Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Synthesis of modified oligonucleotides for prebiotic studies and as novel CoV-2 therapeutics

Milda Nainytė

aus

Tauragė, Lithuania

2021

"Reikia tikėti. Reikia labai tikėti,
Kad, ranką iškėlus, iš dangaus imtų krist mana.
Už kiekvieną stebuklą reikia savim užmokėti –
Savo gyvenimu, meile, širdim ir daina."

J. Marcinkevičius, 1966

Dedicated to my family

Summary

Acknowledgements

My first thank goes to my PhD advisor *Prof. Thomas Carell* for the opportunity, support and guidance he has provided throughout these years I spend in his group. I will undoubtedly miss our laboratory discussions about miscellaneous world topics as well as my research advances. I feel delighted that I had a chance to carry out my PhD work in your laboratory, working on intellectually stimulating topics that gave me an opportunity to become experienced in the field of nucleic acid chemistry and to discover scientific interests that I am looking forward to face in the future.

I would like to thank *Prof. Anja Hoffmann-Röder* for accepting to be my second evaluator. The other committee members, *Prof. Oliver Trapp, Prof. Konstantin Karaghiosoff, Dr. Sabine Schneider* and *Dr. Pavel Kielkowski* are also kindly acknowledged.

I would like to thank *Dr. Tynchtyk Amatov* who was a great teacher and coworker. I am thankful for these hard discussions we had that drove our project further. Your passion for organic chemistry was always a big inspiration for me.

I would also like to thank *Dr. Markus Müller* for all interesting discussions we had, as well as your endless support in all organizational issues.

I thank *Frau Slava Gärtner* for the management of every bureaucratic issue and *Kerstin Kurz* for the help in everyday laboratory needs.

I would like to thank *Dr. Markus Müller*, *Felix Müller*, *Jonas Feldmann* for finding time to comment and help with corrections of my PhD thesis.

A special thanks goes to my co-workers in the "prebiotic subgroup": *Felix Müller, Dr. Luis Escobar, Chun-Yin Chan* (aka *Jamie*) and *Felix Xu*. I am very glad that you all joined to push forward the proto-ribosome project that is so important to me.

I also thank my colleagues *Dr. Franziska Traube* and *Ammar Ahmedani* – parts of the thesis would have been impossible without your enthusiasm and cooperation.

In addition, I would like to thank all the people that I met in Carell group for creating a fantastic working environment. I am very happy that I met so many interesting people here.

I also thank my former advisor *Prof. Viktoras Masevičius* for introducing me to organic chemistry and for his infinite belief in me. It was a huge encouragement for me while being far away from home.

Since my support extends beyond the academic setting, I also would like to thank for my whole family, my wonderful boyfriend and my closest friends for your support and trust in me which gave me strength to reach the end of my PhD. You are main influences on whom I am today and great examples for me during my life.

Publications

- **Milda Nainytė**, Tynchtyk Amatov, Thomas Carell, *Chem. Comm*. **2019**, *55*, 12216-12218.
- **Milda Nainyte**, Thomas Carell, *Helv. Chim. Acta* **2020**, *103*, e2000016.
- **Milda Nainytė**, Felix Müller, Giacomo Ganazzoli, Chun-Yin Chan, Antony Crisp, Daniel Globisch, Thomas Carell, *Chem. Eur. J*. **2020**, *26*, 1-6.

Conferences

- Poster "Towards RNA-templated peptide formation utilizing non-canonical nucleobases", **RNA epigenetics in human diseases**, 09/2019, Cambridge, UK.
- Poster "RNA-templated peptide growth enabled by a labile urea linkage", **Physical chemistry at its best – Leopoldina meeting**, 09/2018, Halle, Germany.
- Poster "Towards RNA-templated peptide formation utilizing non-canonical nucleobases", **Science of early life**, 11/2019, Kloster Seeon, Germany.

Other publications

- M. Mickute, **M. Nainyte**, L. Vasiliauskaite, A. Plotnikova, V. Masevicius, S. Klimasauskas, G. Vilkaitis, *Nucleic Acids Res*. **2018**, *46*, 104.
- A. Osipenko, A. Plotnikova, **M. Nainyte**, V. Masevicius, S. Klimasauskas, G. Vilkaitis, *Angew. Chem. Int. Ed*. **2017**, *56*, 6507.
- V. Masevicius, **M. Nainyte**, S. Klimasauskas, *Curr. Protoc. Nucleic Acid Chem*. **2016**, 64:1.36.1-1-36.13. doi: 10. 1002/0471142700.nc0136s64.

1 Abstract

RNA is one of the major macromolecules known to be essential for all forms of life. Although chemically very close to DNA, RNA strands in most cases are much shorter and to a large extent single stranded. RNA molecules are mainly involved in protein synthesis and transcriptional regulation. In order to fulfill their multiple functions, RNA molecules contain more than just the four canonical nucleosides. It is known that all parts of RNA can be modified by the presence of naturally occurring, post-transcriptionally modified nucleosides. These non-canonical nucleobases were identified in all types of RNAs, such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), as well as in messenger RNAs (mRNAs), small nuclear RNAs (snRNAs), microRNAs (miRNAs) and chromosomal RNAs. Hypermodified bases, as particularly strongly altered nucleobases are widely present in tRNAs. Modified nucleosides in total play important roles in gene expression and they regulate many aspects of RNA function. Some of the modified nucleosides have recently gained importance for the development of RNA based therapeutics.

In the first part of this work, the origin of the translation process was investigated following the generally accepted RNA world theory, which proposes that RNA both stored genetic information and catalyzed chemical reactions. Given the currently seen complexity in the interaction between RNA and proteins, we assumed that RNA and proteins at some point coexisted with RNA catalyzing the formation of peptides and the formed peptides helped for the existence and formation of RNA. Amino acid modified nucleobases (AMNs) positioned in close proximity to the anticodon in tRNA are known to be important for translational fidelity. We believe that these bases, which we call "living molecular fossils" are relics of a prebiotic RNA world and that they were once the key components that allowed RNA to template the formation of peptides. In this thesis, a model based on RNA-templated peptide synthesis, depicted in Figure 1, was developed with the help of the amino acid modified RNA nucleobases (k^2C, k^2U, acp^3U, $(m^6)aa^6A$, mnm^5U, nm^5U).
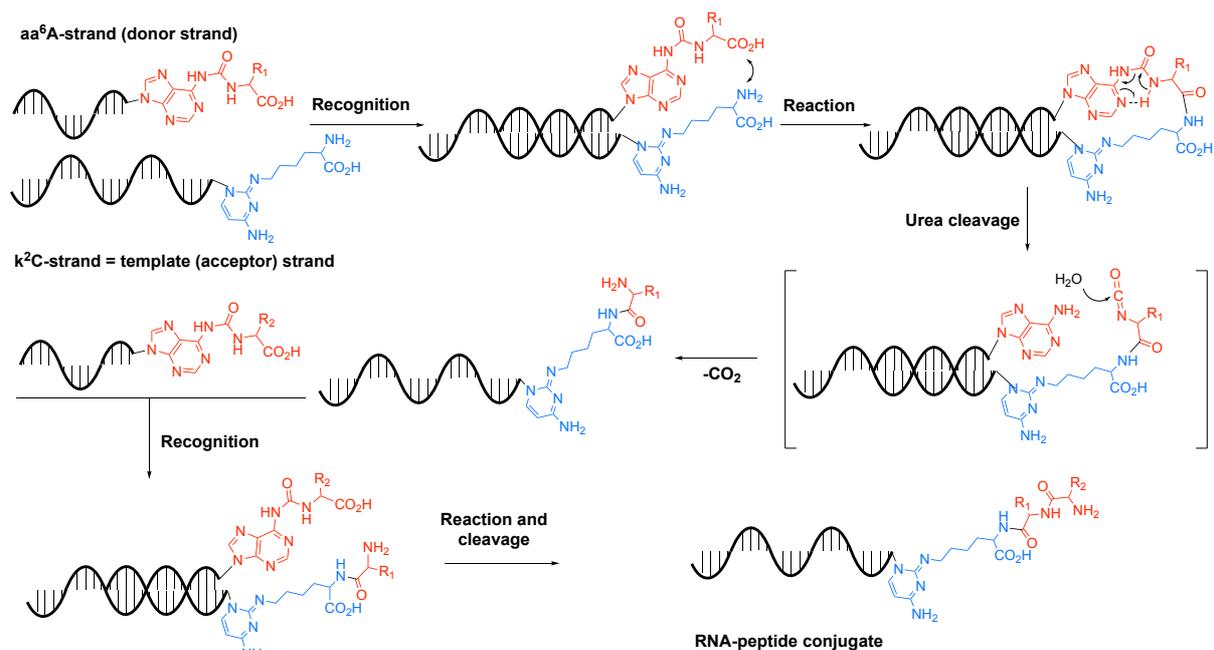
Figure 1. Schematic depiction of an RNA-templated peptide formation assisted by amino acid modified RNA bases.

The feasibility of 5 different RNA modifications (k^2C, k^2U, acp^3U, mnm^5U and nm^5U) as amino acid (aa) acceptors and (m^6)aa^6A bases as amino acid donors were investigated to re-establish a potentially ancient RNA-templated peptide formation system. Although the synthesis of a k^2C phosphoramidite was not achieved, the phosphoramidite building blocks of k^2U, acp^3U, mnm^5U and nm^5U were successfully prepared. Their incorporation into oligonucleotides was achieved. k^2U modified RNA oligonucleotides were indeed able to successfully establish peptide bond formation. We then tested other potential acceptor nucleotides for more reasonable prebiotic plausibility. Peptide bond formation under various conditions was investigated and finally successfully achieved between acp^3U, mnm^5U or nm^5U modified RNA oligonucleotides and several (m^6)aa^6A base containing RNA strands. Initially peptide bridged hairpins were formed, which were subjected to prebiotically plausible urea cleavage conditions, providing either RNA bound peptides or cyclized products, depending on the pH applied. aa^6A bases were in principle suitable for the discovered RNA-templated peptide formation but reactions with m^6aa^6A bases were cleaner. Following our discovered RNA-templated peptide formation, several RNA-peptide conjugates were synthesized as depicted in Figure 2. It was shown that the most promising RNA based amino acid acceptor modification for this approach is mnm^5U (Figure 2d). The results obtained in this study allowed us finally to construct a primitive translational machinery based on modified RNA bases, which

are "living molecular fossils" from the prebiotic RNA world. We could develop a theory about the origin of the ribosome.
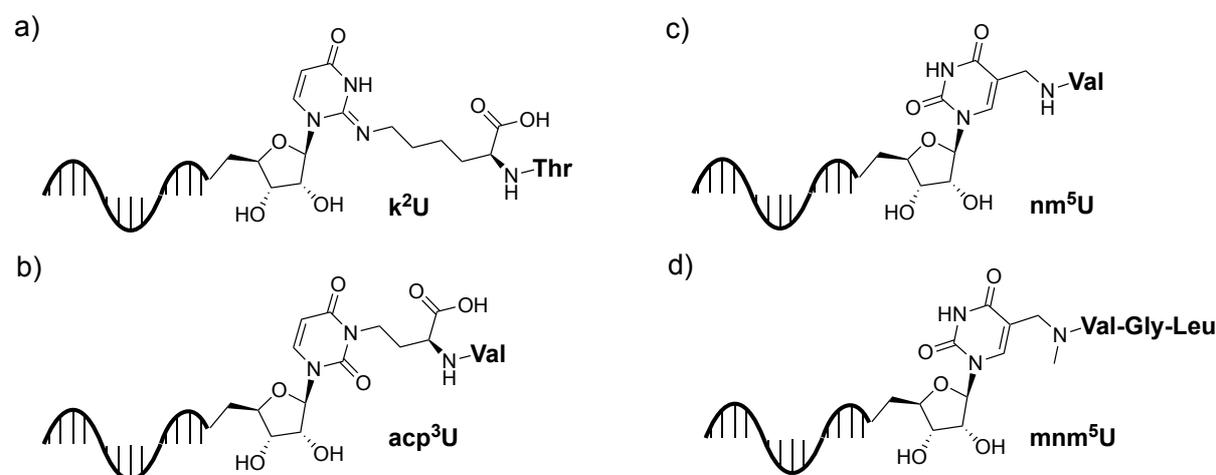
Figure 2. Schematic depiction of RNA-peptide conjugates synthesized via an RNA-templated peptide formation approach.

Because some of the naturally occurring modified RNA bases are today essential for the construction of RNA based medicines, in the second part of the thesis, I designed and synthesized RNA strands containing modified bases to fight against the severe acute respiratory syndrome betacoronavirus 2 (SARS-CoV-2) infection. The complete genome of coronavirus was already sequenced and analyzed. Based on these data, we designed several C_m- and U_m- containing siRNA sequences with additional 3' overhangs able to degrade the CoV-2 genome at the positions encoding the RNA-dependent RNA polymerase and the spike protein.

The biological activity of synthesized siRNAs was investigated by utilizing a dual-luciferase reporter assay. The biological experiments were performed by *Dr. Franziska Traube* and *Ammar Ahmedani*. Some of the prepared modified siRNAs indeed showed strong activity against the CoV-2 virus leading to virus-load efficiency of ~98%.

The here prepared siRNAs contain C_m and U_m to increase the stability of the therapeutic siRNA against nucleases. An alkynyl moiety was furthermore incorporated in order to attach glucose or lipids (cholesterol and arachidonic acid derivative) via "click" chemistry (Scheme 1). This improved the delivery of the siRNA bioconjugates into cells by receptor-mediated cellular uptake. Since the interaction between carbohydrates and carbohydrate-binding proteins is weak, multiple carbohydrate moieties were linked to the oligonucleotides to increase the

binding affinity. We also addressed the question if the positioning of sugar attached on the siRNA influences the knockdown efficiency. Hence, siRNAs containing two and three glucoses on the sense and antisense strands were synthesized (Scheme 1). All synthesized strands (Scheme 1) were subsequently used for biological testing in collaboration with *Dr. Franziska Traube* and *Ammar Ahmedani*.

With all transfected siRNAs we observed very strong silencing effects. Glucose-siRNA showed activity of 91%, cholesterol-siRNA conjugate of 87%, and arachidonoyl-siRNA of 93%. To highlight the advantages associated with bioconjugation we also tested glucose-siRNA conjugates without any transfection reagent. Although the silencing efficiency was lower (roughly 60-70%), in principle glucose conjugation could replace the transfection reagent.
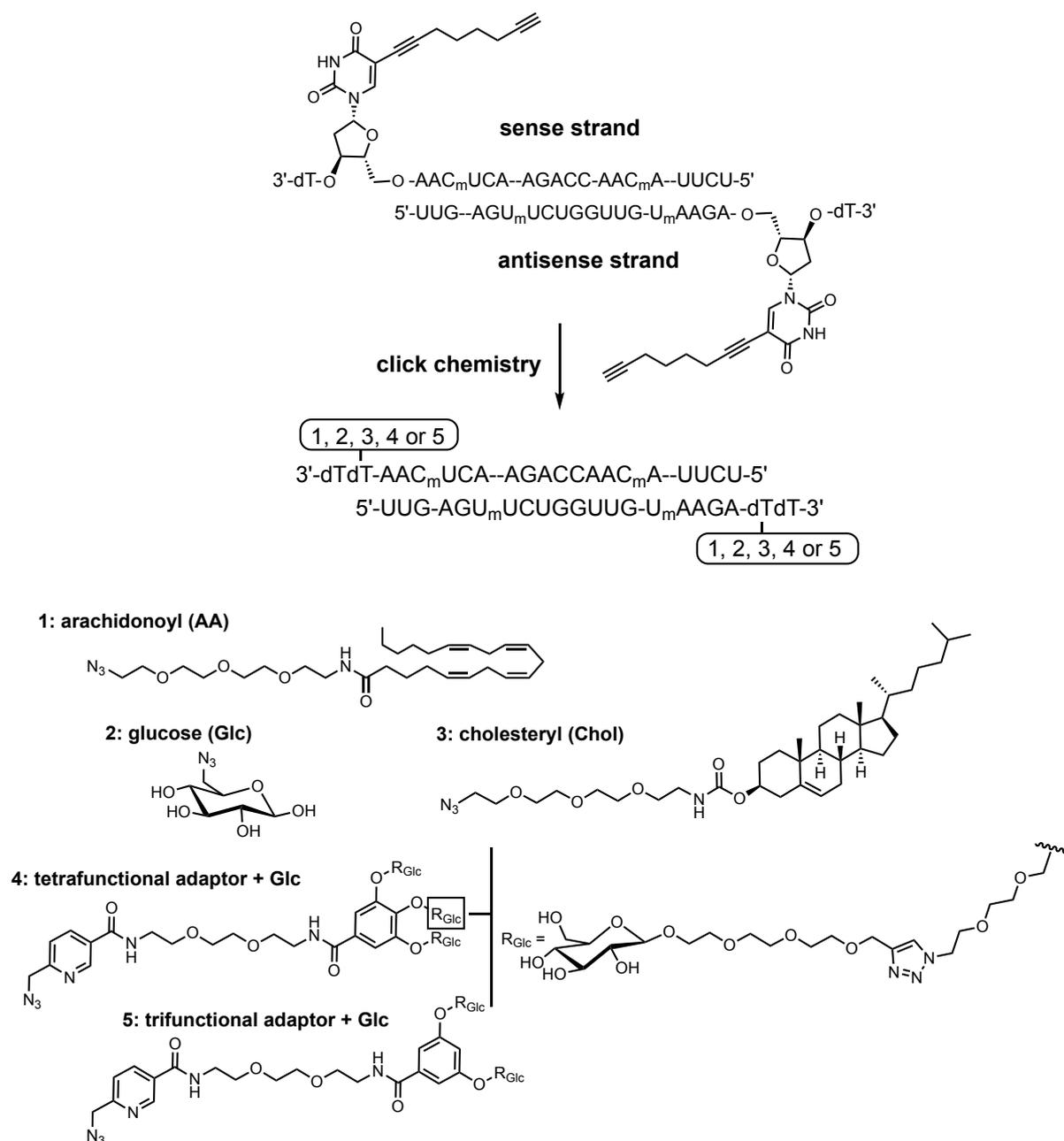
Scheme 1. Synthesis of glucose- and lipid-modified siRNA conjugates.

We also observed that the positioning of the modification on the strand is important. The silencing effect of siRNA bearing one glucose on the sense strand was higher in comparison to the glucose modified antisense strand. This is in contrast to siRNAs bearing multiple glucoses attached. In this case three glucoses on the antisense strand gave better results and showed the activity of ~60%.

2 Introduction

2.1 RNA modifications

There are numerous studies about the central role of RNA in cellular functions. RNA translates the genetic code and decodes it into protein. Moreover, it has various catalytic and regulatory functions.[1-4] microRNAs and transfer RNAs are composed of less than one hundred nucleosides in length, while long non-coding RNAs (lncRNAs), messenger RNAs (mRNAs) and ribosomal RNAs (rRNAs) can contain thousands of nucleosides in length. Such variation in length is a big contributor to RNA versatility as well as to the complex roles that it plays in the cellular environment. RNA is not limited to canonical nucleosides but naturally contains over 100 different post-transcriptional modifications with a significant chemical diversity.[5,6] Some of the modified bases exist in specific locations. For example, the post-transcriptional methylation of uridine occurs entirely at position 54 in tRNA.[7] Some of the modifications are chemically simple, such as 2'-*O*-methylation and pseudouridylation,[8] while the others are more complex like the tricyclic modification of G, wyosine,[9] found at position 37 in tRNA^{Phe} of eucaryotes. Modified bases are found in all types of RNAs, in every structural motif and also in unstructured regions. Non-canonical nucleosides can be recognition determinants for proteins as well as influencers of RNA structure and function. Although the existence of modified nucleosides has been known for over six decades, the knowledge of how modified nucleosides alter the structure, function, and properties of RNAs is still in its infancy due to missing technologies sensitive enough to investigate these modifications. The naturally occurring RNA nucleotides offer, however, a number of application possibilities that are now thoroughly explored in order to improve the properties (e.g. resistance to nucleases,[10,11] physicochemical properties[12,13]) of oligonucleotides particularly for the introduction of them as new pharmaceuticals. Finally, if we consider that the found modified bases are relics of a prebiotic RNA world, studying them may allow us to learn more about the chemistry that led to the evolution of the today existing RNA-protein world.

Part I – Modified RNA bases in prebiotic chemistry

2.2 Towards RNA-templated peptide formation utilizing non-canonical nucleosides

2.2.1 Origin of translation and the emergence of the ribosome

The origin and development of the translation mechanism is one of the major questions in the study of the origin of life, because it reflects a key event during the transition from chemical to biological evolution. Although the contemporary ribosome and the translational process are thoroughly investigated, the origin of translation, as a template-directed process, is still an unsolved problem. The genetic code is universal and the translation process at the ribosome follows principles that are common to all organisms on Earth. This suggests that both, the code and its translation, originated early during evolution.[14]

Simulation experiments of chemical evolution support the idea that catalytic properties of polypeptides[15] and replication activity of polyribonucleotides[16,17] existed before the more or less complete protein synthesis mechanism appeared. One suggestion is that the translation machinery was mostly or even entirely built out of nucleic acid, which was able to express the earliest version of the genetic code. Little or no proteins were needed. This is in agreement with the RNA world hypothesis, which proposes that RNA stored both genetic information and catalyzed chemical reactions potentially in a primitive cell. Later, DNA must have taken over as the genetic material and proteins became the major catalysts.[18-22] Early suggestions were based on ideas that initially no catalyzing enzymes were involved, because the initial tRNAs might have had a special cavity to hold its own amino acid.[23] It was speculated that it would have been possible if tRNA existed in two conformations (*FH* and *hf*) and bound to the messenger RNA with five-base pairs (Figure 3) and not with three as it is in contemporary translation. It was assumed that a tRNA molecule takes up the *hf* conformation once an amino acid is attached. tRNA changes the conformation to *FH* when a peptide gets attached to the tRNA. However, since a sequence of five bases has to be recognized this idea would impose base sequence restriction to encode early messages.[14,24]
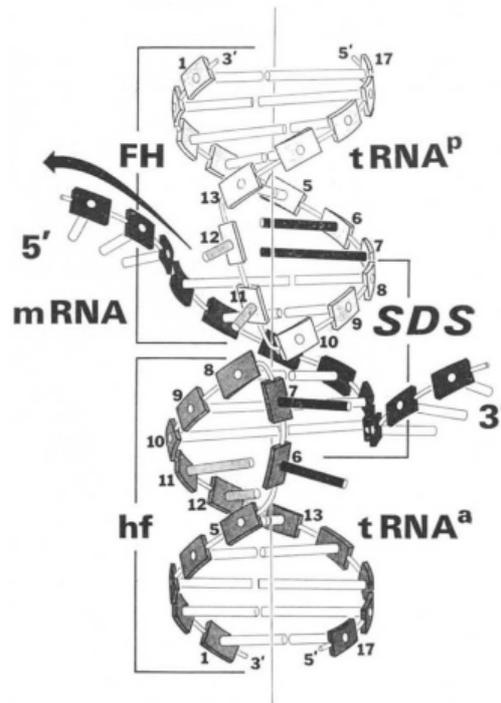
Figure 3. Depiction of the idea of a basic translation complex. The complex includes anticodon arms of both a peptidyl tRNA (tRNA[p]) in the *FH* conformation and an aminoacyl tRNA (tRNA[a]) in the *hf* conformation, together with their respective codons (combined with the anticodons in a central "sextuplet duplex structure", SDS). Adapted from [24].

It was suggested that the early genetic code encoded only a few amino acids. The first amino acids were proposed to be glycine (Gly), serine (Ser), aspartic acid (Asp), asparagine (Asn), although others argue that it was valine (Val), alanine (Ala), aspartic acid and glycine[25]. As said, it seems unlikely that all todays encoded amino acids were available at the time when the genetic code started. For example, tryptophan (Trp) and methionine (Met) are considered later additions as suggested by Crick.[14] At the later stages of the evolution of the genetic code, the primitive ribosome became likely more advanced and might have made it unnecessary for a tRNA to interact with more than three bases. At the same time modifications of the anticodon loop were established, particularly the incorporation of amino acid modified bases, which can be considered to be molecular fossils.[26-29] Today these bases suppress unwanted pairing outside the anticodon triplet.

