duve, C. The second genetic code. *Nature* **333**, 117-118 (1988).
314 Eritja, R., Robles, J., Aviñó, A., Alberico, F. & Pedroso, E. A synthetic procedure for the preparation of oligonucleotides without using ammonia and its application for the synthesis of oligonucleotides containing 0-4-alkyl thymidines. *Tetrahedron* **48**, 4171-4182 (1992).
315 van der Wenden, E. M. *et al.* Ribose-Modified Adenosine Analogs as Potential Partial Agonists for the Adenosine Receptor. *J. Med. Chem.* **38**, 4000-4006 (1995).
316 Zlatev, I., Manoharan, M., Vasseur, J.-J. & Morvan, F. Solid-Phase Chemical Synthesis of 5'-Triphosphate DNA, RNA, and Chemically Modified Oligonucleotides. *Curr. Protoc. Nucleic Acid Chem.* **50**, 1.28.21-21.28.16 (2012).

Abbreviations

aa^{6}A	6-aminocarbamoyladenosines
Abz	3-aminobenzoic acid
aaRS	amino acid tRNA synthetase(s)
AIBN	azobisisobutyronitrile
AICA	4-aminoimidazole-5-carboxamide
AICN	4-aminoimidazole-5-carbonitrile
AMA	ammonia/methylamine solution
AU/a.u.	arbitrary unit
Boc	*t*-butyloxycarbonyl
bp	base pair(s)
BTT	1-benzylthio-1H-tetrazole
CED-Cl	2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite
CPG	controlled pore glass
DAMN	diaminomaleonitrile
DBU	1,8-diazabicyclo(5.4.0)undec-7-ene
DCM	dichloromethane
DIAD	diisopropoyl azodicarboxylate
Diox	1,4-dioxane
DIPA	*N*,*N*-diisopropylamine
DIPEA	*N*,*N*-diisopropyl-*N*-ethylamine

DLS	dynamic light scattering
DMF	*N*,*N*-dimethylformamide
DMT	4,4'-dimethoxytrityl
DMTCl	4,4'-dimethoxytritylchloride
DMTMM	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium
DNA	deoxyribonucleic acid
dsRNA	double strand RNA/RNA duplex
DTBS	di-*tert*-butylsilyl
DTT	dithiolthreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
Egg PC	phosphatidylcholine lipids extracted from chicken egg
FAM	fluorescein amidite
Ga	billion years ago
g^6A	N^6-glycinylcarbamoyladenosine
GC	gas chromatography
geBr	geranylbromide
ges^2U	*S*-geranyl-2-thiouridine
$gmnm^5U$	glycine-coupled 5-methylaminomethyluridine
HBTU	hexafluorophosphate Benzotriazole Tetramethyl Uronium
HDMS	1,1,1,3,3,3,-hexamethyldisilazane
HOBt	hydroxybenzotriazole
HPA	3-hydroxypicolinic acid
HPLC	high performance liquid chromatography
*i*Pr	isopropyl
*i*PrOH	isopropanol
iso-C	isocytidine
iso-m^2C	N^2-methylisocytidine
MALDI	matrix assisted laser detected ionization
m^6A	6-methyladenosine
m^6aa^6A	N^6-methyl-N^6-aminocarbamoyladenosines
MES	2-(N-morpholino)ethanesulfonate
mnm^5U	5-methylaminomethyluridine
MOPS	3-(N-morpholino)propanesulfonate
mRNA	messenger RNA
MS	mass spectroscopy
m/z	mass-to-charge ratio
m^6t^6A	N^6-methyl-N^6-threonylcarbamoyladenosine
NBS	*N*-bromosuccinimide
NEt_3	triethylamine
NMR	nuclear magnetic resonance spectroscopy
nm^5U	5-aminomethyluridine
npe	*p*-nitrophenylethyl
nt.	nucleotide(s)
NTP	nucleoside triphosphate
NNNTP	triplet triphosphate
ON	oligonucleotide
PAGE	polyacrylamide gel electrophoresis
pNPP	*p*-nitrophenylphosphate
POM	pivaloyloxymethyl

Ph	phenyl
Py	pyridine
RNA	ribonucleic acid
ROSINA	Rosetta Orbiter Spectrometer for Ion and Neutral Analysis
rRNA	ribosomal RNA
r.t.	room temperature
ssRNA	single strand RNA
Suc	succinimide
sulfo-NHS	*N*-hydroxysulfosuccinimide
TBAF	tetrabutylammonium fluoride
TBS	*t*-butyldimethylsilyl
TBSCl	*t*-butyldimethylsilylchloride
t-Bu	*t*-butyl
teoc	2-(trimethylsilyl)ethoxycarbonyl
TFA	trifluoroacetic acid
T_m	melting temperature
TMS	trimethylsilyl
TOF	time-of-flight
t_R	retention time
t^6A	N^6-threonylcarbamoyladenosine
t^6dA	N^6-threonylcarbamoyldeoxyadenosine
teoc	(trimethylsilyl)ethoxycarbonyl
THF	tetrahydrofuran
TLC	thin layer chromatography
tRNA	transfer RNA
SELEX	systematic evolution of ligands by exponential enrichment
s^2U	2-thiouridine
UPLC	ultra performance liquid chromatography
UV	ultraviolet
$vmnm^5U$	valine-coupled 5-methylaminomethyluridine
+Lip	in the presence of egg PC liposomes
-Lip	without egg PC liposomes
-ve	negative control
+ve	positive control
⌀	diameter