Ishigami *et al*. proposed that the early translation process evolved in four stages.[30] Initially the amino acids were activated, and the peptide bond formation was catalyzed by ribonucleotides. Then less common amino acids were incorporated by the help of primitive tRNAs and insoluble ribonucleotides. The shorter RNAs were soluble allowing them to form

complexes with amino acids. At this stage it is suggested that an amino acid was attached on the 5'-terminal phosphate of tRNA, because the reaction can occur non-enzymatically. [31,32] In contrast, longer RNAs could have been insoluble due to their higher molecular weight enabling them to serve as adsorbing materials for the soluble aminoacyl polynucleotides. This would bring them into a more concentrated state. Thus, the soluble polynucleotides could have functioned as primitive tRNAs, while the insoluble ones would have been the precursors for mRNA and rRNA. The initiation of templated peptide synthesis is considered as a third step by Ishigami *et al*.[30] Some RNA strands were distributed on the insoluble RNAs, and aminoacylated tRNAs were sometimes base pairing with them, but these interactions were rather random before the specific binding sites for primitive tRNA appeared on the insoluble RNA. Later on, the information transfer from RNA to peptides improved, and mRNA and rRNA differentiated. mRNA had to become independent from the insoluble RNA (considered as a primitive ribosome). In the last stage of evolution more complex proteins, like ribosomal proteins or aminoacyl tRNA synthetase (aaRS) began to participate in translation.

Ribozymes[33,34] as single stranded RNAs that possess catalytic properties after forming secondary structures, similar to the action of protein enzymes, play an important role in the origin of translation.
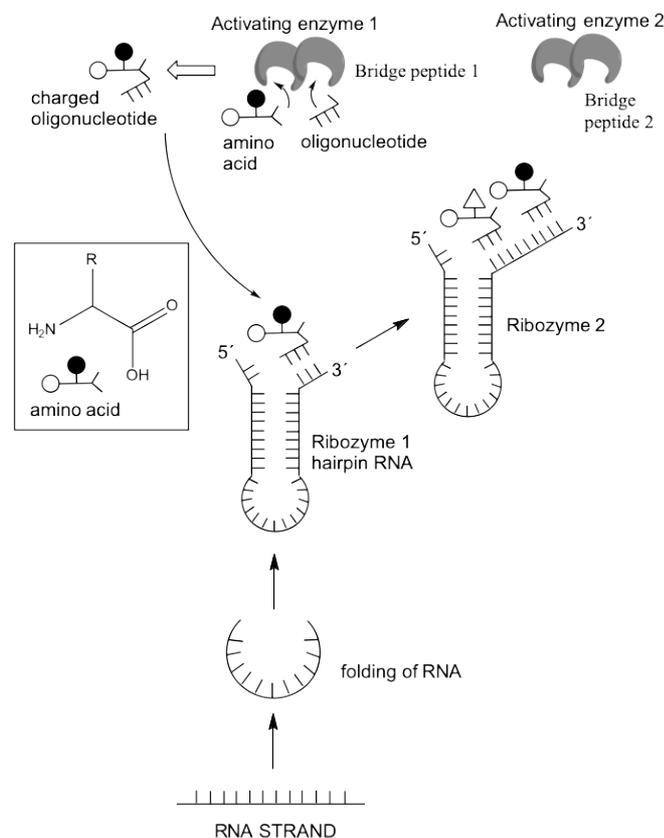
Figure 4. The origin of the ribozyme and its binding to the amino acid. Adapted from [36].

Ribozymes might have even employed an amino acid as a "cofactor" in order to serve as better catalysts (Figure 4).[35] ssRNA can fold into a hairpin loop, with the 3' and 5' ends remaining free. The 3' end could function as an acceptor stem to covalently attach to a specific amino acid (Figure 4). Such a hairpin RNA with a specific terminal sequence may have bound the corresponding amino acid as a "cofactor" to enhance the catalytic efficiency, becoming a ribozyme.[35] Initially, conjugation of one kind of an amino acid and one kind of hairpin RNA would be catalyzed by an activating enzyme, such as a bridge peptide.[37] This is a very short peptide that enables the emergence of an RNA-peptide complex promoting a primitive translation. Such bridge peptide could have been a precursor of the aminoacyl transfer tRNA synthetase. The ribozyme would be involved in a complex formation, bringing two amino acids in close proximity helping the peptide bond to form (Figure 4). The precedence of ribozymes catalyzing peptide bond formation indeed exists. *In vitro* selected ribozymes were shown by Cech to mediate peptide formation.[38] Additionally, an aminoacylated tRNA minihelix might have catalyzed peptide bond formation due to the sequence complementarity between the 3'-CCA sequence of a minihelix and a puromycin-bearing oligonucleotide. However, the proximity of the reactive centers was not sufficient by itself, so that imidazole as a catalyst may have been additionally required.[38] A very interesting approach was suggested by Yarus showing that a tiny ribozyme, GUGGC/GCCU, aminoacylates terminal uridine using phenylalanyl-adenylate (Phe-AMP).[39] Although even the formation of peptidyl-RNA was demonstrated, the approach requires unstable amino acid adenylates.

The enzymatic activity of ribozymes could have been the predecessor of the peptidyl-transferase center (PTC) that is today responsible for peptide bond formation and peptide release. Given the catalytic and replication properties of a ribozyme, the adaptor ribozymes are the precursors of tRNA, which could be the ancestor of all RNAs and the ancient component of the ribosome.[40]

It was proposed that mRNA was made by tRNA. When tRNAs or pre-tRNAs learned how to recognize and react with amino acids, they needed to safely store the information about the assignment of an amino acid. Thus, tRNA began to create mRNA or some kind of its precursor (pre-mRNA).[36] The anticodon of pre-tRNAs acted not randomly but as a template instead to recognize a codon of pre-tRNA that matched using base pairing mechanism. After hybridization, these triplets began to link to form longer strands of pre-mRNA that evolved to mRNA, where the information about the amino acids could have been stored.[36]

One more piece in the primitive translational machinery is missing – the ribosome. It provides the environment for mRNA and tRNA to form a peptide. The ribosome is known to be a ribozyme and its origin is an important question in prebiotic chemistry. Already in 1968[15] Crick was questioning why the contemporary ribosome is mostly made from RNA and why also the adaptor molecules (tRNA) are exclusively RNA based. Both, however, are heavily modified with non-canonical bases. As a potential answer he proposed that "RNA is cheaper to make than a protein". If a ribosome was mostly composed of proteins more ribosomes would be needed in the cell to make additional proteins.

Harish and Caetano-Anolles proposed the accretion model in order to explain the origin and evolution of the ribosome.[40] They believe that it is likely that rRNAs and ribosomal proteins coevolved to construct the ribosome. The authors suggested that ribosomes accreted to grow larger in time, adding the expanding segments, but some parts are more ancient than the others. Phylogenetic work shows that RNAs and proteins formed the ribosome together[40,41] and that ribosomes are molecular fossils from the early RNA-peptide world.[42] However, according to the RNA world hypothesis, peptidyl-transferase center of the large ribosomal subunit gave rise to the ribosome. Ribosomal proteins do not participate directly in the peptide formation but rather act indirectly helping to fold rRNAs in order to increase the accuracy of the ribosome.[43-45]

Although there are various proposals and speculations about the emergence of the primitive translational machinery, there is no consensus. However, *in vitro* simulations of synthetic ribosomes are being investigated and will hopefully shed some new lights on understanding the evolution of protein synthesis.[46-48]

2.3 Contemporary translation mediated by non-canonical nucleobases

2.3.1 The mechanism of contemporary translation

The translation process in all species consists of three main steps: initiation, elongation and termination (Figure 5).[49]
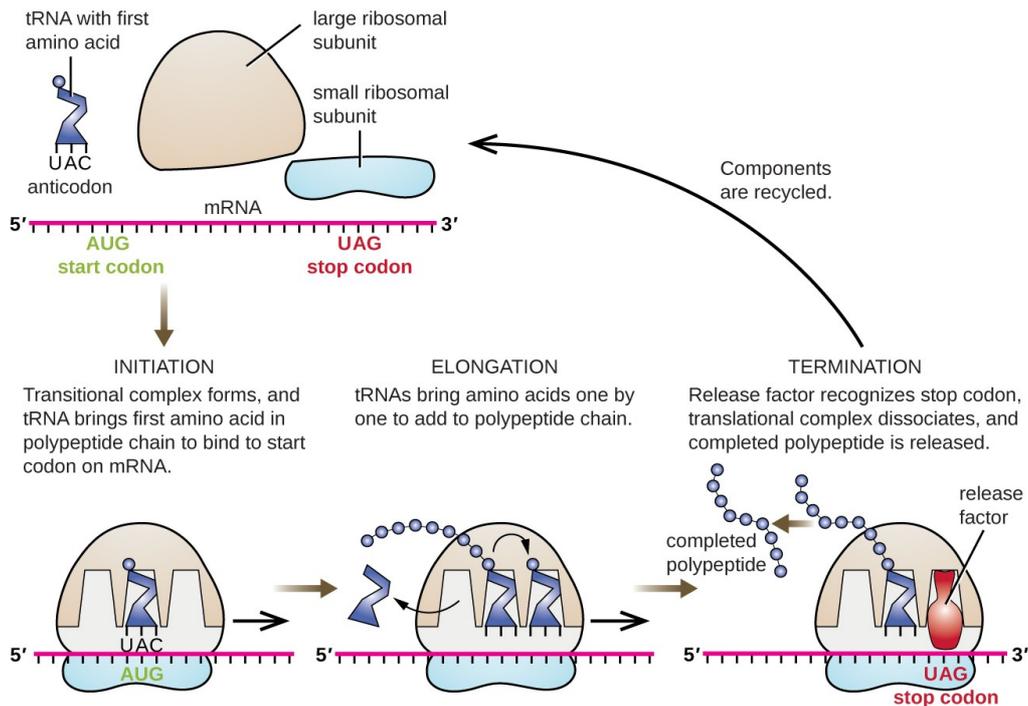
Figure 5. The mechanism of contemporary translation. Adapted from *Lumen Learning*.

During the initiation of protein synthesis, the initiation complex, consisting of the small ribosome subunit, mRNA template, initiation factors, guanosine triphosphate (GTP) and special initiator tRNA (*N*-formyl-methionine, fMet-tRNA^{fMet}, in *E. coli*), has to be formed. The initiator fMet-tRNA^{fMet} binds to the start codon AUG of the mRNA, thus fMet is inserted at the *N* terminus (the beginning of peptide synthesis) of every polypeptide chain in prokaryotes.

The elongation proceeds with single-codon movements (translocation) of the ribosome. The ribosome has three important ribosomal sites: the aminoacyl (A) site that binds incoming charged aminoacyl tRNAs; the peptidyl (P) site that binds charged tRNAs carrying growing peptide chain; the exit (E) site that releases unloaded tRNAs in order to reload a new amino acid carrying tRNA. During each movement, the charged tRNAs enter at the A site, then they move to the P site and subsequently to the E site for removal. Conformation changes cause the ribosome to move by three bases in the 3' direction. An amide bond is formed between the amino group of the amino acid attached to the A site and the ester of the amino acid attached to the P site tRNA. The reaction is catalyzed by peptidyl transferase. One interesting role for the ribosome is that it exposes the ester bond on the P site tRNA for the nucleophilic attack triggered by an induced conformational change in the peptidyl transfer center (PTC) on the A site tRNA.[50] This is a very elegant way of protecting the growing chain from water hydrolysis. The amino acid bound to the P site tRNA is also linked to the growing polypeptide chain.

As the ribosome moves along the mRNA, the former P site tRNA enters the E site, detaches from the amino acid, and is released. Translation is terminated when tRNA is confronted with a nonsense codon (UAA, UAG, UGA).

2.4 Amino acid modified bases as influencers for translational fidelity

2.4.1 tRNA modifications

A mature tRNA molecule is richly decorated with numerous evolutionary conserved nucleoside modifications (Figure 6) that occur in all domains of life and even beyond, as some viruses carry tRNA-like molecules that feature nucleoside modifying enzymes.[51,52]
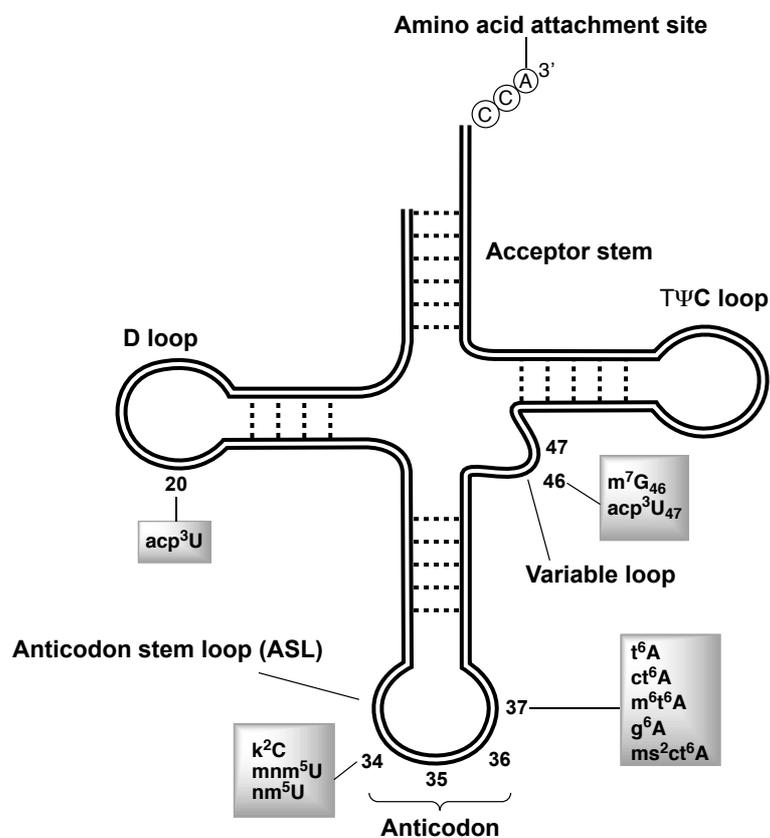
Figure 6. The structure of tRNA.

Nucleoside modifications in tRNA are particularly abundant in the anticodon stem loop (ASL, Figure 6). Most of prokaryotic and eukaryotic tRNAs contain a modified base at position 34, the first position of the anticodon (also known as a wobble position), e.g., lysidine (2-lysyl-cytidine, L or k^2C), 5'-methylaminomethyluridine (mnm^5U) or 5'-aminomethyluridine (nm^5U). These modifications are known to be directly involved in the decoding specificity and the stability of the codon-anticodon base pairing between the mRNA and tRNA at the ribosomal

A site. Here they contribute to the fidelity of translation.[53] Modifications in ASL position 37, 3'-adjacent to the anticodon are also known to play an important role in codon-anticodon interactions.[5,54] A loss of the modified base in this position increases the frequency of frame shifting.[55] Modified nucleobases in other parts of tRNA, e.g. 3-(3-amino-3-carboxy-*n*-propyl)uridine (acp^{3}U) are responsible for folding or tertiary structure stabilization.

2.4.2 Amino acid modified bases

Very often modified bases present in 34th and 37th positions are amino acid modified as shown in Figure 7, although acp^{3}U is found in other positions as well, as already discussed.
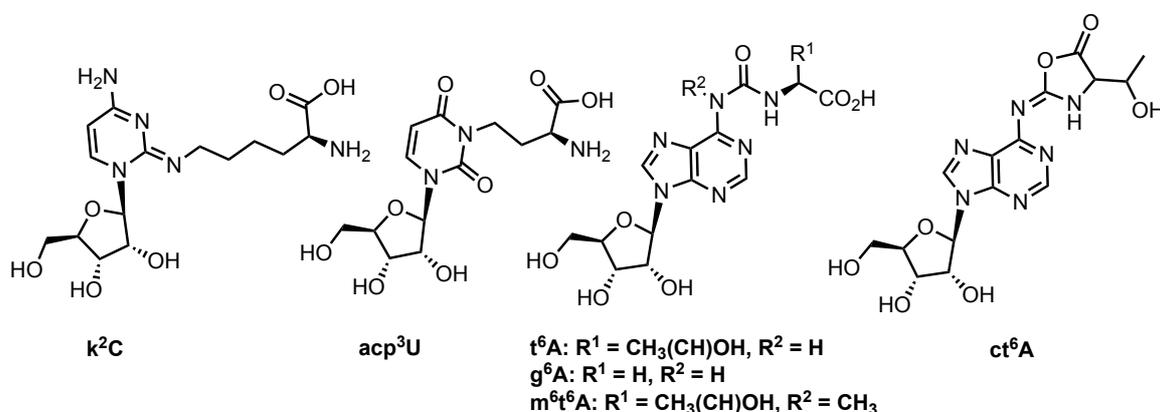
Figure 7. Amino acid modified bases.

For example, the hypermodified nucleoside k^{2}C (Figure 7) is a lysine-modified cytidine derivative that is present in the wobble position of bacterial AUA-specific tRNA^{Ile}. The tRNA containing a CAU anticodon encodes the incorporation of Met when unmodified (Figure 8).[56,57] When k^{2}C is replacing the C, the identity of tRNA^{Ile} bearing the CAU anticodon is switched to Met, and this tRNA^{Met} translates then the AUG codon as Met.[56] Thus, k^{2}C prevents the misrecognition of the AUG codon as isoleucine and that of AUA as methionine.
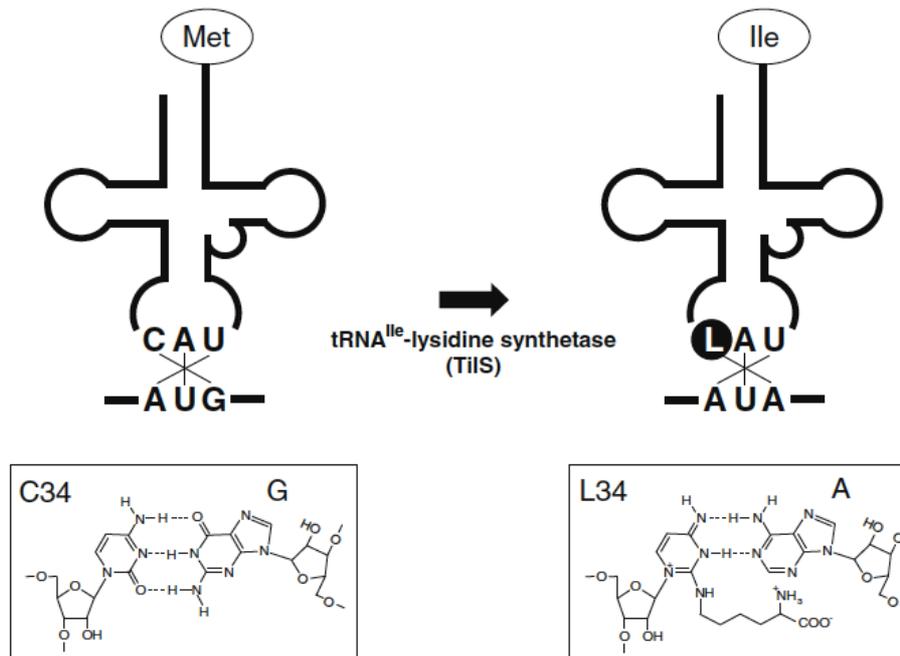
Figure 8. Lysidine changes the codon recognition and consequently the amino acid specificities of tRNA$^{\text{Ile}}$. Adapted from Muramatsu *et al*.[56]

Threonylcarbamoyladenosine (t^{6}A, Figure 7) is a modified nucleoside in the anticodon loop at position 37, which is found in tRNAs in all three kingdoms of life.[54] This hypermodified modification is found in virtually all tRNAs responsible for codons starting with A (ANN codons).[58,59] The bulky structure of t^{6}A supports the formation of the canonical U-turn structure of the anticodon loop[60] by preventing U33-A37 base pairing.[61] t^{6}A plays an important role in maintaining decoding accuracy during protein synthesis, and it is also required for aminoacylation of tRNAs[62] as well as for maintaining the reading-frame.[63] Although the existence of t^{6}A in tRNAs from *Escherichia coli* (*E. coli*) and yeast has been well known for several decades, t^{6}A was recently suggested to be a hydrolyzed artefact of ct^{6}A – a cyclic form of t^{6}A (Figure 7).[64] ct^{6}A is widely distributed in tRNAs from a certain group of bacteria, fungi, plants and some protists,[64] while t^{6}A exists in tRNAs of mammals, archaea and bacteria. In *E. coli* cells, basically all t^{6}A is dehydrated to form ct^{6}A. Thus, ct^{6}A is another modification of t^{6}A that enhances tRNA-decoding activity. In addition to ct^{6}A, the N^6-methylated version of t^{6}A, abbreviated as m^{6}t^{6}A (Figure 7) exists and is found in tRNAs from bacteria, fly, plants and rat.[54,65] In *E. coli*, m^{6}t^{6}A is present at the position 37 of tRNA$^{\text{Thr1}}$(GGU) and tRNA$^{\text{Thr3}}$(GGU), both of which decode ACY codons (Y is a pyrimidine base),[66,67] whereas the isoacceptors tRNA$^{\text{Thr2}}$(CGU) and tRNA$^{\text{Thr4}}$(UGU) contain ct^{6}A[63]. It was shown that the methyl group of m^{6}t^{6}A improves the efficiency of reading the cognate

codon ACC.[68] Glycine can also be occasionally present instead of t[6]A, generating N^6-glycinylcarbamoyladenosine (g[6]A, Figure 7). However, the role of this modification has yet to be elucidated.[69]

Amino acid modified bases can exist not just in the anticodon position. For example, the nucleoside acp[3]U (Figure 7) is present in several *E. coli* and mammalian tRNAs. In *E. coli* it is positioned in the extra loop of tRNAs,[70] while in eukaryotic tRNAs it appears at position 20 in the dihydrouridine loop (D loop).[71] Although the existence of acp[3]U in tRNA and rRNA is conserved, its role is presently unknown. NMR data show that acp[3]U binds Mg^{2+} in small oligonucleotides, which suggests a possible function of acp[3]U in the metal cation-based stabilization of local RNA structures.[72] Interestingly, increased levels of acp[3]U were observed in supernatants from breast carcinoma cells, thus it might serve as a tumor marker.[73]

Modifications mentioned above are composed of an amino acid and a nucleobase, which represent phenotype and genotype, respectively. They are known to be crucial for translational fidelity as mentioned before and hence we suggest that they are relics of an ancient code, when RNA hairpin structures that harbored primordial anticodons (the ancestor of tRNA) were strongly associated with or possibly even "charged" with amino acids.[26-29] Thus, these bases could reflect historical aspects of the evolution of the genetic code.

Part II – Therapeutic applications of modified RNA bases

2.5 RNA therapeutics

Some of the modified nucleosides today enable to utilize RNA as a new therapeutic agent that operates on the genomic level. In this context, advances in next-generation sequencing technology help to identify the genetic roots of many common diseases.[74] RNA-targeting therapeutics[75-77] such as antisense oligonucleotides (ASOs), aptamers, small interfering RNAs (siRNAs), microRNAs (miRNAs) and synthetic mRNAs can be used to control expression of disease-relevant genes and pave the way to treat various illnesses. However, in order to turn this concept into a clinical reality means that obstacles – to make the molecules more potent and less immunogenic, and to deliver drugs specifically to cells and tissues – have to be overcome.

Now, after years of careful investigation, the field is making progress. In 2016, *Nusinersen*, an ASO that can influence the splicing of mRNA, was approved by the U. S. Food and Drug administration (FDA or USFDA). It became the first pharmaceutical to treat spinal muscular atrophy. In 2018, *Patisiran* was marketed as a first small interfering RNA-based drug used to treat hereditary transthyretin amyloidosis (hATTR), a rare disease of the liver. Then in 2019, *Givosiran* witnessed its approval as another siRNA drug for the treatment of adults with acute hepatic porphyria. These recent successes proved the clinical utility of RNA-targeting therapeutics. Currently, dozens of new oligonucleotide-based potential drug candidates are under clinical investigation for disease indications including neurodegeneration, metabolic and cardiovascular disorders as well as various types of cancers.

2.6 RNA interference

RNA interference (RNAi) is a gene silencing process in which RNA molecules suppress gene translation because of the messenger RNA (mRNA) molecules being degraded. In 1998, Andrew Fire and Craig Mello published their discovery of mRNA degradation, induced after double-stranded RNA (dsRNA) is entering the cell.[78] In 2006 they shared a Nobel prize. Double-stranded RNA activates a biochemical mechanism, which induces degradation of single stranded mRNA molecules bearing a genetic code identical to that of the double-stranded RNA. When mRNA molecules are degraded, the translation of the corresponding gene is silenced, blocking the production of the corresponding protein. RNA interference occurs in plants, animals and humans. The process is used for the regulation of gene expression,

participates in defense against parasitic nucleotide sequences, viruses and class I transposons (genetic elements that can shift to different locations within a genome), as well as finds utility in the RNA-targeting therapeutics field.

2.7 microRNA and siRNA

microRNA (miRNA) and small interfering RNA (siRNA) are two types of RNA molecules that are pivotal to RNA interference. The first small RNA, known as microRNA, and its regulatory function was discovered in 1993 while screening nematodes.[79,80] miRNAs are defined as a class of small, endogenous RNAs, which are of 21-25 nucleotides (nts) in length. miRNAs play an important regulatory role in animals and plants by targeting specific mRNAs for degradation or inhibition of translation and thus provide a potential novel class of therapeutics. miRNAs are produced from hairpin shaped precursors.[81] In animals, miRNAs are synthesized from primary miRNAs (pri-miRNAs) in two steps. The synthesis is performed by two different RNase III-type proteins: Drosha in the nucleus and Dicer in the cytoplasm.[82] In plants the maturation of miRNA occurs entirely in the nucleus and is performed by a single RNase III enzyme, DCL1 (Dicer-like 1).[83] After maturation, miRNAs are bound by multiprotein component complex referred to as RISC (RNA-induced silencing complex). The catalytic part of RISC, the Argonaute (Ago) protein, associated with miRNA, targets mRNAs and acts as a posttranscriptional regulator.[82]

In 1999, siRNA, also known as short interfering RNA or silencing RNA, was discovered in plants.[84] Soon afterwards, Tuschl and colleagues noticed that synthetic siRNAs could induce RNA interference in mammalian cells.[85] Figure 9 represents the structure of siRNA. It is a double-stranded RNA of ~20-30 base pairs in length.
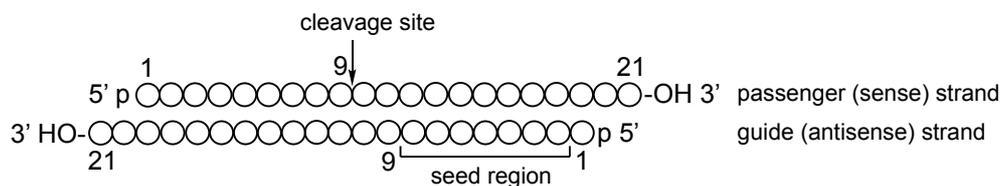
Figure 9. The schematic depiction of siRNA duplex: 21 base pairs of RNA core and 2 nucleotide 3' overhang on each strand.

It has two 2 nt overhang on the end of each strand at the 3' terminus. siRNA is similar to miRNA and operates within the RNAi pathway. The recognition of the target mRNA by siRNA is conferred by the "seed region" (positions 2-8 on the antisense strand). siRNAs also inhibit

the expression of specific genes with complementary nucleotide sequences by degrading the mRNA after transcription, thus preventing translation.

2.8 The mechanism of RNA interference

RNA interference is induced both by miRNAs, as regulators of endogenous genes, and by siRNAs, as defenders of genome integrity that respond to invasive nucleic acids such as RNA-containing viruses,[86] or retrotransposons[87]. siRNAs and miRNAs share a very similar mechanism of gene silencing as shown in Figure 10.
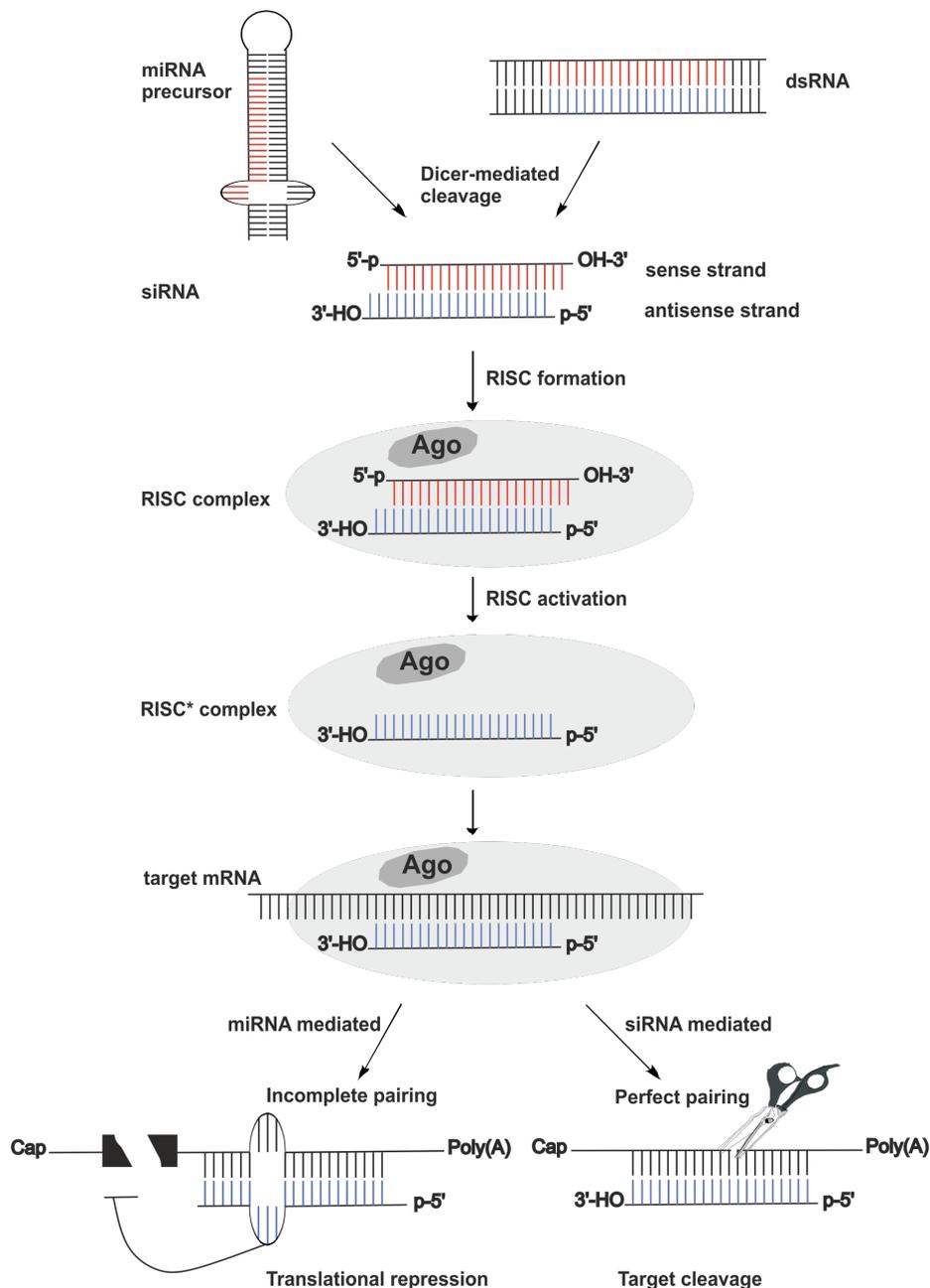
Figure 10. RNAi mechanism by siRNAs and miRNAs.

The RNAi mechanism can be divided into two stages: the initiation and the effector stages. During the initiation, RNA strands are processed by endoribonuclease Dicer into short (~20-30 nt) fragments with the two nucleotide overhangs on the 3' ends. The mechanism involves just the effector phase when chemically or enzymatically synthesized siRNAs are used.[88] Then during the effector stage, the RNA-induced silencing complex (RISC) is formed after which one of the siRNA strands (sense) is cleaved and dissociated from the complex, while the other (antisense) remains in the complex. Such activated complex RISC* binds to the complementary target mRNA and cleaves it (Figure 10).

In the first stage of RISC assembly, the R2D2 protein (in *Drosophila*) or its analog (e.g. TRBP in *Homo Sapiens*), which contains two dsRNA binding domains and a Dicer binding domain, binds to the thermodynamically more stable 5' end of the duplex, causing subsequent binding of Dicer,[89] whose domain has a specificity for 3' overhangs.[90] At the end of RISC assembly, the sense strand is cut by Argonaute-2 (Ago2, a catalytic part of RISC*), resulting in the formation of RISC*. The orientation of the Dicer-R2D2 heterodimer relative to the siRNA determines the selection of which strand is included in RISC*. R2D2 interacts with the thermodynamically more stable end of the duplex, thus the most active siRNAs are those with the 5' end of the sense strand having higher melting temperature than the 5' end of the antisense strand. Ago2 can cut both sense and mRNA strands,[91] however it is also known that the dissociation of the siRNA by human Ago2 can occur without its cleavage.[92,93] This complex might direct either mRNA translational repression (miRNA-mediated) or mRNA target cleavage (siRNA-mediated), depending on the degree of complementarity between the ~21–23 nt RNA and the mRNA.

2.9 Hurdles for siRNA

Shortly after the discovery of siRNAs, their potential as therapeutics was deeply investigated in clinical trials. However, various biological obstacles had to be overcome in order to successfully use siRNAs for therapeutic purposes. siRNAs are polyanions, which means that penetration directly through the hydrophobic cell membrane is not possible. Thus, siRNAs have to be taken up by the cell through endocytosis. The next challenge is then to escape from the enclosing endosome. Finally, when siRNAs enter the cytoplasm they can be degraded by cytoplasmic ribonucleases as reported by Whitehead *et al*.[94] Also due to the division of cells, the concentration of siRNAs can get decreased through dilution.

Initially it was thought that siRNAs act specifically and lack immunogenic properties.[95-100] However, after more data were published several types of off-targeting effects, elicited by different mechanisms, were discovered as depicted in Figure 11.[101] Off-target activity can make the analysis of phenotypic effects in gene silencing experiments rather complicated. Also, it can induce unwanted toxicities. There are several categories of siRNA off-targeting effects. First, so called microRNA-like off-targeting effects are caused by imperfect pairing of the siRNA strands with untranslated regions (3' UTRs) of cellular mRNAs resulting in a multiple-site cleavage and/or translational block. The sequence of the 5' end of the guide (antisense) strand is very important for off-target transcript silencing. This part of the strand is very similar to the seed region of microRNAs, which is important for target recognition. Each microRNA regulates plenty of mRNA targets, each with 3' UTR sequence complementarity to the 5' end of the microRNA antisense strand.[102-105] The regulation of microRNAs is similar to the capacity of siRNAs to control numbers of transcripts enriched for seed region motifs. siRNAs and microRNAs share the same silencing mechanism, and the off-targeting effects of siRNAs are induced by siRNAs entering the natural microRNA pathway. It was shown that off-targeting effects are increased at higher concentrations of siRNAs. An optimal siRNA concentration that could maintain full on-target-silencing was not found.[106,107]
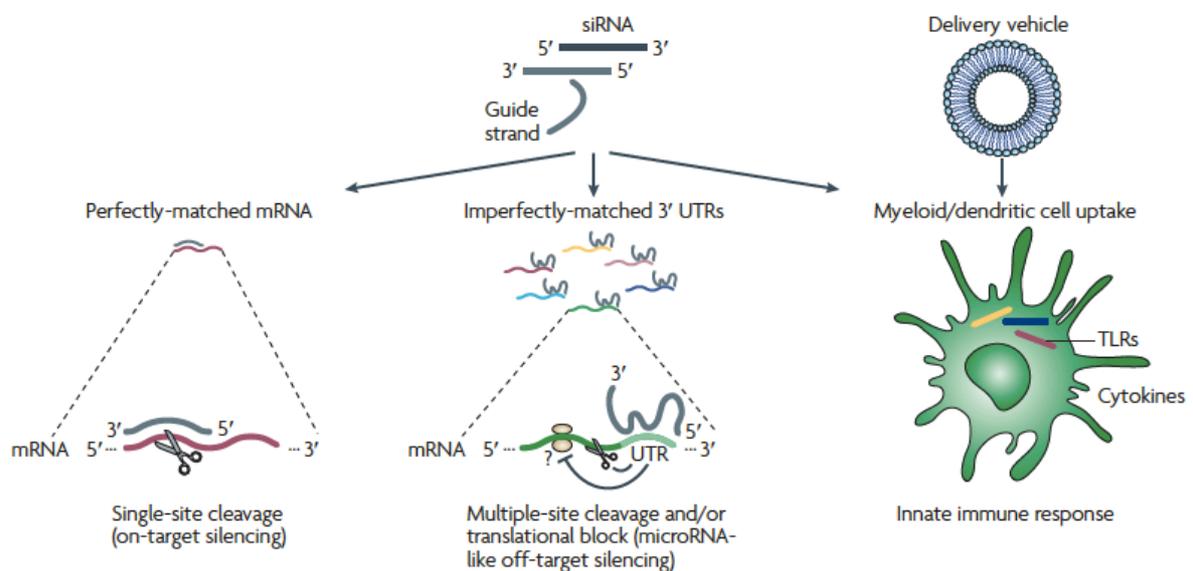
Figure 11. Off-targeting effects caused by siRNAs. Micro-RNA-like off-target silencing is sequence-specific and is caused by imperfect pairing of siRNA strands with sequence motifs that appear mainly in 3' UTR regions of cellular mRNAs. Other off-targeting effects of siRNAs result from innate immune responses to either oligonucleotides or the delivery vehicles. TLR: toll-like receptor. Adapted from Jackson *et al*.[108]

The other type of off-targeting effect can be triggered by siRNAs and/or their delivery system. Mammalian immune cells express toll-like receptors (TLRs) that recognize pathogen-associated molecules, including cell wall components, flagella and bacterial or viral nucleic acids.[109,110] Initially it was thought that siRNAs shorter than 30 nt are not big enough to induce the response of the immune system.[111] But later studies revealed that short synthetic siRNAs can induce unanticipated, non-specific effects.[112,113] siRNAs trigger the immune system response mainly by activating TLR7 and TLR8, which identify RNAs in the endosomal compartment.[114,115] The recognition of siRNAs by TLR7 and TLR8 is sequence dependent. Although it is not clear what sequences are recognized by these proteins, it is known that U and G rich sequences, together with 5'-UGU-3' motifs have high immunostimulatory activity.[116,117] Also AU-rich sequences can trigger an immune response by preferentially activating TLR8.[118] The vast majority of native siRNAs causes an immune response. It is challenging to eliminate immunostimulatory activity even after careful design of the sequence. Moreover, the structure of the siRNA is also very important. TLR3 recognizes the duplex form of siRNA, whereas TLR7 and TLR8 identify both single and double strands of siRNA. Thus, the immunostimulatory effect can be caused by any form of siRNA and in some cases single strands induce even stronger activity in comparison with double strands.[113,115,119,120] The length of the siRNAs, together with certain sequences also play a role in immune cell activation. Hornung *et al.* identified the 9mer motif of the sense strand (GUCCUUCAA) that induces the immune response through interferon-alpha (IFN-α). The minimal length of such RNA has to be around 19 bases, since 12mer and 16mer strands containing that 9mer sequence showed much lower immune activity.[113]

The delivery system can also cause an immune response. A study by Judge *et al*. showed that lipid delivery vehicle increased the production of interferons and cytokines. This was not observed with unformulated siRNA or lipids alone. [114] Such lipid delivery systems protect the siRNAs from nuclease degradation giving extended circulation times in blood and a facilitated uptake by endocytosis. As internalization and endosomal maturation are important for the activation of the innate immune system due to the presence of TLRs in the endosome,[119] delivery vehicles that require the endosomal pathway are likely to induce the immune response. Thus, delivery strategies that use different pathways have the potential to bypass the immunostimulatory effect. For example, siRNAs delivered directly to the cytoplasm by electroporation did not induce the immune response, while the same siRNAs packaged in cationic lipids caused the release of cytokines (small proteins important in cell signaling).[119]

2.10 The role of chemical modifications

Identification of potential immunostimulatory motifs could help for a rational design of synthetic siRNAs that could avoid activating the immune system. However, removing the putative immunostimulatory motifs can also cause the loss of activity. In addition, not all sequences that can induce the immune response are known. Chemical modifications are widely used to block the recognition of siRNAs by the immune system, to influence nuclease-resistance or conformation of the RNA helix.[121-123] Figure 12 represents the most commonly used modifications that improve the properties of siRNAs.
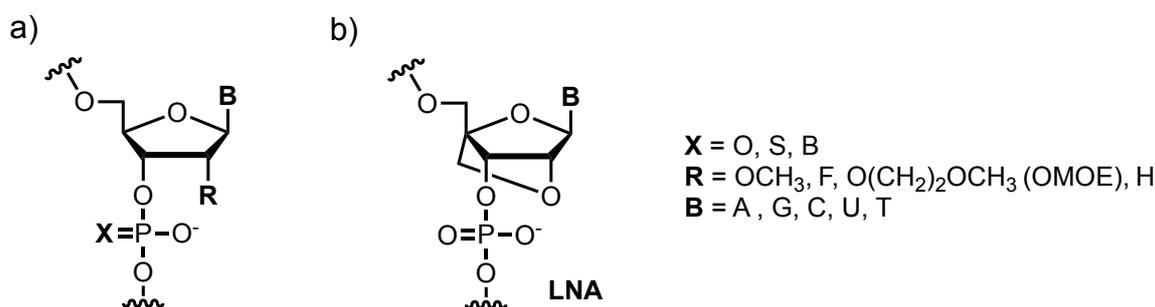
Figure 12. Chemical modifications of siRNAs. LNA – locked nucleic acid.

Studies show that even small chemical modifications can alter immunological properties without disrupting the potency of siRNAs.[122] Eberle *et al.* reported that thymidine as a modification can be introduced into siRNAs to inhibit interferon secretion without affecting gene-silencing activity.[123] The data are in accordance with the report published by Chiu stating that 2'-OH in the ribose backbone is not absolutely necessary for RNAi.[124] Substitution of the 2'-position of the ribose provides nuclease-resistant siRNA, since the 2'-OH group participates in the cleavage of RNA by endoribonucleases.[125] The size of the 2'-substituent plays an important role. A small methyl residue favors the 3'-*endo* ribose conformation, providing the A-form RNA, which is known to be important for RNAi. An investigation of the effect of 2'-sugar modifications (2'-F, 2'-OMe and 2'-OMOE ($O(CH_2)_2OCH_3$)) on both sense and antisense strands was performed by Prakash *et al.*[126] The study of the antisense strands revealed that the activity of the strand highly depends on where the modification is positioned. siRNAs with modified ribonucleotides at the 5'-end of the antisense strand were less active in comparison with 3'-modified ones. However, highly modified siRNA with 2'-F can lead to toxicity. This is why *Alnylam Pharmaceuticals* in 2016 discontinued the clinical trials of the conjugate of *N*-acetylgalactosamine (GalNAc) and siRNA with 50% of 2'-F modifications after severe cardiotoxicity was revealed.[127] In general, the 2'-F moiety was well tolerated in the antisense

strand, while the 2'-OMe showed significant change in the activity depending on the position of the modification. The 2'-OMOE modification (introduced to evaluate the contribution of modification size) in the antisense strand resulted in less active siRNA constructs, irrespective of the position of the modification in the strand. A positional preference of the modification, e.g., 2'-OMe and 2'-OMOE, in the sense strand was not observed. These properties support the strategies to make siRNAs druggable,[128] however full or even 50% replacement of 2'-OH to 2'-OMe leads to inhibition of gene silencing.[129]

Various nucleic acid analogues with structural changes in the furanose ring are also known to protect siRNAs from nucleases. Among them, locked nucleic acids (LNA, Figure 12b) were thoroughly investigated. LNAs significantly increase the melting temperature of siRNA, however the antisense strand is very sensitive to this modification which may cause total inhibition of RNAi.[130]

Another type of modification strategy is the derivatization of the phosphate backbone as was shown in Figure 12a. Oxygen can be replaced by sulfur[131,132] or boron[133] to form phosphothioates and boranophosphates, respectively. They were shown to protect siRNAs from nucleases.[134] On the other hand, these modifications inhibit the RNAi process to some extent.[135] Phosphothioates are known to protect the siRNA from exonucleases,[136] thus this modification is very valuable when introduced in the terminal positions of siRNAs.[134] Phosphothioates can enter the cells by clathrin-dependent endocytosis without the use of transfection agents.[136] Nevertheless non-specific interactions of siRNAs containing phosphothioates with serum proteins, cell receptors can occur,[137] which limit the clinical potential of these compounds. Although boron derivatives were shown to be very effective protecting units against nuclease mediated degradation,[133] the method to synthesize large quantities of boron-modified siRNAs is still lacking.

Natural siRNAs or miRNAs have 5'-phosphates, while synthetic duplexes have 5'-OH groups that get phosphorylated inside the cells by cellular kinases.[138] The presence of a 5'-phosphate on the antisense strand is essential for RNA interference.[139] Chemical modifications of the first nucleotide of the 5' end of the antisense strand can interrupt intracellular phosphorylation.[140-142] However, the introduction of metabolically stable phosphates (e.g. 5'-methylenephoshonate,[143] 5'-(*E*)-vinylphosphonate[144], Figure 13) can restore the activity.[145]

5'-methylenephosphonate   *E*-vinylphosphonate

R = OCH₃, H
B = A , G, C, U, T
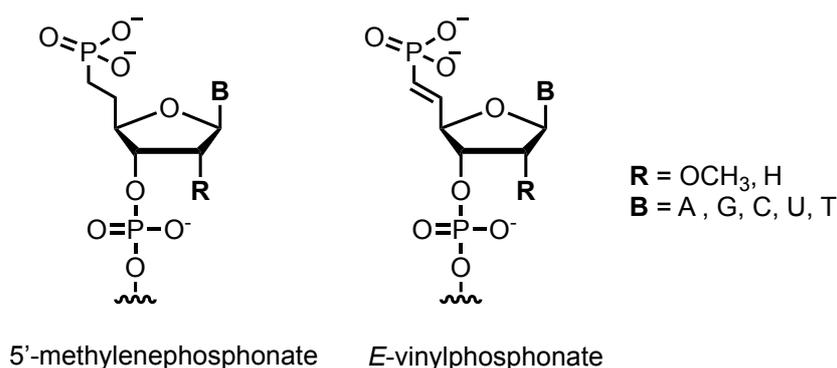
Figure 13. Schematic depiction of metabolically stable phosphates.

2.11 Delivery systems

Specific and efficient delivery of siRNAs is still an unsolved issue and is considered to be a major obstacle for siRNAs to reach clinical use. Naked siRNAs are unstable in the bloodstream and cannot efficiently cross cell membranes, meaning that they have short lifetimes *in vivo*. Thus, the delivery (transfection) system has to be carefully designed in order to enable siRNAs to function as therapeutics. Figure 14 depicts the extra- and intracellular challenges that siRNAs face in order to reach their targets. Naked siRNAs are rapidly degraded in blood by RNases, filtrated by kidneys, absorbed by liver scavenger receptors. They can activate the innate immune system.[146]
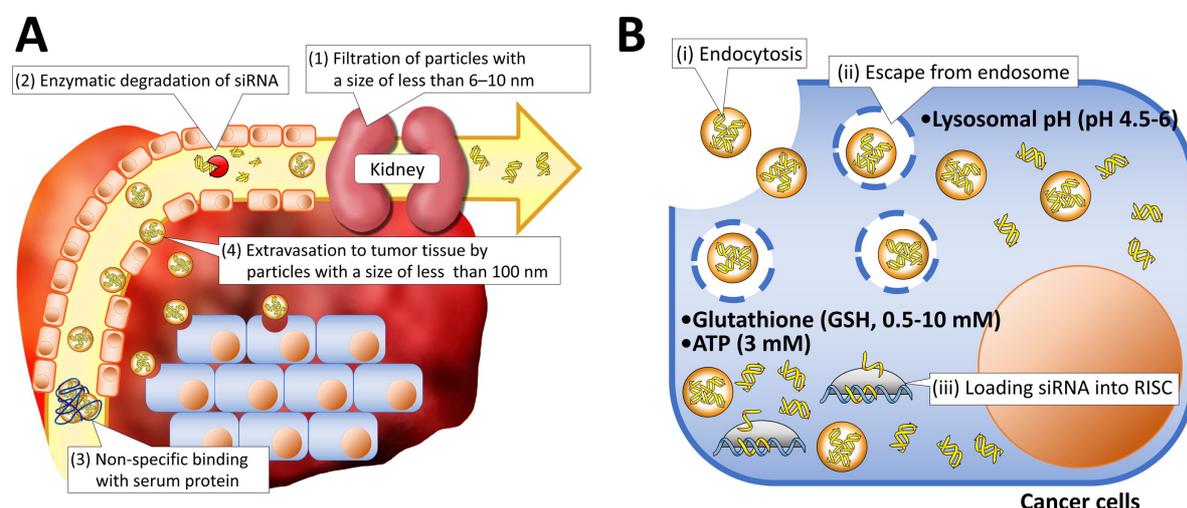
Figure 14. Schematic depiction of delivery challenges in the (a) extracellular and (b) intracellular regions. Figure adapted from Kim *et al*.[147]

The subcellular fate of nucleic acids depends on the mechanism of entry into cells. For example, in receptor-mediated endocytosis nucleic acids are encapsulated in endosomes after

their internalization. Most endosomes fuse with membrane-bound organelles, called lysosomes. In order to reach their targets, nucleic acids need to escape from the endosomes or lysosomes. A number of delivery vehicles have been investigated to improve the internalization including nanoparticles because they protect naked siRNAs from nucleases. They can be used for targeted delivery by attaching target-specific ligands to their surfaces. There are several advantages of nanoparticles. The size of the particles can be controlled, they are inert, they have enhanced circulation time allowing them to penetrate more efficiently. They can also be tracked and imaged. On the other hand, nanoparticles have poor solubility in water, poor hydrophobicity, as well as limited bioavailability.

Liposomes are nano-sized phospholipid bubbles that are commonly used for a therapeutic delivery.[146] In an aqueous solution, these materials can form liposomes, in which lipid bilayer forms a sphere with an aqueous core, that can host a nucleic acid.[147] Liposomes can also have an amorphous structure, where lipids and nucleic acids are interspersed. Felgner and colleagues showed already more than 30 years ago that lipofection (liposome transfection) can be successfully applied for transfection of nucleic acids using cationic lipid DOTMA (*N*-[1-(2,3-dioleyloxy)propyl]-*N,N,N*-trimethylammonium chloride, Figure 15a).[148] Various derivatives of DOTMA have been made[149] and commercialized (e.g. lipofectamine, oligofectamine). More recently, a stable nucleic acid-lipid particle (SNALP, Figure 15b) formulation was shown to be efficient *in vivo*. For example, Morrisey and co-workers showed the suppression of Hepatitis-B (HBV) replication by the delivery of an siRNA-SNALP system that targeted the RNA of HBV.[150]
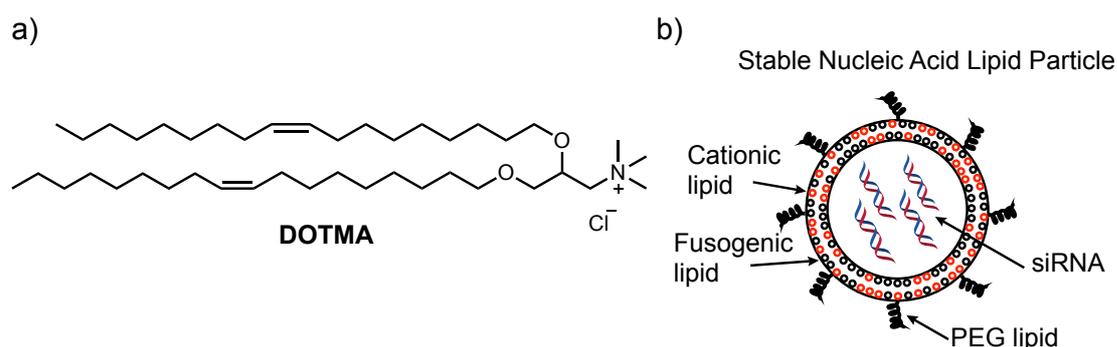
Figure 15. Schematic depiction of (a) DOTMA and (b) stable nucleic acid-lipid particle (SNALP).

Although liposomes are the most commonly used nucleic acid delivery system, there are concerns regarding their safety for a therapeutic use. The immunostimulatory effect of certain cationic lipids has been reported both *in vitro* and *in vivo*,[151-153] and some synthetic agents

have been found to induce gene expression changes that might increase the off-targeting effects of siRNAs.[154,155] Nevertheless, liposomes show promising results for future therapeutic use, as evidenced by the drugs that received regulatory approval from the FDA.[156] Moreover, *Alnylam Pharmaceuticals* developed *Patisiran*,[157] the first siRNA based drug against hereditary transthyretin-mediated (hATTR) amyloidosis. The drug is formulated in a lipid nanoparticle targeted to inhibit hepatic TTR synthesis.

2.11.1 Polymers

Linear or branched cationic polymers bind and condense nucleic acids into stabilized nanoparticles, thus they are used as transfection reagents.[158,159] It was also shown that such materials stimulate nonspecific endocytosis and endosomal escape.[154] One of the polymers used as a delivery agent for siRNAs is polyethylenimine (PEI).[160] It can be used for various local siRNA delivery applications. Pain receptors in rat faced a selective knockdown by intrathecal injection of PEI-siRNA complexes.[161] PEI was also effective in a subcutaneous mouse tumor model. The systemic administration of complexed, but not naked siRNAs, led to the delivery of the intact siRNAs into the tumors which reduced their growth.[162] However, high molecular masses and high doses of PEI are known to be toxic.[163,164]

Cyclodextrin-based polymeric nanocarriers can be also used for the delivery of siRNAs. The systemic delivery of sequence-specific siRNAs against the EWS-FLI1 gene product (Ewing's sarcoma fusion gene) was investigated by a nonviral delivery system that caused the inhibition of tumor growth in a murine model of metastatic Ewing's sarcoma.[165] The delivery system was composed of cyclodextrin-containing polycations to bind and coat siRNA, while transferrin was used as a targeting ligand for entering the transferrin receptor-expressing tumor cells. Removal of the ligand or the use of a control siRNA sequence did not show any antitumor effects.[165]

2.11.2 Bioconjugates

The delivery systems covered above were based on non-covalent interactions, but siRNAs can also be delivered as part of a bioconjugate. The possible conjugates can be biomolecules that can specifically bind to receptors on the cell membrane. Binding molecules are folate,[166] antibodies,[167-169] peptides,[170,171] aptamers[172,173] and carbohydrates[174]. Molecules that penetrate into the cell by natural transport mechanisms (e.g. cholesterol,[175,176] vitamins[177]) can also serve in such conjugates, as well as molecules that interact with the membrane non-

specifically (e.g. positive electrostatic charge and hydrophobicity).[178,179] The structure of the linker that connects the biomolecule with siRNA also plays an important role in the efficiency of RNAi. In particular, linkers that can be cleaved off after the conjugate has entered the cell, thereby preventing the inhibition of RNAi associated with the disruption of RISC assembly formation. Such linkers can contain disulfide,[180] thioether[179] or pH sensitive bonds,[181,182] or they may contain bonds that can be cleaved photochemically[183].

2.11.2.1 Bioconjugates with lipids

Cholesterol and lipids were suggested as ligands for conjugation with siRNAs due to their interaction with the cell membrane. In addition, their lipophilic properties allowed endogenous transport mechanisms into the cells.[184] Cholesterol is transported into cells via lipoprotein particles, which are low-density and high-density lipoproteins (LDL and HDL), respectively,[185] which bind to corresponding receptors. Cholesterol-conjugated siRNAs can form complexes with HDL and LDL particles and then penetrate into the cell by receptor-mediated endocytosis.[186] Subsequently the cholesterol-siRNA and lipoprotein conjugates dissociate, and the siRNA enters the cytoplasm to participate in gene silencing. Cholesterol is not unique in its ability to connect siRNAs to lipoprotein particles. Other highly lipophilic molecules, such as long-chain fatty acids are also effective in binding to lipoproteins and mediating siRNA uptake into cells. A critical factor determining the affinity of fatty acid–conjugated siRNAs to lipoprotein particles is the length of the alkyl chain, which is a major determinant of lipophilicity. One interesting approach applied in order to silence rabies virus (RABV) nucleoprotein (N protein) and phosphoprotein (P protein) using anandamide (arachidonoylethanolamine, AEA) modified siRNA dendrimers (Figure 16) was reported by the Carell group.[187] The dendrimer enters the cell via cannabinoid receptors, which are expressed on immune and neural cells. Thus, this approach enables the targeted delivery of siRNAs into these sensitive cells that typically are hard to transfect.[188]
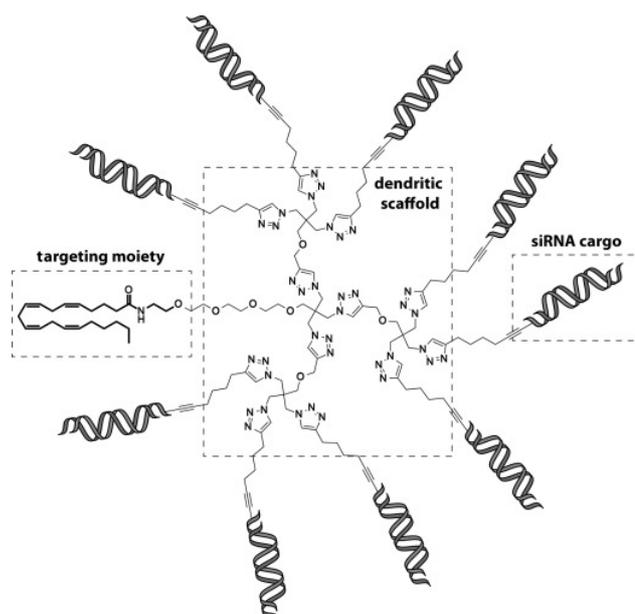
Figure 16. A dendritic siRNA nanostructure with an anandamide targeting moiety. Adapted from Brunner *et al*.[187]

A drawback of lipid-conjugated siRNAs is that usually they are not delivered to specific tissues and most of the material will be delivered to primary clearance tissues, including liver, kidney and spleen. Nevertheless, lipid-siRNAs could be used to silence targets in disease tissues, or targets that are widely expressed.[189,190]

2.11.2.2 Bioconjugates with peptides

The ability of peptides to interact specifically with proteins on the cell surface due to specific elements in their tertiary structure can be employed for targeted delivery of siRNAs.[191-193] For example, a conjugate of siRNA and the cyclic peptide "D-(Cys-Ser-Lys-Cys)-Gly" mimicking insulin-like growth factor 1 (IGF-1), successfully allowed penetration into breast cancer cells expressing the IGF-1-specific receptor. The siRNA conjugate achieved to silence the target gene by 60% without the use of transfection reagents.[191] The cyclo(Arg-Gly-Asp) (cRGD) peptide is known to bind to αvβ3 integrin receptors that play an important role in human cell metastasis and tumor-induced angiogenesis. During the animal studies it was shown that cRGD can specifically direct conjugated siRNA into αvβ3-expressing cells, resulting in 55% knockdown of targeted genes, which decreased the tumor growth.[193] Many peptide and siRNA conjugates were made, and such delivery approach seems to be promising. However, the toxic and immunogenic effects of such conjugates remain an issue.

2.11.2.3 Bioconjugates with carbohydrates

Being thoroughly investigated by biotechnology companies, carbohydrates, mostly *N*-acetylgalactosamine (GalNAc)- modified siRNA conjugates seem to allow the delivery of RNAi therapeutics. Tris-GalNAc binds to the asialoglycoprotein receptors (ASGPRs) that are predominantly expressed on hepatocytes (main tissue on the liver), which leads to rapid endocytosis.[174,194] Manoharan has demonstrated that GalNAc-conjugated siRNAs can provoke strong RNAi-mediated gene silencing in hepatocytes *in vitro* and *in vivo* without using transfection reagents.[174] Moreover, high efficiency was reached while using therapeutically relevant doses. The high number and recycling of ASGPR receptors contribute to the delivery of the GalNAc-siRNA conjugate. After the conjugate binds to ASGPR, it quickly gets internalized into clathrin-coated endosomes. When the endosomal pH drops, ASGPR separates from the conjugate and is recycled back to the cell surface while GalNAc-siRNA conjugate stays in the endosome. Then endosomal glycosidases cleave GalNAc from siRNA. The vast majority of siRNA then remains in the endosome and only a very small amount (less than 1%) escapes the endosomal lipid bilayer membrane via an unknown mechanism to induce an RNAi response. *Givosiran*[195,192] (Figure 17) developed by *Alnylam Pharmaceuticals* is a GalNAc-siRNA conjugate approved by the FDA in 2019 as a drug used for a treatment of acute hepatic porphyrias. The molecule is fully modified and is lacking the overhang on the 3' end of the sense strand.
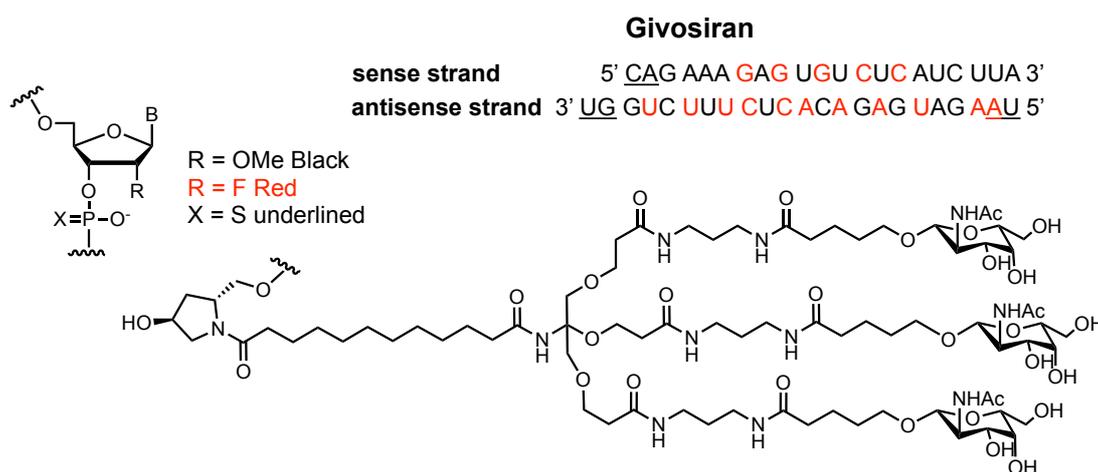
Figure 17. Schematic depiction of *Givosiran*. GalNAc moiety is attached to the 3' end of the sense strand.

Conjugates with other carbohydrates as delivery agents were also investigated, although the data were not very promising. Aviñó *et al*. tested siRNA duplexes against tumor necrosis factor carrying one, two or four glucose or galactose residues at the 5' terminus of the sense strand.[196] However, the carbohydrate-modified siRNAs had no inhibitory properties, unless oligofectamine as a transfection reagent was used.

2.12 Future outlook

The use of chemical modifications in siRNAs allowed to reduce the off-targeting effects. Bioconjugation strategies improved the ability of siRNA to penetrate in certain cells without transfection reagents. Systematic delivery of RNAi therapeutics to the liver reached a clinical reality after *Patisiran* was marketed. Many current siRNA-drug candidates target rare diseases. The therapeutic potential of siRNA could be expanded when the delivery to non-liver and non-kidney tissues becomes feasible in the clinical settings. The low bioavailability of siRNA conjugates and unfavorable pharmacokinetics, which, together with the rather high cost of obtaining such drugs in therapeutically relevant quantities, impede the transfer of drugs from the laboratory bench to the clinic.

3 Aim of the project

Naturally occurring non-canonical amino acid modified RNA bases were proposed by Carell to be living fossils of a prebiotic RNA world. The goal of the first part of the thesis was to investigate the possibility whether the amino acid modified nucleosides such as k^2C, acp^3U, t^6A, $(m)nm^5U$ are able to form peptides in an RNA-templated manner as a kind of a proto-ribosome. To achieve this, we planned to prepare the phosphoramidite building blocks of the mentioned modified bases and to incorporate them into RNA oligonucleotides. Then we planned to investigate potential RNA-templated peptide synthesis. This was thought to allow us to get insights into the principles of how the translational machinery might have evolved.

The aim of the second part of the thesis was triggered by the coronavirus pandemic. The idea was to utilize modified RNA bases for the chemical derivatization of siRNAs in order to use them as a treatment for SARS-CoV-2 infection. Incorporation of U_m and C_m into siRNAs to increase the stability against nucleases was planned. By connecting siRNA with biomolecules via "click" chemistry, we aimed to create a delivery platform for siRNA bioconjugates for entering lung cells via receptor-mediated uptake or natural transport mechanisms.

4 Published work

4.1. Synthesis of an acp³U phosphoramidite and incorporation of the hypermodified base into RNA

Milda Nainytė, Tynchtyk Amatov, Thomas Carell, *Chem. Comm.* **2019**, *55*, 12216-12218.

Introduction and summary

RNA contains a vast majority of bases and many of them are modified. 3-(3-amino-3-carboxy-*n*-propyl)uridine (acp³U) is a hypermodified base that is found in the tRNAs and rRNA of prokaryotes and eukaryotes. Although the existence of acp³U in tRNA and rRNA is conserved, its function is still unknown. It was proposed that acp³U complexes Mg-ions, which could stabilize the regional RNA structure. As a hypermodified base, in which a nucleoside is covalently connected to an amino acid, acp³U is a natural nucleoside between genotype and phenotype and hence also of particular importance for theories about the origin of life. Herein, we reported the development of a phosphoramidite building block and of a solid phase protocol that allows to synthesize RNA containing acp³U.

In this work we designed and synthesized a phosphoramidite building block acp³U. We also developed an efficient solid phase RNA synthesis protocol that allowed us to prepare RNA containing this base.

Declaration of contribution

In this work I was responsible for the design, synthesis and purification of all the compounds that were synthesized. Furthermore, I synthesized all oligonucleotides and analyzed them by MALDI-TOF mass spectrometry and HPLC analysis. I also performed the measurements of melting points.

Authorization

COMMUNICATION

Check for updates

Synthesis of an acp³U phosphoramidite and incorporation of the hypermodified base into RNA†

Milda Nainytė, Tynchtyk Amatov and Thomas Carell ⓘD *

acp³U is a hypermodified base that is found in the tRNAs of prokaryotes and eukaryotes and also in the ribosomal RNA of mammals. Its function has so far been unknown but it is speculated that acp³U complexes Mg ions, which may contribute to the stabilization of the RNA structure. As a hypermodified base in which a nucleoside is covalently connected to an amino acid, acp³U is a natural nucleoside between genotype and phenotype and hence is also of particular importance for theories about the origin of life. Herein, we report the development of a phosphoramidite building block and of a solid phase protocol that allows synthesis of RNA containing acp³U.

RNA contains a vast majority of modified bases.[1] Until recently, it was believed that such bases are predominantly present in transfer-RNAs (tRNAs) and ribosomal-RNAs (rRNAs), but with the discovery that messenger-RNAs (mRNAs) also contain modified bases, the whole field has regained enormous attention.[2–5] Modified bases can be divided into two groups. Most modified nucleosides (first group) deviate from the canonical bases by just small structural changes such as methylation(s) or exchange of oxygen by sulfur atoms. Other modified bases can be categorized into the group of hypermodified bases, which feature large structural changes compared to the canonical bases.[6] The biosynthesis of the bases in this group requires complex and often not completely understood biosynthesis machinery.

Related to the question of the origin of life, those hypermodified bases, which are modified with amino acids, are of particular interest.[7] These nucleosides feature properties of the genotype (information encoding potential) and the phenotype in a sense that they are in principle able to catalyse chemical transformations similar to the amino acids present in proteins. Indeed, hypermodified bases, which are "charged" with an amino acid, such as acp³U, hn/g/t⁶A, tm⁵s²U, k²C and agm²C, were suggested to be relics of an ancient code (Fig. 1).[8–10] As

such, they could have been part of an early translational system, within the concept of a prebiotic RNA-peptide world.

In order to investigate the properties of the hypermodified, amino acid charged nucleosides, it is essential to have phosphoramidite building blocks and solid phase synthesis procedures for their incorporation into RNA. Here we report the development of a phosphoramidite building block for acp³U (Fig. 1) together with an efficient solid phase RNA synthesis protocol that allows preparing RNA containing this hypermodified base acp³U (**1**).
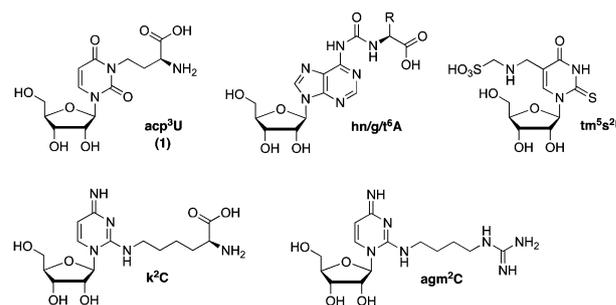
Fig. 1 Depiction of hypermodified bases that are charged with an amino acid.

For the synthesis of the phosphoramidite building block of **1** (**1-PA**), we first prepared the properly protected ʟ-homoserine **2**. The design of the appropriate protecting group strategy for ʟ-homoserine **2** was guided by the need to have sufficient stability during RNA solid phase oligonucleotide synthesis. Also, the protecting groups need to be easily removable after the synthesis. Importantly, this has to be possible without racemization of the ʟ-amino acid. Based on previous studies regarding the compatibility of protecting groups with solid phase oligonucleotide synthesis, we initially selected 9-fluoromethoxycarbonyl (Fmoc)[11,12] and 4-nitrophenylethyl (npe)[13] for deeper investigation. *N*-Fmoc protection was successfully achieved, while attempts to form ester **3a** (Scheme 1) lead to the formation of a stable lactone.[14] Although, in principle, it was possible to obtain compound **3a** *via* blocking of the γ-OH group, experiments in parallel to protect the amino- and

Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany. E-mail: thomas.carell@lmu.de
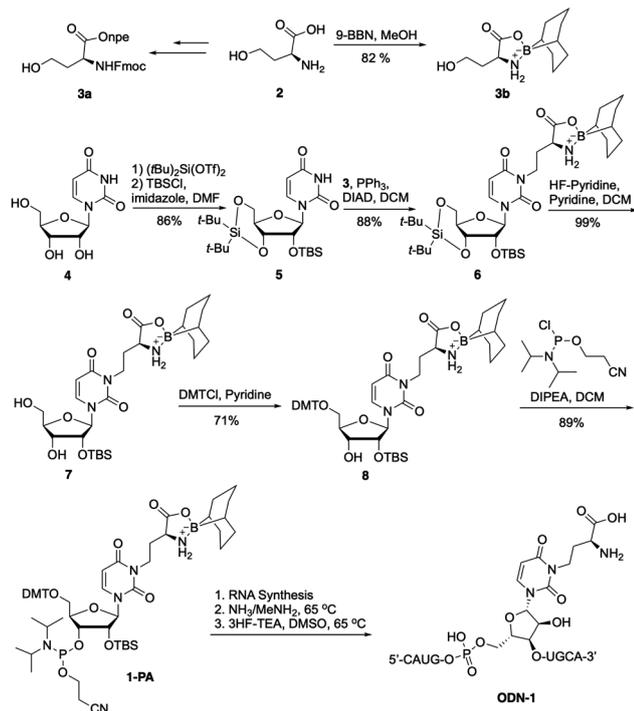
Scheme 1 Synthesis of the acp³U phosphoramidite (**1-PA**) with a 9-BBN protecting group and incorporation of **1-PA** into the oligonucleotide **ODN1**.

carboxy groups simultaneously utilising 9-BBN gave better yields and allowed avoiding lactonization. Stirring of **2** in MeOH under reflux with 9-BBN for 4 h provided the protected L-homoserine building block **3b** in 82% yield.[15] In parallel, uridine **4** was first 5′-3′-silyl-protected followed by protection of the still free 2′-OH with TBS-Cl using imidazole as the base.[16–18] The fully protected uridine building block **5** was obtained in 86% yield. We next connected the protected L-homoserine unit **3b** with **5** to afford **6** (88% yield) using a Mitsunobu reaction with Ph₃P and DIAD in dichloromethane.[19] Subsequent deprotection of the cyclic 5′-3′ protecting group with HF in pyridine furnished compound **7** (99% yield), which was then converted into the 5'-DMT protected acp³U compound **8**.[20–23] Reaction of **8** with 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite[24] gave finally the target compound **1-PA** in excellent 89% yield.

We next investigated the incorporation of **1-PA** using standard RNA solid phase oligonucleotide synthesis.[25–31] The main question was how the 9-BBN group would behave during the complex protocol. To the best of our knowledge, the 9-BBN protecting group has never been used before in solid phase oligonucleotide chemistry. We used a standard RNA synthesis protocol and were surprised that **1-PA** could be incorporated into strands such as **ODN1** just like the canonical bases. Neither an extension of the coupling time nor double coupling often required to obtain a decent yield for the insertion of a modified unit was required. The 9-BBN protecting group did not create problems, making it ideally suited for the incorporation of amino acid modified bases. We next cleaved the synthesized strand from the CPG-solid support material under concomitant deprotection of the base protecting groups with an aqueous solution of NH₃ and MeNH₂ followed by deprotection of the 2′-OTBS group with triethyl-

amine trihydrofluoride in DMSO over 1.5 h at 65 °C. This leads to complete 2′-OTBS deprotection. Under these conditions we also noted cleavage of the 9-BBN group. An additional deprotection step with **ODN1**, to our delight, was not required.

Fig. 2a shows the raw HPL chromatograms obtained directly after cleavage of **ODN1** from the support and of all the protecting groups. The inset depicts **ODN1** after one round of HPLC purification using a C18-column and a water (acetic acid, triethylamine buffer)–acetonitrile gradient. The MALDI-TOF spectrum depicted in Fig. 2b shows a clean spectrum with the expected *m/z*-value for $[M–H]^-$ of 2913 (*m/z*-value expected = 2913 for $[M–H]^-$). Both the already clean raw-HPLC chromatogram and the mass spectrum prove the successful incorporation of acp³U (**1**) into RNA.
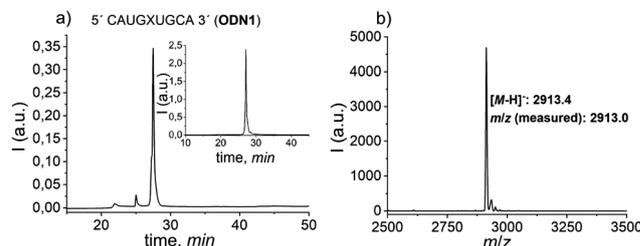
Fig. 2 (a) Raw-HPL chromatogram of **ODN1**, with the inset showing the HPL chromatogram of purified **ODN1**. (b) MALDI-TOF mass spectrum of **ODN1** after purification.

In order to show that our protocol also allows the preparation of longer RNA strands containing even two acp³U modifications, we finally successfully prepared the 22-mer RNA strand **ODN4** (seq. 5′-GACUGAC**acp³U**CGUAGC**acp³U**AACUCAU-3′). The synthesis went as expected very well. The purity and structural integrity were assessed by HPLC and MALDI-TOF (ESI†). We continued our studies with **ODN1**. For acp³U we expected a strong destabilizing effect on duplex formation due to the substitution of the N3 required for standard Watson–Crick base pairing. In order to investigate to which extent the L-homoserine at N3 causes duplex destabilization, we measured melting temperatures for **ODN1**: **ODN2** relative to the standard duplex **ODN3:ODN2**. The obtained melting curves are depicted in Fig. 3. Clearly evident is that acp³U reduces the melting temperature dramatically by an unusual 24 °C, proving the extremely strong duplex destabilizing nature of this amino acid charged hypermodified base.
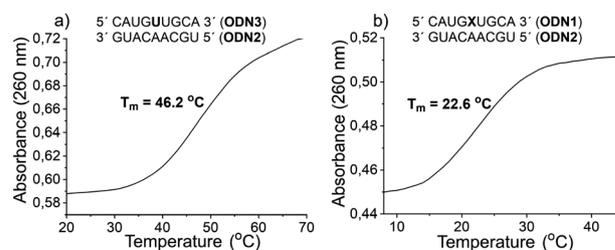
Fig. 3 (a) Depiction of the melting curve obtained with the duplex ODN3:ODN2 as the reference. (b) Depiction of the melting curve obtained with the duplex ODN1:ODN2.

In summary, we report here the development of a phosphoramidite building block for the hypermodified base acp^{3}U. We show that the acp^{3}U base can be almost perfectly incorporated into small and even long oligonucleotides using the building block **1-PA** with the 9-BBN protecting group for the L-homoserine amino acid. We also show and quantify that the destabilizing effect of acp^{3}U is unusually high.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Howkawicz, K. M. Rother, M. Helm, J. M. Bujnicki and H. Grosjean, *Nucleic Acids Res.*, 2013, **41**, D262–D267.
2 W. V. Gilbert, T. A. Bell and C. Schaening, *Science*, 2016, **352**, 14018.
3 B. S. Zhao, I. A. Roundtree and C. He, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 31–42.
4 S. R. Jaffrey, *Nat. Struct. Mol. Biol.*, 2014, **21**, 945–946.
5 E. M. Harcourt, A. M. Kietrys and E. T. Kool, *Nature*, 2017, **541**, 339–346.
6 M. T. Wattsa and I. Tinoco, *Biochemistry*, 1978, **17**, 2455–2463.
7 C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier and T. Carell, *Angew. Chem., Int. Ed.*, 2018, **57**, 5943–5946.
8 M. Di Giulio, *J. Theor. Biol.*, 1998, **191**, 191–196.
9 H. Grosjean, V. de Crecy-Lagard and G. R. Björk, *Trends Biochem. Sci.*, 2004, **29**, 519–522.
10 E. Szathmary, *Trends Genet.*, 1999, **15**, 223–229.
11 P. Gaytan, J. Yanez, F. Sanchez, H. Mackie and X. Soberon, *Chem. Biol.*, 1998, **5**, 519–527.
12 H. Schirmeister-Tichy, G. G. Alvarado and W. Pfleiderer, *Nucleosides Nucleotides*, 1999, **18**, 1219–1220.
13 F. Himmelsbach, B. S. Schulz, T. Trichtinger, R. Charubala and W. Pfleiderer, *Tetrahedron*, 1984, **40**, 59–72.
14 A. J. Ozinskas and G. A. Rosenthal, *J. Org. Chem.*, 1986, **51**, 5047–5050.
15 W. H. Dent III, W. R. Erickson, S. C. Fields, M. H. Parker and E. G. Tromiczak, *Org. Lett.*, 2002, **4**, 1249–1251.
16 C. Chaix, A. M. Duplaa, D. Molko and R. Teoule, *Nucleic Acids Res.*, 1989, **17**, 7381–7393.
17 J. Stawinski, R. Strömberg, M. Thelin and E. Westman, *Nucleic Acids Res.*, 1988, **16**, 9285–9298.
18 S. M. Ching, W. J. Tan, K. L. Chua and Y. Lam, *Bioorg. Med. Chem.*, 2010, **18**, 6657–6665.
19 M. de Champdore, L. De Napoli, G. Di Fabio, A. Messere, D. Montesarchio and G. Piccialli, *J. Chem. Soc., Chem. Commun.*, 2001, 2598–2599.
20 B. M. Trost and C. G. Caldwell, *Tetrahedron Lett.*, 1981, **22**, 4999–5002.
21 M. Sekine, S. Limura and K. Furusawa, *J. Org. Chem.*, 1993, **58**, 3204–3208.
22 H. Schaller, G. Weimann, B. Lerch and H. G. Khorana, *J. Am. Chem. Soc.*, 1963, **85**, 3821–3827.
23 K. L. Agarwal, A. Yamazaki, P. J. Cashion and H. G. Khorana, *Angew. Chem., Int. Ed. Engl.*, 1972, **11**, 451–459.
24 N. Usman, K. K. Ogilvie, M. Y. Jiang and R. J. Cedergren, *J. Am. Chem. Soc.*, 1987, **109**, 7845–7854.
25 H. Seliger, D. Zeh, G. Azuru and J. B. Chattopadhyaya, *Chem. Scr.*, 1983, **22**, 95–102.
26 M. H. Caruthers, D. Dellinger, K. Prosser, A. D. Brone, J. W. Dubendorff, R. Kierzek and M. Rosendahl, *Chem. Scr.*, 1986, **26**, 25–30.
27 R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner and S. M. Freier, *Biochemistry*, 1986, **25**, 7840–7846.
28 I. Hirao, M. Ishikawa and K. Miura, *Nucleic Acids Symp. Ser.*, 1985, **16**, 173–176.
29 H. Tanimura, T. Fukuzawa, M. Sekine, T. Hata, J. W. Efcavitch and G. Zon, *Tetrahedron Lett.*, 1988, **29**, 577–578.
30 H. Tanimura, M. Sekine and T. Hata, *Nucleosides Nucleotides*, 1986, **5**, 363–383.
31 H. Tanimura and T. Imada, *Chem. Lett.*, 1990, 1715–1718.

4.2. Synthesis and Incorporation of k²U into RNA

Introduction and summary

Lysidine (k²C) is a cytidine nucleobase in which a lysine amino acid is attached to the C^2 position. k²C is an amino acid modified base that combines genotype and phenotype due to a nucleobase and an amino acid combination. Lysidine was suggested to be a molecular fossil from an early RNA world and thus it makes the compound particularly interesting in the context of origin of translation.

In this work we designed and synthesized the uridine analogue of k²C (abbreviated as k²U), which in principle is a deamination product of k²C that likely occurred under prebiotic conditions. We showed that the base can be successfully incorporated into RNA strands. After melting point and NMR studies we showed that k²U exists in an isoC-like tautomeric form and thus dramatically destabilizes the RNA duplexes.

Declaration of contribution

In this work I performed the design, synthesis and purification of the synthesized compounds.

Authorization

FULL PAPER

Synthesis and Incorporation of k²U into RNA

Milda Nainyte[a] and Thomas Carell*[a]

[a] Department of Chemistry, Ludwig-Maximilians-Universität, Butenandtstr. 5–13, DE-81377 Munich,
e-mail: Thomas.carell@lmu.de

Dedicated to Prof. *François Diederich* on the occasion of his retirement

Lysidine (k²C) is one of the most modified pyrimidine RNA bases. It is a cytidine nucleoside, in which the 2-oxo functionality of the heterocycle is replaced by the ε-amino group of the amino acid lysine. As such, lysidine is an amino acid-containing RNA nucleoside that combines directly genotype (C-base) with phenotype (lysine amino acid). This makes the compound particularly important in the context of theories about the origin of life and here especially for theories that target the origin of translation. Here, we report the total synthesis of the U-derivative of lysidine (k²U), which should have the same base pairing characteristics as k²C if it exists in the isoC-like tautomeric form. To investigate this question, we developed a phosphoramidite building block for k²U, which allows its incorporation into RNA strands. Within RNA, k²U can base pair with the counter base U and isoG, confirming that k²U prefers an isoC-like tautomeric structure that is also known to dominate for k²C. The successful synthesis of a k²U phosphoramidite and its use for RNA synthesis now paves the way for the preparation of a k²C phosphoramidite and RNA strands containing k²C.

Keywords: lysidine, modified RNA bases, RNA, nucleosides, origin of life, prebiotic chemistry.

Introduction

RNA contains a vast variety of modified nucleosides.[1] Many of the modified bases are just methylated versions of the canonical nucleosides A, C, G and U, but other are highly modified. This is achieved with the help of dedicated biosynthesis machineries.[2] From a prebiotic point of view in which genotype-phenotype discussions address the question of whether life started with nucleic acids or peptides, the most interesting modified nucleosides are those which are modified with amino acids.[3] These molecules are chemical structures between genotype and phenotype. They directly merge the properties of nucleic acids with those of amino acids. As such, RNA containing these amino acid-modified bases could have been a central element for the origin of life and for the origin of translation as already discussed by *Grosjean* and others.[4–7] RNA containing these amino acids could establish catalytic properties next to information encoding functions. Catalytic RNAs in turn

are the central elements in all discussions about how life emerged from simple starting materials.[8,9]

One of the most interesting amino acid modified nucleosides is lysidine (k²C),[10,11] which is a cytidine base to which a lysine amino acid is attached with the ε-amino group to the C(2) position (*Figure 1*). Agmatidine (agm²C)[12,13] is a close relative, which features the guanidinium group found in the amino acid arginine (*Figure 1*). We also want to mention the amino acid modified RNA base puromycin in which the amino acid is attached to the C(3′)-OH group. This base is in use for ribosomal studies.[14,15] In our effort to investigate the properties of RNA containing amino acid-modified nucleosides as units that equip such RNA potentially with peptide-like properties, we were interested to study the lysine modified uridine base. This compound is a close derivative of lysidine, which we name k²U. k²U is in principle the deamination product of k²C which has likely occurred under prebiotic conditions, where the bases were exposed to potentially hot aqueous environments.[16,17] Under
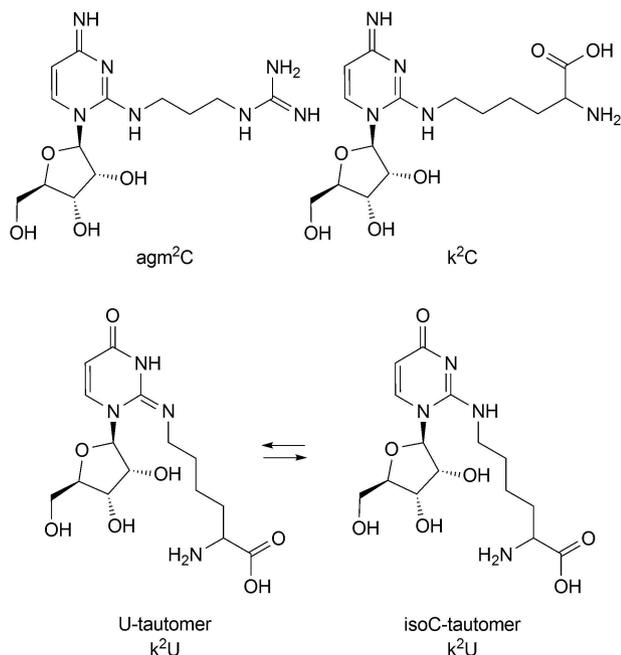
Figure 1. Depiction of agmatidine (agm^{2}C), lysidine (k^{2}C) and of the potential prebiotically relevant deamination product (k^{2}U) of k^{2}C.

these conditions, deamination of k^{2}C to k^{2}U is an expected process. k^{2}U can in principle exist is two tautomeric states. In one it has U-type base pairing properties, while in the second it should behave like an isoC similar to k^{2}C, which also exists in a quinoid-like tautomeric structure (*Figure 1*)[10,18–23]. In this latter scenario, k^{2}U and k^{2}C are supposed to have similar base pairing properties. A major difference could be that while k^{2}C is protonated, k^{2}U is likely not. k^{2}C protonation is supposed to occur at N(6) and hence it does not change the base pairing properties.[10,24]

Results and Discussion

To investigate how k^{2}U is affecting the structure and properties of RNA, we started the synthesis of k^{2}U as its phosphoramidite building block (**k^{2}U-PA**) and investigated procedures that allow its incorporation into RNA. The synthesis of the **k^{2}U-PA** is depicted in *Scheme 1*. The synthesis was started with carboxybenzoyl- (Cbz) protected lysine **1**, which we converted with 1-({[2-(trimethylsilyl)ethoxy]carbonyl}oxy)pyrrolidine-2,5-dione into the Cbz- and Teoc-protected lysine compound **2**. Subsequent protection of the carboxy group with 2-(trimethylsilyl)ethanol furnished the fully protected lysine amino acid **3**. After Cbz-deprotection,

we obtained the amino acid coupling partner **4**. At the nucleoside side, we treated uridine **5** under *Mitsunobu* conditions with DIAD and PPh$_3$ to obtain the literature-known 2′,3′-protected cyclouridine compound **6**.[25] Coupling of this intermediate with the lysine building block **4** in the presence of LiCl and DBU furnished the protected lysine coupled uridine derivative **7**. We next protected the primary 5′-OH group with DMTCl and the secondary 2′-OH group with TBSCl to give the intermediate **9**.[26–28] Compound **9** was finally converted into the phosphoramidite building block using a standard procedure.[29] Purification of the **k^{2}U-PA** was difficult due to its high polarity. We needed to use a rather polar mixture of dichloromethane/acetone (8:3) as the mobile phase for the chromatographic purification. This provided, however, the target compound **k^{2}U-PA** in a total yield of 12% in just eight steps in a good but not excellent purity.

To study the properties of k^{2}U in RNA, we next inserted the compound into an RNA strand using solid phase RNA synthesis.[30–33] The oligonucleotide synthesis was performed using standard coupling conditions. This allowed us to achieve a coupling yield of 30%. Importantly, the incorporation of the next canonical bases, *e.g.*, coupling of uridine to k^{2}U, was not affected. We achieved elongation yields that reached typically 95–98%. All together the method provided enough material for all further studies. Optimization of the k^{2}U-incorporation yield was consequently not performed. In addition, we noted that the purity of the obtained strands was high allowing rapid separation of the target k^{2}U containing oligonucleotide. After the RNA synthesis, we removed the Teoc- and TMSE-protecting groups from the lysine moiety with saturated ZnBr$_2$ solution in isopropanol/nitromethane (1:1) at r.t. over night. It is interesting that the RNA strand is stable under these quite *Lewis* acidic conditions. Our observation, however, agrees with an earlier report.[34] Own attempts to cleave the Teoc- and TMSE-protecting groups with HF·Et$_3$N complex provided only partial deprotection which gave a mixture of products. In our hands, the reported ZnBr$_2$ method gave superior results.[33] RNA degradation was not observed.

We subsequently cleaved the oligonucleotide from the solid support and removed the protecting groups from the canonical bases with a mixture of methylamine and ammonia (1:1) at 65 °C for 5 min. *Figure 2* shows the sequence of the prepared oligonucleotide together with the raw HPL-chromatogram and the MALDI-TOF spectrum. These data prove the integrity

TMSE: 2-(trimethylsilyl)ethyl
Teoc: [2-(trimethylsilyl)ethoxy]carbonyl
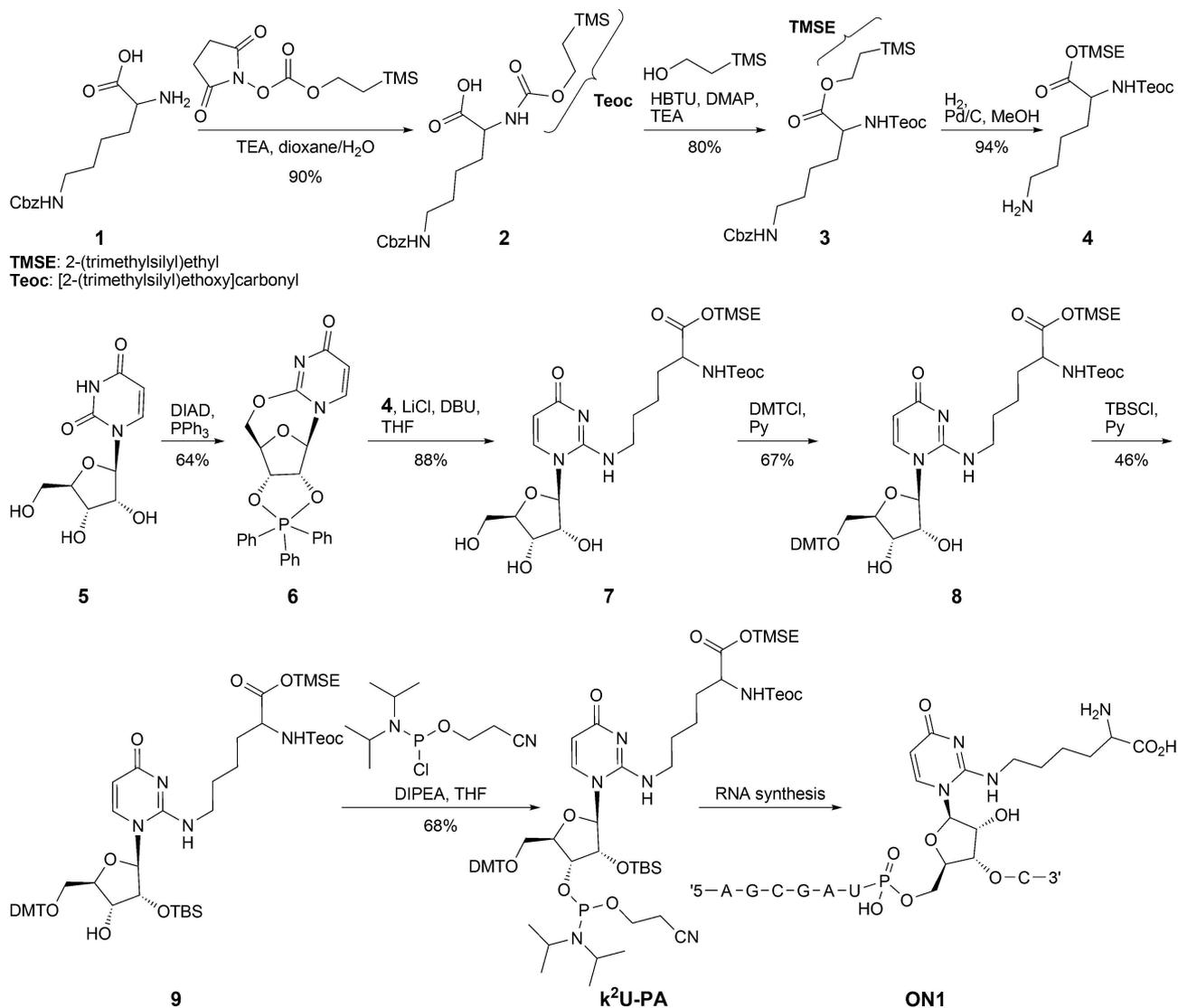
Scheme 1. Synthesis of k²U and of its phosphoramidite building block **k²U-PA**.
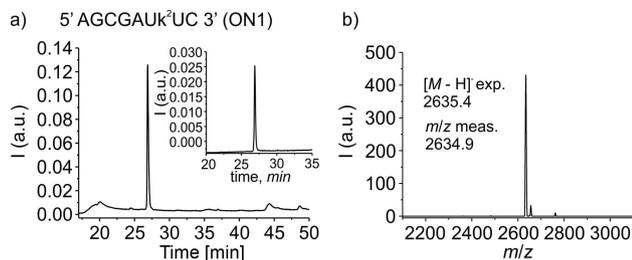
Figure 2. a) Sequence of the k²U-containing RNA strand together with the raw-HPL chromatogram of **ON1**. The inset shows the HPL-chromatogram of the purified **ON1**. b) MALDI-TOF mass spectrum of **ON1** after purification.

and the high purity of the synthesized k²U-containing RNA strand.

We finally studied how the k²U base affects the stability of the RNA duplex. *Figure 3* shows the melting curves of the k²U-containing oligonucleotide (k²U:A base pair, red) in comparison to the unmodified RNA duplex containing a U:A base pair (black). The table in *Figure 3* depicts all melting points measured for k²U facing any of the four canonical bases, together with all possible combinations of canonical bases as reference strands. While the U:A reference strand melts at 42 °C, replacement of U by k²U reduces the melting temperature (k²U:A) to just 32 °C, which is a dramatic destabilization. This is unexpected and not explainable with a U-type tautomeric structure, because the lysine residue can in principle point out of the large shallow groove of the RNA-duplex in A-
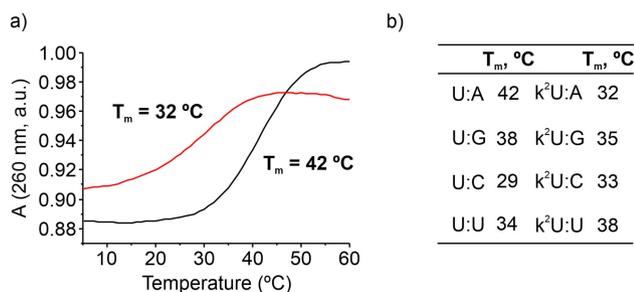
Figure 3. a) Depiction of melting curve with the duplex containing U:A base pair (black), and k^2U:A (red); b) table of all melting points measured for k^2U.

conformation. Because the C(2)-O atom, which is replaced by the lysine residue, does not take part in the H-bonding to the A-counterbase as depicted in *Figure 1*, the lysine residue should not affect base pairing so strongly if k^2U would exist in the U-tautomeric form. Further melting point studies in which we exchanged systematically the counter base showed the k^2U is unable to undergo any productive base pairing. The only slight stabilization that we detected is in the k^2U:U situation (38 °C), which is compared to the U:U (34 °C) situation stabilized by 4 °C.

This rather large global destabilization is best explained, if we assume that the k^2U base exists indeed not in the typical U-tautomer but in the hemiquinoid tautomeric state known from isoC (*Figure 4*). As such, k^2U behaves like an isoC tautomer, which indeed prohibits k^2U to form productive base pairs with any of the other canonical bases particularly with the purine bases A and G. The slight interaction with U is in this scenario explained because the U counterbase can get engaged with isoC-tautomeric k^2U with two H-bonds. To investigate this isoC type
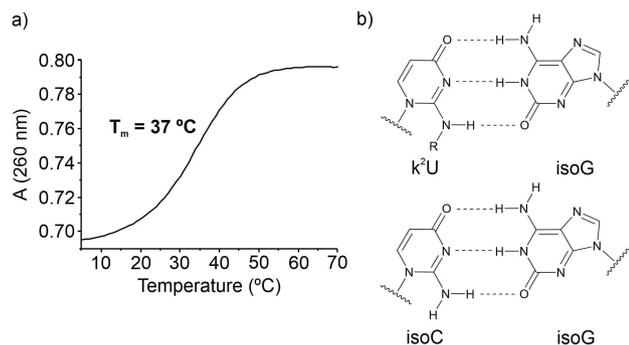
structure of k^2U in more detail, we created a duplex in which k^2U is facing isoG as the counterbase. Indeed, in this situation we measure a higher melting point of 37 °C (*Figure 4*) close to the k^2U:U situation (38 °C), in fully agreement with the idea that k^2U is a lysine modified isoC derivative that has an k^2C like tautomeric structure. Further support for the idea that the k^2U base exists predominantly in the quinoid tautomeric structure comes from NMR data. In our compounds with the k^2U base the typical NH ^{1}H-NMR signal around $\delta = 9.5$ ppm was not observed. Instead we observed the NH signal at around $\delta = 7.0$ ppm, in line with the quinoid tautomeric structure of k^2U (*Figure 4*). This shift to around $\delta = 7.0$ ppm agrees with literature data about such compounds.[35]

Conclusions

Here, we report the development of a phosphoramidite building block of k^2U which is a close deamination-based relative of the non-canonical base k^2C. We show that the modified base can be incorporated into RNA strands using standard phosphoramidite chemistry in combination with a three-stage deprotection protocol, in which we first cleave the protecting groups at the lysine residue, followed by deprotection of the nucleobases and cleavage from the resin. Melting point studies and NMR data show that k^2U exists in a tautomeric state that resembles the situation in isoC. As such, k^2U destabilizes RNA duplexes dramatically allowing only limited interactions with U and most importantly isoG as counterbases.

Experimental Section

General Methods

Chemicals were purchased from *Sigma–Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros organics* and used without further purification. Strands containing canonical bases and isoG were purchased from *Metabion*. The solvents were of reagent grade or purified by distillation, unless otherwise specified. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel *F254* TLC plates from *Merck KGaA*. Flash chromatography was performed on *Geduran® Si60* (40–63 μm) silica gel from *Merck KGaA*. NMR spectra were recorded on *Bruker AVIIIHD 400* spectrometers (400 MHz). ^{1}H-NMR shifts were calibrated to the residual solvent resonan-

Figure 4. a) Depiction of melting curve with the duplex containing k^2U:isoG base pair; b) depiction of base pairing of isoC:isoG and k^2U:isoG.

ces: (D$_6$)DMSO (2.50 ppm), CDCl$_3$ (7.26 ppm), (D$_6$) acetone (2.05 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: (D$_6$)DMSO (39.52 ppm), CDCl$_3$ (77.16 ppm), (D$_6$)acetone (29.84 ppm). All NMR spectra were analyzed using MestRENOVA 10.01.1 from Mestrelab Research S. L. High resolution mass spectra were measured on the spectrometer *MAT 90* (ESI) from *Thermo Finnigan GmbH*. Analytical RP-HPLC was performed on an analytical HPLC *Waters Alliance* (*2695 Separation Module*, *2996 Photodiode Array Detector*) equipped with the column *Nucleosil 120 – 2C18* from *Macherey Nagel* using a flow of 0.5 ml/min, a gradient of 0 – 30% of buffer *B* in 45 min was applied. Preparative RP-HPLC was performed on a HPLC *Waters Breeze* (*2487 Dual λ Array Detector*, *1525 Binary* HPLC Pump) equipped with the column *VP 250/32C18* from *Macherey Nagel* using a flow of 5 ml/min, a gradient of 0 – 25% of buffer *B* in 45 min was applied for the purifications. Oligonucleotides were purified using the following buffer system: buffer *A*: 100 mM NEt$_3$/AcOH (pH 7.0) in H$_2$O and buffer *B*: 100 mM NEt$_3$/AcOH in 80% (*v/v*) acetonitrile. The pH values of buffers were adjusted using a *MP 220* pH-meter (*Metter Toledo*). Oligonucleotides were detected at wavelength: 260 nm. Melting profiles were measured on a *JASCO V-650* spectrometer. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0. For strands containing artificial bases, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a *Bruker Autoflex II*. For MALDI-TOF measurements, the samples were desalted on a 0.025 μm *VSWP* filter (*Millipore*) against ddH$_2$O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

N^6-[(Benzyloxy)carbonyl]-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysine (2).

The reaction was performed according to the published procedure.[36]

To a stirred suspension of cbz-protected lysine (**1**) (3 g, 10.7 mmol, 1 equiv.) in dioxane/water (1 : 1) mixture, TEA (2.23 ml, 16 mmol, 1.5 equiv.) was added. To the resultant solution, Teoc–OSu (3.2 g, 12.3 mmol, 1.1 equiv.) was added and the mixture was left to stir overnight at room temperature. Afterwards, the mixture was diluted with water, acidified with saturated KHSO$_4$ solution and extracted with diethyl ether. The combined organic layers were washed with water, dried over Na$_2$SO$_4$ and evaporated *in vacuo*. Yield: 85%. ^{1}H-NMR (400 MHz, CDCl$_3$): 7.35 – 7.33 (*m*, 5 H);

5.50 – 5.45 (*m*, 1 H); 5.09 (*s*, 2 H); 4.41 – 4.32 (*m*, 1 H); 4.15 (*t*, *J* = 8.4, 2 H); 3.22 – 3.13 (*m*, 2 H); 1.86 – 1.72 (*m*, 2 H); 1.51 – 1.40 (*m*, 4 H); 0.97 (*t*, *J* = 8.4, 2 H); 0.02 (*s*, 9 H). ^{13}C-NMR (101 MHz, CDCl$_3$): 176.4; 168.9; 156.8; 136.6; 128.6; 128.4; 128.3; 66.9; 63.7; 53.5; 40.6; 31.9; 29.5; 22.3; 17.7; −1.4. HR-ESI-MS: 425.2103 (C$_{20}$H$_{33}$N$_2$O$_6$Si$^+$, [*M* + H]$^+$; calc. 425.2108).

2-(Trimethylsilyl)ethyl N^6-[(benzyloxy)carbonyl]-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (3).

To the stirred solution of compound **2** (1.2 g, 2.82 mmol, 1 equiv.) in CH$_2$Cl$_2$/DMF (1 : 1) mixture, HBTU (1.61 g, 4.2 mmol, 1.5 equiv.), TEA (0.788 ml, 5.65 mmol, 2 equiv.) and DMAP (0.034 g, 0.27 mmol, 0.1 equiv.) were added. Then, 2-(trimethylsilyl)ethanol (0.486 ml, 4.2 mmol, 1.2 equiv.) was added and the mixture was left to stir overnight at room temperature. Then, CH$_2$Cl$_2$ was evaporated and product was extracted with diethyl ether and washed with water. The organic phase was dried and evaporated. The residue was purified by flash chromatography eluting with hexane/AcOEt (4 : 1). Yield: 80%. ^{1}H-NMR (400 MHz, CDCl$_3$): 7.35 – 7.30 (*m*, 5 H); 5.19 (*d*, *J* = 8, 1 H); 5.08 (*s*, 2 H); 4.81 (br. *s*, 1 H); 4.32 – 4.26 (*m*, 1 H); 4.24 – 4.18 (*m*, 2 H); 4.14 (*t*, *J* = 9, 2 H); 3.18 (*q*, *J* = 6.6, 2 H); 1.82 – 1.62 (*m*, 2 H); 1.55 – 1.50 (*m*, 2 H); 1.42 – 1.34 (*m*, 2 H); 1.00 – 0.96 (*m*, 4 H); 0.04 (*s*, 9 H); 0.02 (*s*, 9 H). ^{13}C-NMR (101 MHz, CDCl$_3$): 172.8; 156.6; 156.4; 136.7; 128.6; 128.4; 128.2; 66.7; 63.9; 63.5; 53.7; 40.7; 32.4; 29.5; 22.4; 17.8; 17.5; −1.36; −1.39. HR-ESI-MS: 525.2817 (C$_{25}$H$_{45}$N$_2$O$_6$Si$_2^+$, [*M* + H]$^+$; calc. 525.2816).

2-(Trimethylsilyl)ethyl N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (4).

Cbz-protected lysine derivative **3** (0.7 g, 1.33 mmol, 1 equiv.) was dissolved in methanol. Pd/C (0.096 g) was added, and the mixture was stirred at room temperature under H$_2$ at atmospheric pressure. After 2 h, the mixture was filtered through *Celite* and evaporated. Yield: 94%. ^{1}H-NMR (400 MHz, CDCl$_3$): 5.18 (*d*, *J* = 8, 1 H); 4.33 – 4.26 (*m*, 1 H); 4.23 – 4.18 (*m*, 2 H); 4.14 (*t*, *J* = 8.5, 2 H); 2.69 (br. *s*, 2 H); 1.80 – 1.34 (*m*, 6 H); 0.98 (*q*, *J* = 8.5, 4 H); 0.04 (*s*, 9 H); 0.03 (*s*, 9 H). ^{13}C-NMR (101 MHz, CDCl$_3$): 172.9; 156.4; 64.0; 63.4; 53.8; 42.2; 32.6; 28.1; 22.6; 17.8; 17.4; −1.37; −1.40. HR-ESI-MS: 391.2444 (C$_{17}$H$_{39}$N$_2$O$_4$Si$_2^+$, [*M* + H]$^+$; calc. 391.2448).

2′,3′-*O*-(Triphenylphosphoranediyl)-*O*^{2},5′-cyclouridine (= (3a*R*,4*R*,12*R*,12a*R*)-2,2,2-Triphenyl-4,5,12,12a-tetrahydro-2*H*,3a*H*,8*H*-4,12-epoxy-2λ^5-[1,3,2]dioxaphospholo[4,5-*e*]pyrimido[2,1-*b*][1,3]-oxaz-

ocin-8-one; **6**). The reaction was performed according to the published procedure.[25]

To a stirred solution of uridine **5** (5 g, 20.5 mmol, 1 equiv.) in THF, PPh₃ (16.1 g, 61.4 mmol, 3 equiv.) was added. Then DIAD (12.1 ml, 61.4 mmol, 3 equiv.) was added dropwise, and the reaction was left to stir at room temperature overnight. Afterwards the precipitate was filtered, washed with THF and water, and then dried. Yield: 64%. ^{1}H-NMR (400 MHz, (D₆)DMSO): 8.05 (*d*, *J* = 7.5, 1 H); 7.67 – 7.51 (*m*, 2 H); 7.47 – 7.17 (15 H); 5.96 – 5.85 (*m*, 2 H); 4.84 (*dd*, *J* = 6.6, 13.2, 1 H); 4.73 (*s*, 1 H); 4.60 – 4.46 (*m*, 2 H); 4.16 (*d*, *J* = 12.7, 1 H). ^{13}C-NMR (101 MHz, (D₆)DMSO): 170.7; 157.0; 144.8; 143.6; 130.8; 128.5; 127.6; 109.0; 97.8; 85.4; 76.8; 74.4; 72.7. HR-ESI-MS: 487.1412 ($C_{27}H_{24}N_2O_5P^+$, $[M+H]^+$; calc. 487.1423).

2-(Trimethylsilyl)ethyl N^6-(4-oxo-1-pentofuranosyl-1,4-dihydropyrimidin-2-yl)-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (7).

To the dry LiCl (0.043 g, 1.02 mmol, 5 equiv.) THF was poured followed by the addition of DBU (0.156 g, 1.02 mmol, 5 equiv.) and amine **4** (0.16 g, 0.41 mmol, 2 equiv.). Then, cyclouridine derivative **6** (0.1 g, 0.2 mmol, 1 equiv.) was added, and the mixture was left to stir for 2 h at room temperature. In time, the solution became pink and clear. Afterwards, the mixture was evaporated, and the residue was purified by flash chromatography eluting with MeCN/H₂O (95:5, *v/v*). Yield: 88%. ^{1}H-NMR (400 MHz, (D₆)DMSO): 7.82 (*s*, 1 H); 7.58 (*d*, *J* = 7.7, 1 H); 7.44 (*dd*, *J* = 7.7, 4.6, 1 H); 7.12 (*t*, *J* = 5.2, 1 H); 5.53 (*d*, *J* = 7.7, 1 H); 5.50 (*d*, *J* = 7.0, 1 H); 5.44 (*t*, *J* = 4.6, 1 H); 5.39 (*d*, *J* = 7.0, 1 H); 5.29 (*d*, *J* = 4.3, 1 H); 4.16 – 4.08 (*m*, 4 H); 4.04 – 3.97 (*m*, 4 H); 3.20 (*q*, *J* = 6.7, 2 H); 2.74 (*t*, *J* = 7.6, 1 H); 1.66 – 1.43 (*m*, 4 H); 1.37 – 1.26 (*m*, 2 H); 0.95 – 0.89 (*m*, 4 H); 0.02 (*s*, 18 H). ^{13}C-NMR (101 MHz, (D₆)DMSO): 172.7; 163.5; 156.5; 150.9; 140.9; 101.9; 87.7; 73.7; 70.0; 62.7; 62.0; 60.9; 53.9; 51.9; 30.8; 30.1; 26.7; 22.6; 17.5; 16.9; −1.32; −1.35. HR-ESI-MS: 617.3028 ($C_{26}H_{49}N_4O_9Si_2^+$, $[M+H]^+$; calc. 617.3038).

2-(Trimethylsilyl)ethyl N^6-(1-{5-O-[bis(4-methoxyphenyl)(phenyl)methyl]pentofuranosyl}-4-oxo-1,4-dihydropyrimidin-2-yl)-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (8).

To the solution of nucleoside **7** (0.45 g, 0.73 mmol, 1 equiv.) in dry pyridine, DMTCl (0.37 g, 1.1 mmol, 1.5 equiv.) was added. The mixture was left to stir overnight at room temperature. The product was extracted with AcOEt and washed with sat. NaHCO₃ solution. The combined

organic layers were evaporated, and the residue was purified by flash chromatography eluting with CH₂Cl₂/MeOH (10:1, *v/v*) containing 0.02% of pyridine. Yield: 67%. ^{1}H-NMR (400 MHz, (D₆)acetone): 7.86 (*d*, *J* = 7.8, 1 H); 7.49 – 7.44 (*m*, 2 H); 7.38 – 7.20 (*m*, 8 H); 7.02 (*t*, *J* = 5.3, 1 H); 6.92 – 6.87 (*m*, 4 H); 6.48 (*d*, *J* = 8.2, 1 H); 5.71 (*d*, *J* = 4.3, 1 H); 5.57 (*d*, *J* = 7.7, 1 H); 4.69 (br. *s*, 1 H); 4.49 – 4.47 (*m*, 1 H); 4.43 – 4.41 (*m*, 1 H); 4.23 (*q*, *J* = 3.5, 1 H); 4.18 – 4.04 (*m*, 5 H); 3.77 (*s*, 6 H); 3.47 (*d*, *J* = 3.5, 2 H); 3.41 – 3.30 (*m*, 2 H); 1.83 – 1.52 (*m*, 4 H); 1.48 – 1.38 (*m*, 2 H); 1.00 – 0.92 (*m*, 4 H); 0.03 (*s*, 9 H); 0.02 (*s*, 9 H). ^{13}C-NMR (101 MHz, (D₆)acetone): 173.4; 172.2; 159.7; 157.3; 154.3; 145.7; 139.1; 136.4; 136.2; 131.0; 128.9; 128.7; 127.7; 114.0; 106.4; 92.1; 87.5; 85.4; 75.5; 71.2; 63.6; 63.1; 55.6; 55.1; 41.9; 32.2; 23.9; 18.3; 17.9; −1.3; −1.4. HR-ESI-MS: 919.4326 ($C_{47}H_{67}N_4O_{11}Si_2^+$, $[M+H]^+$; calc. 919.4345).

2-(Trimethylsilyl)ethyl N^6-(1-{5-O-[bis(4-methoxyphenyl)(phenyl)methyl]-2-O-[*tert*-butyl(dimethyl)silyl]pentofuranosyl}-4-oxo-1,4-dihydropyrimidin-2-yl)-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (9).

DMT-protected nucleoside **8** (0.1 g, 0.1 mmol, 1 equiv.) was dissolved in pyridine. Then, imidazole (0.022 g, 0.3 mmol, 3 equiv.) and TBSCl (0.049 g, 0.3 mmol, 3 equiv.) were added. The mixture was left to stir overnight at room temperature. Then, the mixture was poured into saturated NaHCO₃ solution and extracted with CH₂Cl₂. Organic layers were evaporated, and the residue was purified by flash chromatography eluting with CH₂Cl₂/acetone (8:3, *v/v*) containing 0.02% of pyridine. Yield: 46%. ^{1}H-NMR (400 MHz, (D₆)acetone): 7.49 – 7.44 (*m*, 2 H); 7.38 – 7.29 (*m*, 9 H); 6.91 (*d*, *J* = 8.9, 5 H); 5.67 (*d*, *J* = 5.3, 1 H); 5.53 (*d*, *J* = 7.7, 1 H); 4.45 (*t*, *J* = 5.3, 1 H); 4.41 – 4.35 (*m*, 1 H); 4.21 – 4.14 (*m*, 4 H); 4.13 – 4.06 (*m*, 4 H); 3.79 (*s*, 6 H); 3.55 (*dd*, *J* = 11.0, 3.0, 1 H); 3.38 (*dd*, *J* = 11.0, 3.0, 1 H); 3.31 (*t*, *J* = 6.7, 2 H); 1.81 – 1.67 (*m*, 2 H); 1.65 – 1.50 (*m*, 2 H); 1.47 – 1.34 (*m*, 2 H); 1.02 – 0.95 (*m*, 2 H); 0.86 (*s*, 9 H); 0.11 (*s*, 3 H); 0.05 (*s*, 9 H); 0.03 (*s*, 9 H). ^{13}C-NMR (101 MHz, (D₆)acetone): 173.4; 170.9; 159.7; 157.2; 154.3; 145.6; 138.9; 136.2; 136.1; 131.0; 128.9; 128.7; 127.8; 114.0; 107.8; 91.4; 87.7; 85.9; 75.9; 71.7; 64.2; 63.5; 63.0; 55.5; 55.1; 41.9; 32.2; 26.1; 24.1; 18.7; 18.3; 17.9; −1.37; −1.41; −4.5; −4.7. HR-ESI-MS: 1033.5205 ($C_{53}H_{81}N_4O_{11}Si_3^+$, $[M+H]^+$; calc. 1033.5210).

2-(Trimethylsilyl)ethyl N^6-[1-(5-O-[bis(4-methoxyphenyl)(phenyl)methyl]-2-O-{*tert*-butyl(dimethyl)silyl]-3-O-{(2-cyanoethoxy)[di(propan-2-yl)amino]phosphanyl}pentofuranosyl)-4-oxo-1,4-dihydropyrimidin-2-yl]-N^2-{[2-(trimethylsilyl)ethoxy]car-

bonyl}lysinate (**k²U-PA**). Compound **9** (0.1 g, 0.09 mmol, 1 equiv.) was co-evaporated three times in dry pyridine. Then, the compound was dissolved in dry THF and cooled to 0 °C. DIPEA (67 µl, 0.39 mmol, 4 equiv.) and 2-cyanoethyl-*N,N*-diisopropyl chlorophosphoramidite (54 µl, 0.24 mmol, 2.5 equiv.) were added under nitrogen atmosphere. The mixture was brought to room temperature and stirred for 2.5 h. Then, the reaction was quenched by the addition of saturated $NaHCO_3$ solution and extracted with CH_2Cl_2. The combined organic layers were dried and concentrated. The residue was purified by flash chromatography eluting with CH_2Cl_2/acetone (8:3, *v/v*) containing 0.3% of pyridine. Yield: 68%. ^{31}P-NMR (162 MHz, (D_6)acetone): 149.9; 148.6. HR-ESI-MS: 1233.6278 ($C_{62}H_{98}N_6O_{12}PSi_3^+$, $[M+H]^+$; calc. 1233.6288).

Synthesis and Purification of Oligonucleotide

The oligonucleotide was synthesized on a 1 µmol scale using a DNA automated synthesizer (*Applied Biosystems 394 DNA/RNA Synthesizer*) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleotides were purchased from *Glen Research* and *Sigma-Aldrich*. Oligonucleotide containing k²U nucleoside was synthesized in DMT-OFF mode using phosphoramidites (Bz–A, Dmf-G, Ac–C, U) with BTT in MeCN as an activator, DCA in CH_2Cl_2 as a deblocking solution and Ac_2O in pyridine/THF as a capping reagent. After the synthesis the solid support was treated with saturated solution of $ZnBr_2$ in ^iPrOH/$MeNO_2$ (1:1, *v/v*, 1 ml) and left overnight at room temperature. Then the beads were washed with water and 0.1 M EDTA solution. The cleavage and deprotection of CPG-bound oligonucleotide were performed with aq. NH_4OH/$MeNH_2$ solution (1:1, *v/v*, 1 ml) for 5 min at 65 °C. The resin was removed by filtration, washed with H_2O and the solution was evaporated under reduced pressure. The residue was subsequently heated with a solution of triethylamine trihydrofluoride (125 µl) in DMSO (50 µl) at 65 °C for 1.5 h. Upon cooling in an ice bath, AcONa (3 M, 25 µl) and BuOH (1 ml) were added. The resulting suspension was vortexed and cooled in a freezer (−80 °C) for 1 h. After the centrifugation, supernatant was removed, and the remaining oligonucleotide pellet was dried under vacuum. The oligonucleotide was analyzed and purified using RP-HPLC. The structural integrity of RNA was analyzed by MALDI-TOF mass measurement.

Author Contribution Statement

All syntheses and characterization of the compounds and measurements were carried out by *M. N. T. C.* designed the study and supervised the experimental work. *M. N.* and *T. C.* wrote the manuscript.

References

[1] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Howkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, 'MODOMICS: a database of RNA modification pathways – 2013 update', *Nucleic Acids Res.* **2013**, *41*, D262–D267.

[2] M. T. Watts, I. Tinoco, 'Role of hypermodified bases in transfer RNA. Solution properties of dinucleoside monophosphates', *Biochemistry* **1978**, *17*, 2455–2463.

[3] H. Grosjean, E. Westhof, 'An integrated, structure- and energy-based view of the genetic code', *Nucleic Acids Res.* **2016**, *44*, 8020–8040.

[4] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, 'Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth – Generation by Prebiotic Methylations and Carbamoylations', *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946; *Angew. Chem.* **2018**, *130*, 6050–6054.

[5] M. Di Giulio, 'Reflections on the Origin of the Genetic Code: A Hypothesis', *J. Theor. Biol.* **1998**, *191*, 191–196.

[6] H. Grosjean, V. de Crécy-Lagard, G. R. Björk, 'Aminoacylation of the anticodon stem by a tRNA-synthetase paralog: relic of an ancient code?', *Trends Biochem. Sci.* **2004**, *29*, 519–522.

[7] E. Szathmáry, 'The origin of the genetic code: amino acids as cofactors in an RNA world', *Trends Genet.* **1999**, *15*, 223–229.

[8] T. R. Cech, A. J. Zaug, P. J. Grabowski, '*In Vitro* splicing of the ribosomal RNA precursor of tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence', *Cell* **1981**, *27*, 487–496.

[9] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, 'The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme', *Cell* **1983**, *35*, 849–857.

[10] S. Yokoyama, T. Muramatsu, N. Horie, K. Nishikawa, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, T. Miyazawa, 'A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*', *Pure Appl. Chem.* **1989**, *61*, 573–576.

[11] B. J. Kopina, C. T. Lauhon, 'Efficient Preparation of 2,4-Diaminopyrimidine Nucleosides: Total Synthesis of Lysidine and Agmatidine', *Org. Lett.* **2012**, *14*, 4118–4121.

[12] D. Mandal, C. Köhrer, D. Su, S. P. Russell, K. Krivos, C. M. Castleberry, P. Blum, P. A. Limbach, D. Söll, U. L. RajBhandary, 'Agmatidine, a modified cytidine in the anticodon of archaeal tRNA^Ile, base pairs with adenosine but not with guanosine', *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2872–2877.

[13] Y. Ikeuchi, S. Kimura, T. Numata, D. Nakamura, T. Yokogawa, T. Ogata, T. Wada, T. Suzuki, T. Suzuki, 'Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea', *Nat. Chem. Biol.* **2010**, *6*, 277–282.

[14] S. V. Melnikov, N. F. Khabibullina, E. Mairhofer, O. Vargas-Rodrigues, N. M. Reynolds, R. Micura, D. Söll, Y. S. Polikanov, 'Mechanistic insights into the slow peptide bond formation with D-amino acids in the ribosomal active site', *Nucleic Acids Res.* **2019**, *47*, 2089–2100.

[15] S. Sothiselvam, S. Neuner, L. Rigger, D. Klepacki, R. Micura, N. Vázquez-Laslop, A. S. Mankin, 'Binding of Macrolide Antibiotics Leads to Ribosomal Selection against Specific Substrates Based on Their Charge and Size', *Cell Rep.* **2016**, *16*, 1789–1799.

[16] R. Shapiro, 'Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life', *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4396–4401.

[17] H. Lönnberg, R. Käppi, 'Competition between the hydrolysis and deamination of cytidine and its 5-substituted derivatives in aqueous acid', *Nucleic Acids Res.* **1985**, *13*, 2451–2456.

[18] R. M. Voorhees, D. Mandal, C. Neubauer, C. Köhrer, U. L. RajBhandary, V. Ramakrishnan, 'The structural basis for specific decoding of AUA by isoleucine tRNA on the ribosome', *Nat. Struct. Mol. Biol.* **2013**, *20*, 641–643.

[19] A. Soma, Y. Ikeuchi, S. Kanemasa, K. Kobayashi, N. Ogasawara, T. Ote, J. Kato, K. Watanabe, Y. Sekine, T. Suzuki, 'An RNA-Modifying Enzyme that Governs Both the Codon and Amino Acid Specificities of Isoleucine tRNA', *Mol. Cell* **2003**, *12*, 689–698.

[20] Y. Ikeuchi, A. Soma, T. Ote, J. Kato, Y. Sekine, T. Suzuki, 'Molecular Mechanism of Lysidine Synthesis that Determines tRNA Identity and Codon Recognition', *Mol. Cell* **2005**, *19*, 235–246.

[21] C. Switzer, S. E. Moroney, S. A. Benner, 'Enzymatic incorporation of a new base pair into DNA and RNA', *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.

[22] J. A. Piccirilli, S. A. Benner, T. Krauch, S. E. Moroney, 'Enzymatic incorporation of a new base into DNA and RNA extends the genetic alphabet', *Nature* **1990**, *343*, 33–37.

[23] T. Suzuki, K. Miyauchi, 'Discovery and characterization of tRNA^Ile lysidine synthetase (TilS)', *FEBS Lett.* **2010**, *584*, 272–277.

[24] T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, S. Yokoyama, 'Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification', *Nature* **1988**, *336*, 179–181.

[25] J. Kimura, K. Yagi, H. Suzuki, O. Mitsunobu, 'Studies on Nucleosides and Nucleotides. VIII. Preparation and Reactions of Triphenylphosphoranediylnucleosides', *Bull. Chem. Soc. Jpn.* **1980**, *53*, 3670–3677.

[26] P. T. Gilham, H. G. Khorana, 'Studies on Polynucleotides. A New and General Method for the Chemical Synthesis of the C_5''-C_3'' Internucleotidic Linkage. Synthesis of Deoxyribo-dinucleotides', *J. Am. Chem. Soc.* **1958**, *80*, 6212–6222.

[27] M. Smith, D. H. Rammler, I. H. Goldberg, H. G. Khorana, 'Studies on Polynucleotides. XIV. Specific Synthesis of the C_3''-C_5'' Interribonucleotide Linkage. Syntheses of Uridylyl-(3''→5'')-Uridine and Uridylyl-(3''→5'')-Adenosine', *J. Am. Chem. Soc.* **1962**, *84*, 430–440.

[28] K. K. Ogilvie, A. L. Schifman, C. L. Penney, 'The synthesis of oligoribonucleotides. III. The use of silyl protecting groups in nucleoside and nucleotide chemistry. VIII', *Can. J. Chem.* **1979**, *57*, 2230–2238.

[29] N. Usman, K. K. Ogilvie, M.-Y. Jiang, R. J. Cedergren, 'Automated Chemical Synthesis of Long Oligoribonucleotides Using 2'-O-Silylated Ribonucleoside 3'-O-Phosphoramidites on a Controlled-Pore Glass Support: Synthesis of a 43-Nucleotide Sequence Similar to the 3'-Half Molecule of an *Escherichia coli* Formylmethionine tRNA', *J. Am. Chem. Soc.* **1987**, *109*, 7845–7854.

[30] R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, S. M. Freier, 'Polymer-Supported RNA Synthesis and Its Application to Test the Nearest-Neighbor Model for Duplex Stability', *Biochemistry* **1986**, *25*, 7840–7846.

[31] I. Hirao, M. Ishikawa, K. Miura, 'Solid-phase synthesis of oligoribonucleotides', *Nucleic Acids Symp. Ser.* **1985**, *16*, 173–176.

[32] H. Tanimura, T. Fukazawa, M. Sekine, T. Hata, J. W. Efcavitch, G. Zon, 'The practical synthesis of RNA fragments in the solid phase approach', *Tetrahedron Lett.* **1988**, *29*, 577–578.

[33] H. Tanimura, M. Sekine, T. Hata, 'Chemical Synthesis of RNA Fragments Related to the C4 N Hypothesis', *Nucleosides Nucleotides* **1986**, *5*, 363–383.

[34] F. Ferreira, F. Morvan, 'Silyl protecting groups for oligonucleotide synthesis removed by a $ZnBr_2$ treatment', *Nucleosides Nucleotides Nucleic Acids* **2005**, *24*, 1009–1013.

[35] A. A.-H. Abdel-Rahman, E. B. Pedersen, C. Nielsen, 'Synthesis of 5-methylisocytidine derivatives', *Monatsh. Chem.* **1996**, *127*, 455–459.

[36] R. E. Shute, D. H. Rich, 'Synthesis and Evaluation of Novel Activated Mixed Carbonate Reagents for the Introduction of the 2-(Trimethylsilyl)ethoxycarbonyl(Teoc)-Protecting Group', *Synthesis* **1987**, *4*, 346–349.

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

Milda Nainytė, Felix Müller, Giacomo Ganazzoli, Chun-Yin Chan, Antony Crisp, Daniel Globisch, Thomas Carell, *Chem. Eur. J.* **2020**, *26*, 14856-14860.

Introduction and summary

The RNA-peptide co-evolution hypothesis describes the parallel existence of peptides and RNA. Such tight coexistence of peptides and RNA is seen in the contemporary ribosomes, where the RNA part is responsible for peptide synthesis and proteins ensure translational fidelity. It seems plausible that some of the components from the extinct RNA-peptide world are still present until today. Indeed, amino acid modified bases found in tRNA were suggested to be molecular fossils. The most abundant amino acid modified bases are adenosines, containing an amino acid attached, via a urea connector, to the N^6-amino group of the heterocycle.

In this work we reported the synthesis of amino acid containing adenosine derivatives together with the evaluation of their physicochemical properties. We showed that these bases do not establish base pairing and have to be placed outside the pairing positions that are required for RNA folding. As such they behave like anchors allowing the connection of an amino acid to RNA structures independent of the counterbase.

Declaration of contribution

In this work I performed the synthesis of the Thr^{6}A, Ser^{6}A, Val^{6}A, Asp^{6}A, t^{6}dA phosphoramidite building blocks. The phosphoramidite of the glycine derivative (g^{6}A) and its incorporation into RNA oligonucleotide was performed by *Felix Müller*. *Giacomo Ganazzoli* synthesized the His^{6}A phosphoramidite building block. Phe^{6}A phosphoramidite building block and its incorporation into RNA oligonucleotide was made by *Chun-Yin Chan*. All the other oligonucleotides discussed in the publication were synthesized by me. I also performed the melting point measurements showing that aa^{6}A bases alone do not establish stable base pairing unless they are found in loop structures.

Authorization

Europe
European Chemical
Societies Publishing

▌ **Oligonucleotides**

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

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

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

The RNA-peptide co-evolution hypothesis describes the emergence of self-replicating molecules that contained amino acids and RNA.[1] At the macromolecular level, this tight coexistence of peptides and RNA is established in the ribosome, where encoding and catalytic RNA is supported by proteins.[2] Although

we cannot delineate how such an early RNA-peptide world may have looked like, it seems not too implausible to assume that some of the molecular components may have survived until today as vestiges of this extinct world.[3] tRNAs derived from all three kingdoms of life contain a large number of modified bases,[4] and some of them are indeed modified with amino acids.[3] The most wide spread amino acid modified bases are adenosine nucleosides, in which the amino acid is linked via urea connector to the N^6-amino group of the heterocycle as depicted in Figure 1a. Particularly ubiquitous are adenosine modifications containing the amino acids threonine (t6A)[5–7] and glycine (g6A),[8] together with hn6A.[9,10] Based upon recent phylogenetic analyses and the fact that t6A is found in all three kingdoms of life, it has been suggested that such amino acid modified bases were already present in the last universal common ancestor (LUCA), from which all life

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

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

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

Figure 1. (a) Depiction of the amino acid modified A-bases (aa6A) together with computer visualizations that show how such bases may reside in an (b) A-form RNA duplex and a (c) B-form DNA duplex. The sequence used for the visualization is: 5′-CAUAUAUAUAUG-3′ with A = g6A.

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

forms descended.[11–14] t⁶A is for example today found in nearly all ANN decoding tRNAs.[15] We recently reported a plausible prebiotic route to some of these amino acid modified A-bases, which strengthens the idea that they could indeed be living chemical fossils of the extinct RNA-peptide world.[16] Despite the interesting philosophical genotype-phenotype dualism that characterizes these structures and their contemporary importance for the faithful decoding of genetic information, a general synthesis of aa⁶A modified bases (Figure 1a) and a systematic study of their properties is lacking.

Here we report the synthesis of a variety of aa⁶A nucleosides with canonical amino acids (aa = Asp, Gly, His, Phe, Thr,[17] Ser, Val), their incorporation into DNA and RNA and an investigation of how they influence the physicochemical properties of oligonucleotides. We were particularly interested to study how they might affect the stability of RNA and DNA. The computer visualization shows that in A-form RNA (Figure 1b), the amino acid part of the aa⁶A base would need to reside inside the helix, shielded from the outside. In the B-form DNA one could imagine a decoration of the major groove with the amino acid side chains as depicted in Figure 1c.

In the Schemes 1 and 2 we show the synthesis of the different urea linked amino acid A-derivatives (aa⁶A). We first prepared the amino acid components for the coupling to the A-nucleoside (Scheme 1). Our starting points for Thr⁶A, Ser⁶A and Asp⁶A were the free amino acids 1–3, in which we first transformed all carboxylic acids into the p-nitrophenylethyl esters (npe, 4–6).[17] The hydroxy groups of the Thr and Ser compounds were finally protected as TBS-ethers to give the final products 7 and 8 (Scheme 1a). For Val, Gly and Phe we started with the Boc-protected amino acids 9–11, which we also converted into the npe-esters 12–14 using Mitsunobu type
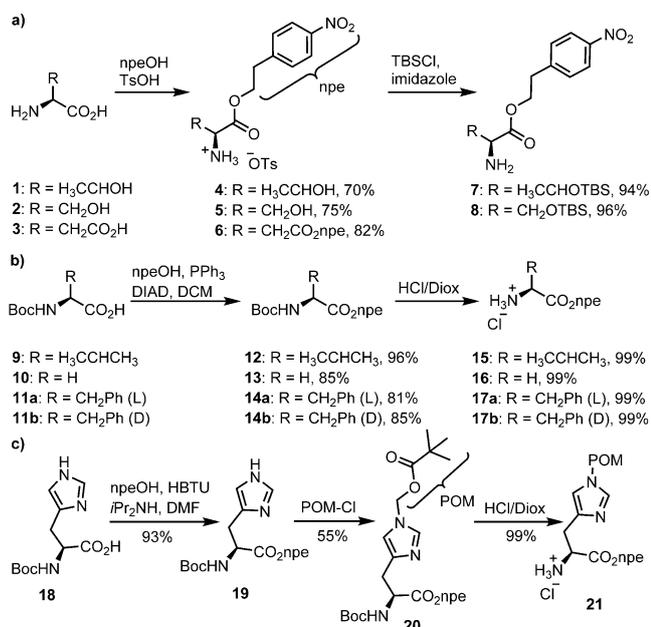
Scheme 1. Synthesis of the amino acid building blocks as needed for the coupling to the nucleoside A to give Thr⁶A, Ser⁶A, Asp⁶A, Val⁶A, Gly⁶A, Phe⁶A and His⁶A.

ON1: Thr⁶A; **ON2**: Gly⁶A; **ON3**: Val⁶A
ON4: His⁶A; **ON5**: Asp⁶A; **ON6**: L-Phe⁶A; **ON7**: D-Phe⁶A
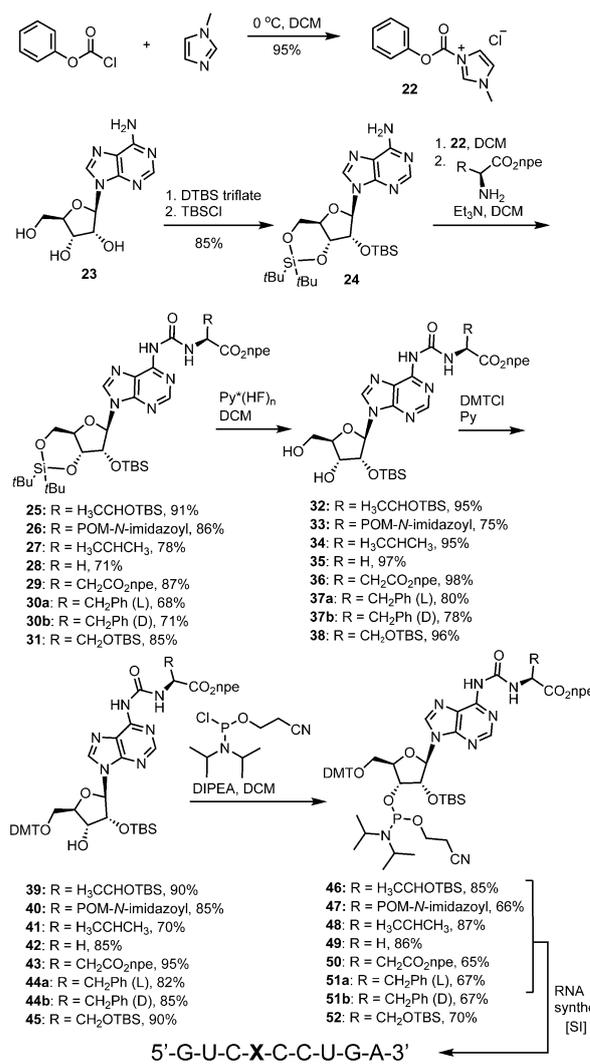
Scheme 2. Synthesis of phosphoramidite building blocks of Thr⁶A, Ser⁶A, Asp⁶A, Val⁶A, Gly⁶A, Phe⁶A and His⁶A and their incorporation into RNA.

chemistry[18] followed by acidic (4 M HCl in dioxane) Boc-deprotection to give the amino acid products 15–17 (Scheme 1b).[19] For His⁶A, we again started with the Boc-protected amino acid 18 (Scheme 1c) and used HBTU activation to generate the npe ester 19. Protection of the imidazole Nᵀ with POM-chloride followed again by Boc-deprotection furnished the ready to couple amino acid 21.

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

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

cyclic silylether with HF·pyridine complex,[22,23] protection of the 5′-OH group with dimethoxytritylchloride (DMTCl)[24] allowed the final conversion of the compounds into the corresponding phosphoramidites **46–52**. Standard solid phase RNA chemistry[25–31] was subsequently employed to prepare RNA strands containing the individual aa⁶A nucleosides stably embedded. The standard RNA synthesis protocol did not require any adjustment. In all cases we observed fair coupling of the aa⁶A phosphoramidites and no decomposition during deprotection. Deprotection required three steps. First, with DBU in THF at r.t. for 2 h we cleaved the npe-protecting group. Second, we deprotected the bases and cleaved from the solid support with aqueous NH₃/MeNH₂. Finally, we removed the 2′-silyl group with HF in NEt₃.

In order to investigate how aa⁶A bases would affect the stability of DNA duplexes we also prepared as a representative molecule t⁶dA as depicted in Scheme 3. To this end we first acetyl-protected dA **53**,[32] performed the coupling of the protected threonine with the activated carbonate **22**, cleaved the acetyl groups and converted the nucleoside subsequently into the 5′-DMT protected phosphoramidite **57**. The purification of compound **57** was quite difficult due to its high polarity. We needed to use rather polar mixture of EtOAc/Hex (2/1) as the mobile phase for the chromatographic separation. This provided the phosphoramidite **57**, however the material had a lower purity in comparison to the RNA phosphoramidites. Nevertheless, solid phase DNA synthesis and deprotection of the DNA strand **ODN1** proceeded again smoothly and in high yields.
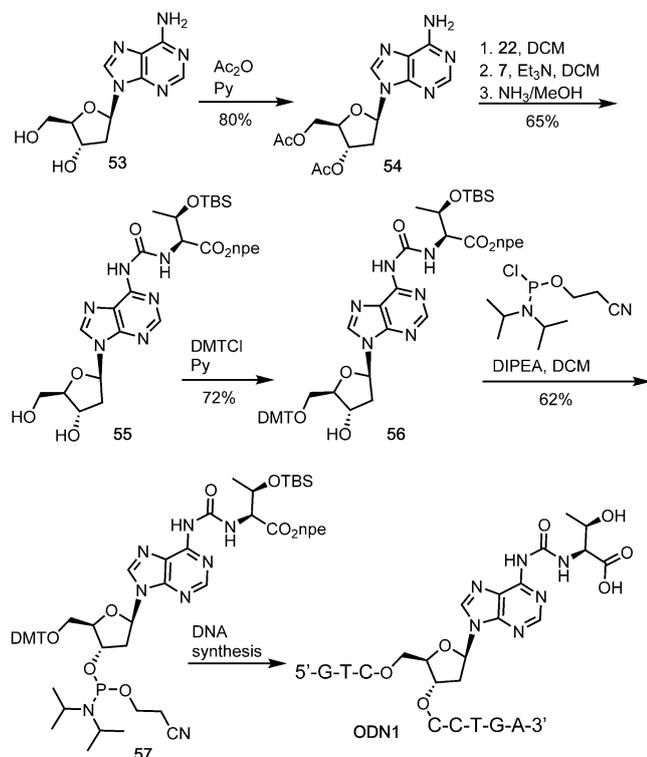
Figure 2a shows as an example the raw HPL-chromatograms of **ON1** (RNA strand with embedded t⁶A) and the corresponding chromatogram after purification (inset) together with the obtained MALDI-TOF mass spectrum (Figure 2b). The chromatograms of the raw material show a good quality of the obtained RNA material. The analytical chromatogram after purification and the MALDI-TOF data prove the purity of the finally obtained RNA oligonucleotide and the integrity of the t⁶A-containing RNA strand.

Figure 2c and 2d show the same data set for the t⁶dA containing DNA oligonucleotide (**ODN1**), proving again the successful synthesis of t⁶dA containing oligonucleotide. The aa⁶(d)A nucleosides can exist in two different conformations.[33] The first, s-trans, maintains the Watson–Crick hydrogen bonding capabilities with the urea amino acid oriented towards the imidazole ring system (Figure 3a). This allows formation of a Hoogsteen type 7-membered ring H-bond with the N^7. In the corresponding s-cis-conformation, the urea amino acid orients towards the Watson–Crick side thereby establishing a typically strong intramolecular 6-membered H-bond with N^1 (Figure 3b). In order to investigate if the embedding of the amino acid would enforce s-trans-conformation and hence Watson–Crick H-bonding, we measured melting points of all aa⁶A containing RNA strands and of the t⁶dA containing DNA strand hybridized to the corresponding counter strands (Figure 3). In the RNA:RNA situation we noted for all aa⁶A strands that we investigated, a single clear melting point, showing that only one conformer of the aa⁶A base likely exists in the RNA:RNA duplexes. In situation where the aa⁶A base exists in two different stable conformations, one would expect a more complex melting behaviour. In all cases we saw that the melting point is strongly reduced by 10–15 °C. When we embedded two aa⁶A building blocks into a short RNA strand no duplex formation

Scheme 3. Synthesis of t⁶dA phosphoramidite and its incorporation into DNA.
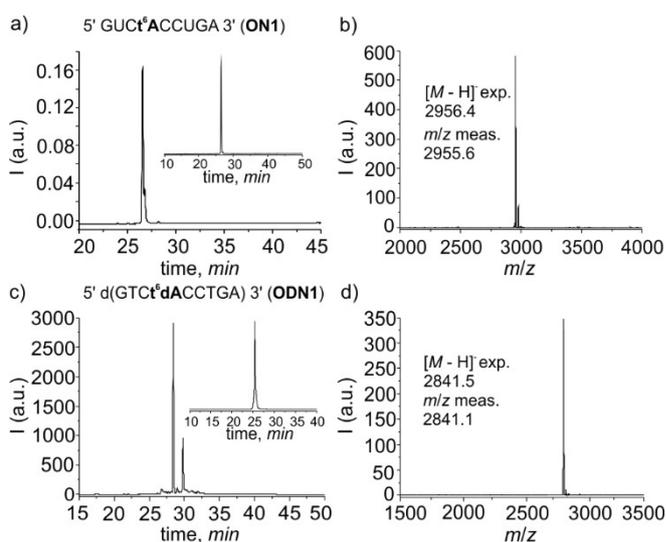
Figure 2. (a) Raw-HPL chromatogram of **ON1**, with the inset showing the HPL-chromatogram of purified **ON1**; (b) MALDI-TOF mass spectrum of **ON1** after purification; (c) raw-HPL chromatogram of **ODN1**, with the inset showing the HPL chromatogram of purified **ODN1**; (d) MALDI-TOF mass spectrum of **ODN1** after purification.

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing
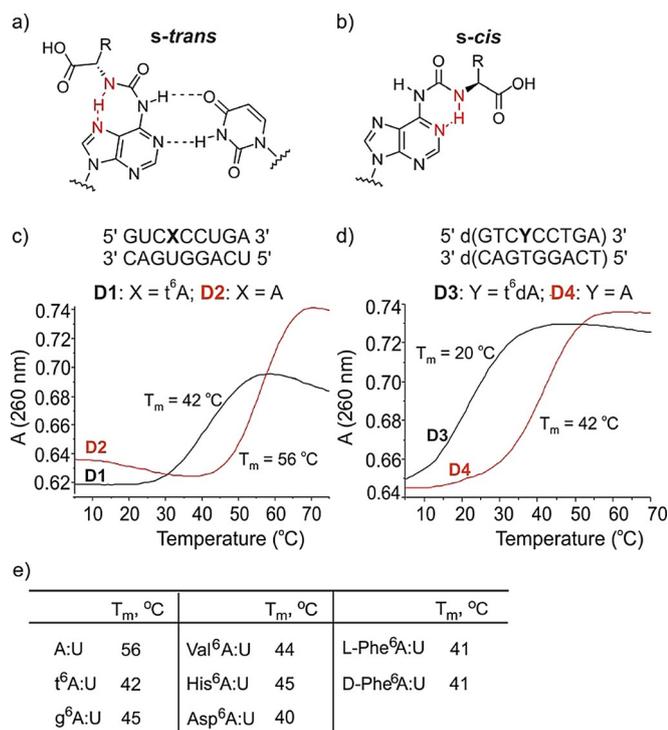
Figure 3. (a, b) Possible conformation, base pairing and intramolecular H-bond of aa⁶A; (c, d) melting curves measured for t⁶A containing RNA:RNA duplexes and of a t⁶dA containing DNA:DNA duplex in comparison with the duplexes containing canonical (d)A:(d)T base pairs; (e) table of the determined melting points.

was obtained. Even stronger reduction of the melting point was observed for the DNA duplex containing one t⁶dA. Here, we also saw just one sharp melting point and a reduction of the T_m by over 20 °C. These data show that the aa⁶A bases and among them t⁶A and g⁶A are unable to base pair. Although we have no direct proof of the structure the data argue for a preferred s-*cis*-conformation (Figure 3 b) in agreement with the literature.[34]

This conclusion is also supported by the observation that irrespective of the chirality of the attached amino acid (L- versus D-Phe⁶A), we measured the same melting temperature. This would not be expected if the s-*trans*-conformation and base pairing would be possible. These data suggest that aa⁶A nucleosides within RNA position a given amino acid outside the A-form helix in an unpaired situation and hence independent from the counterbase. As such, multiple aa⁶A containing RNA strands would be structures in which the RNA part is decorated by the amino acid side chains. In order to show that RNA-structures containing multiple amino acids as representatives of an RNA-peptide world can stably form, we prepared two RNA duplexes (Figure 4). In the first (**D5**), we placed three t⁶A bases as extra bases in an otherwise undisturbed RNA duplex. Indeed, now the stability of this duplex was indistinguishable from the same construct containing just canonical bases (**D6**). Finally, we prepared an RNA duplex **D7**, in which we placed the amino acids Ser-Asp-His directly next to each other to simulate what is known in the peptide world as the catalytic triad
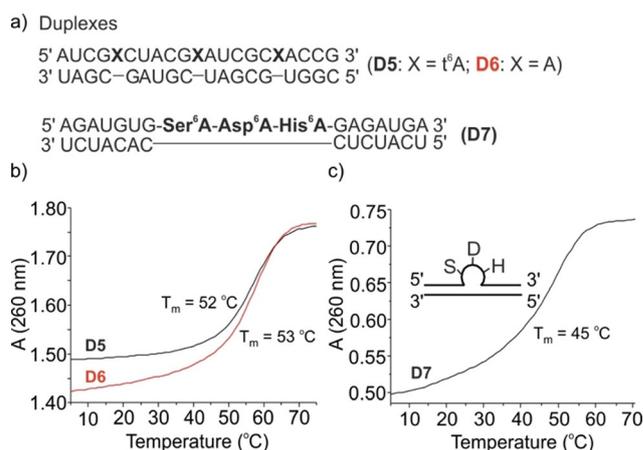
Figure 4. (a) Depiction of the RNA structures containing aa⁶A nucleobases in extrahelical positions forming either three little bulges or assembling a Ser-Asp-His triad known as the catalytic triad in serine proteases; (b, c) depiction of melting curves of duplexes **D5**, **D6**, **D7**; S: serine, D: aspartate, H: histidine.

present in serine peptidases.[35] Again in this case a stable duplex structure forms with the three aa⁶A bases creating a loop. Although we do not show any catalytic activity here, we believe that it is easily imaginable that if these amino acids are properly positioned in a stably folded RNA the structure could gain catalytic properties.

The melting data show, that aa⁶A bases alone are unable to establish base pairing, which hinder them to encode sequence information. On the other side, these bases allow the incorporation of amino acids into RNA structures irrespective of the counterbase. Because RNAs are mostly stably folded structures in which many bases are not involved in any base pairing or establish no Watson–Crick interactions the amino acid adenosine nucleosides allow the stable incorporation of amino acid functionality into RNA.

In summary, here we investigated the synthesis and properties of aa⁶A nucleoside-amino acid conjugates, some of which (t⁶A, g⁶A, hn⁶A) are today found as key components in the tRNAs of many species. In these tRNAs the aa⁶A nucleosides reside at the general purine position 37 adjacent to the anticodon loop. They are not involved in base pairing but fine tune the codon-anticodon interaction to enable faithful translation of information into a peptide sequence.[36] Here we show that these bases are indeed unable to base pair. They have to be placed outside the pairing regime that is needed for RNA folding. As such they function as anchors that allow the connection of amino acid to RNA structures independent of the counterbase. The side chains are then available to equip RNA with additional functions that might have been beneficial in an early RNA-peptide world. The fact that aa⁶A nucleosides are stable structures and until today broadly found in today's RNA make them prime candidates to develop idea about the chemical constitution of the vanished RNA-peptide world.

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

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

5 Unpublished results

Part I

5.1 Chemical studies on the emergence of the proto-ribosome

Since amino acid modified nucleosides (AMNs) can form under prebiotic conditions,[29] we assume that amino acid (aa) modified bases such as aa^{6}A bases were likely present early on during chemical evolution.[197] Our concept involves that these bases, which one could consider to be "living fossils" of a prebiotic RNA world, were once the key units that allowed RNA to control peptide formation. We assume that they established an early stage proto-ribosome. The basic chemical idea is depicted in Figure 18.
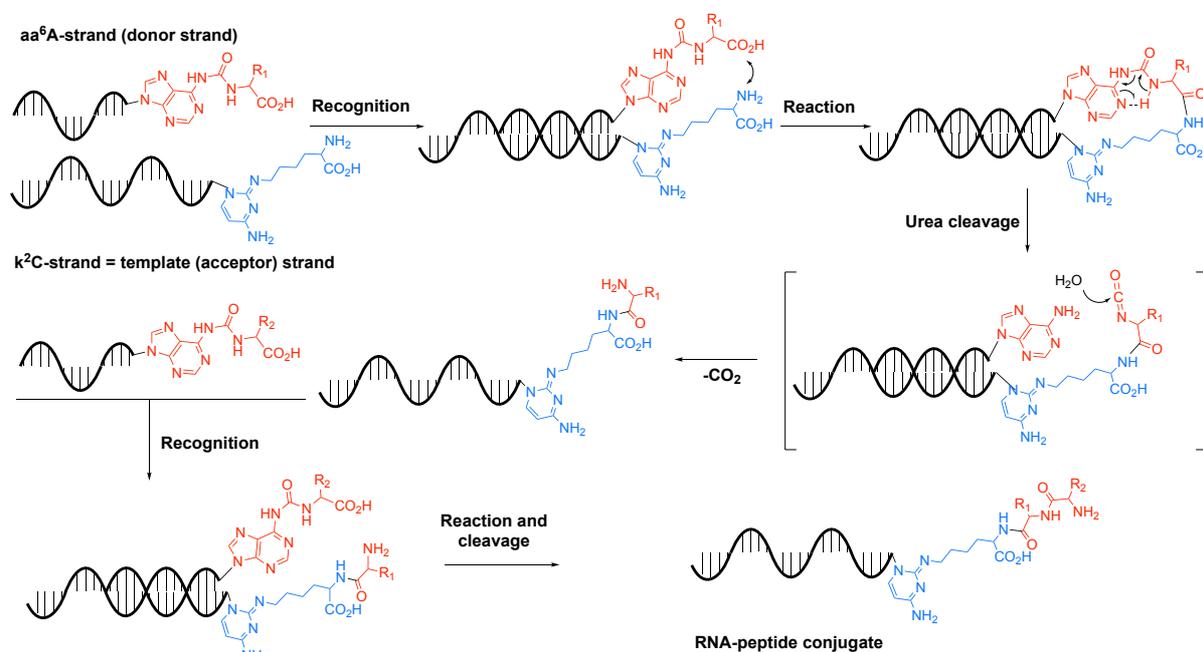
Figure 18. Schematic depiction of a synthetic cycle based on amino acid modified RNA nucleosides that form peptides in an RNA-templated manner.

In the depicted concept, the amino acid modified adenosines (aa^{6}A), found today in anticodon position 37, contain a carboxyl group. It could serve as an amino acid donor (shown in red, Figure 18). While the amino group containing RNA modifications located today in anticodon position 34 (e.g. k^{2}C, shown in blue, Figure 18) could serve as amino acid acceptors. The peptide bond could be formed within two complementary strands between an activated carboxyl and the amino groups in an RNA-templated manner to give first a hairpin intermediate. Afterwards hydrolysis of the urea linkage, which should be labile under thermal

62

conditions, could take place. During this reaction a reactive isocyanate is formed, which is captured by water. This could then be followed by a subsequent decarboxylation, which leads to the formation of a (n+1)-peptide. Release of the first aa^{6}A containing donor strand and binding of a new one could give the (n+2)-peptide and so forth as represented in Figure 18. One can imagine that at some point such RNA-peptide conjugates gain catalytic activity.

Several types of naturally occurring modified bases (Figure 19) can be used for setting up the primitive translational machinery. Theoretically, amino acid modified bases such as t^{6}A, m^{6}t^{6}A, g^{6}A and k^{2}C are excellent candidates for an RNA-templated peptide bond formation process as depicted in Figure 18. Initially it was decided to investigate lysidine (k^{2}C) for the template (acceptor) strand and t^{6}A for the donor strand.
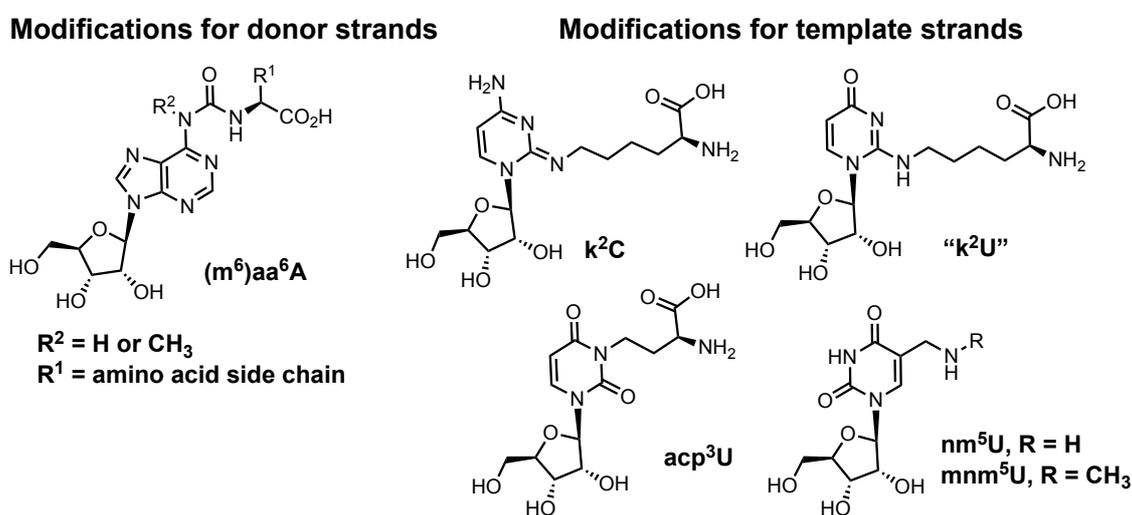
Figure 19. Modifications used for the donor and acceptor (template) strands.

5.1.1 k^{2}C as a candidate for an RNA-templated peptide formation

To prove that amino acid modified nucleosides like t^{6}A and k^{2}C could in principle participate in an RNA-templated peptide formation, phosphoramidite building blocks of t^{6}A and k^{2}C (Figure 20) were synthesized. The synthesis of **t^{6}A-PA** was performed according to the procedures presented in chapter 4.3. The phosphoramidite was incorporated into the 7-mer RNA strand 5'-**t^{6}AAUCGCU**-3' (**ON1**), which is a part of tRNA^{Ile} anticodon stem loop (ASL) containing both k^{2}C and t^{6}A at 34th and 37th positions, respectively.
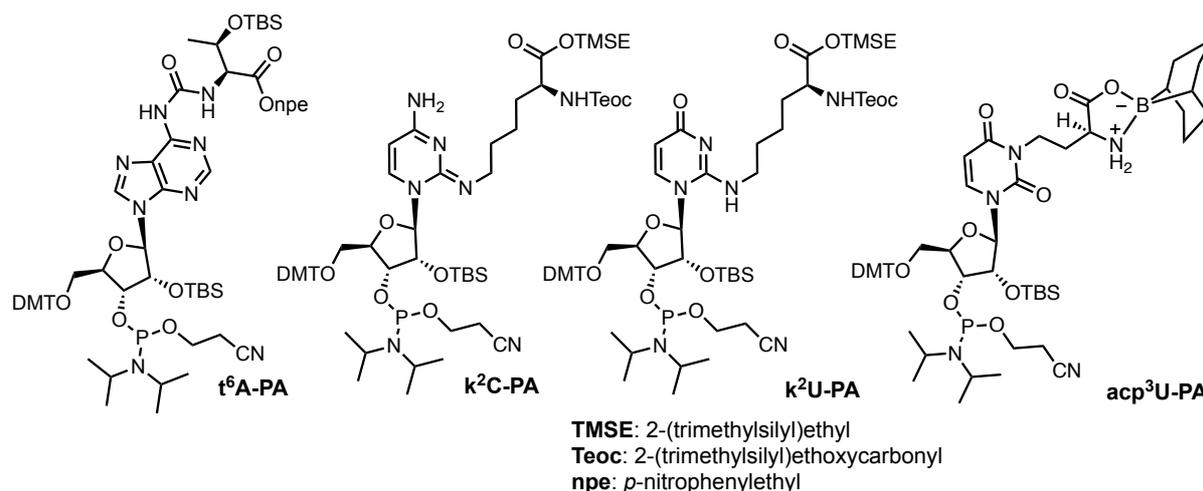
Figure 20. Phosphoramidite building blocks of t⁶A, k²C, k²U and acp³U.

TMSE: 2-(trimethylsilyl)ethyl
Teoc: 2-(trimethylsilyl)ethoxycarbonyl
npe: p-nitrophenylethyl

The synthesis of **k²C-PA** was rather challenging. Initially the acetonide protected compound **1** was synthesized according to published procedures.[198,199] Although the coupling of compound **1** with the lysine building block **A** (the synthesis was presented in chapter 4.2) in the presence of LiCl and DBU furnished the cytidine derivative **2**, various attempts to deprotect the acetonide protecting group provided the target product **3** only in a very low yield (Scheme 2a). We then decided to use the phosphoranediylcyclocytidine derivative **4**[198] for the coupling with the lysine building block **A** (Scheme 2b). The coupling reaction took place with liberation of triphenylphosphine oxide, forming cytidine derivative **5**. Then the protection of the 5'-OH group with DMTCl was performed. However, only small amounts of the target product **6** were isolated. Also, the NMR data looked very complex, potentially due to possible tautomerism.
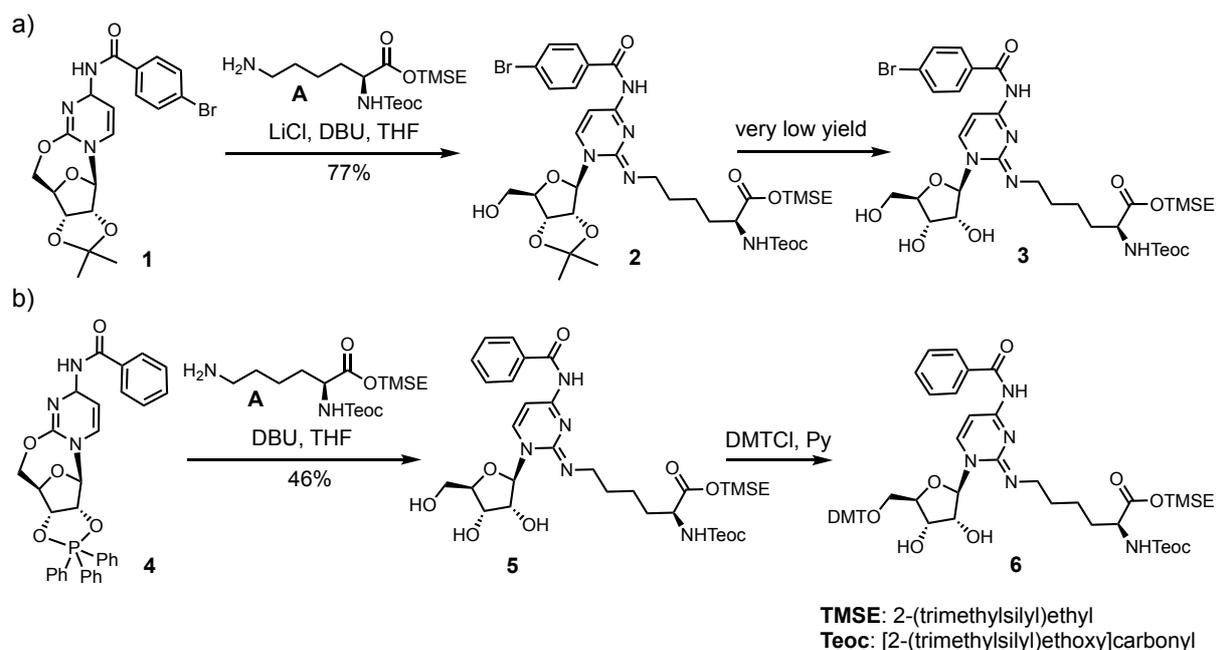
Scheme 2. Attempts to synthesize **k²C-PA**.

Because of these synthetic difficulties, the analog, k²U (Figure 20) was prepared as discussed in chapter 4.2 and incorporated into the RNA oligonucleotide (5'-AGCGAU**k²U**C-3', **ON2**) in order to investigate the peptide bond formation. We also attempted to use k²U derivative **7** (the synthesis is shown in chapter 4.2) to prepare a convertible nucleoside **9**, which after incorporation into RNA strand could be transformed into k²C containing oligonucleotide as shown in Scheme 3. BOP ((benzotriazole-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate) was used in order to activate the nucleobase. The formed BOP derivative was directly used in the next step without prior purification. The BOP moiety was displaced with 4-chlorophenol to yield compound **8**, however the yield reached only 15%. Attempts to increase the yield were not successful. The synthesis of **k²C-PA** was consequently not investigated further.

Scheme 3. Potential synthesis of convertible nucleoside **9**.

Next, the peptide growth (Figure 21) was performed using EDC hydrochloride (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) or DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride) as an activator at pH 5 and pH 7, respectively. However, the EDC reagent was quickly substituted by DMTMM because it led to the formation of numerous side products resulting in complicated reaction mixtures, which were observed by MALDI-TOF. Carbodiimides are reactive towards phosphate groups and, additionally, they are problematic due to generation of unreactive *N*-acyl urea derivatives, when the peptide formation kinetics is unfavored. On the contrary, the DMTMM reagent reacted chemoselectively with the carboxylic acid, which was investigated by keeping the acceptor and donor strand as single strands in a MOPS buffer (pH 7) together with DMTMM. The samples were analyzed by HPLC after 24 h and showed no change in composition.

Since the melting point of the 7-mer double strand is expected to be below room temperature, the peptide **P1** bond formation between **ON1** and **ON2** (Figure 21) was initially performed at 4 °C. At this temperature, the reaction was rather slow as monitored by MALDI-TOF. After 72 h starting material still remained. Side reactions, however, were not observed.

Figure 21. Peptide **P1** formation between k^{2}U and t^{6}A containing strands **ON2** and **ON1**.

These experiments suspected that the concept of an RNA-templated peptide bond formation between two amino acid modified bases is valid. However, the synthesis of **k²U-PA** and of the RNA strands containing it was rather complicated. k^{2}U is also not present in biological systems. The synthesis of **k²C-PA** was even more complex and inefficient. Thus, different acceptor nucleotides were selected for further investigations.

5.1.2 Mechanistic considerations of urea cleavage

To successfully achieve the formation of peptide chains in an RNA-templated manner, the urea cleavage conditions were next investigated in detail. Potential mechanistic considerations about it are shown in the figure below. Figure 22a depicts a potential thermal and Figure 22b a metal-mediated urea cleavage. Catalytic metal-mediated urea bond cleavage was also discussed in the literature.[200] It is important to mention that our thermal urea cleavage experiments (Figure 22a) were performed in three different devices: thermomixer, incubator and in the PCR machine (thermo cycler). The results greatly differed depending on the device that was used. The best results were achieved by performing the reaction in the PCR machine because the whole reaction vial including the lid can be heated, which ensures homogenous heating of the whole reaction mixture. Urea cleavage in other devices lead to degradation of RNA. During the reaction, the liquid phase condensed on the lid and the RNA then was kept in the dry state which could explain why RNA underwent degradation in other devices than the PCR machine.
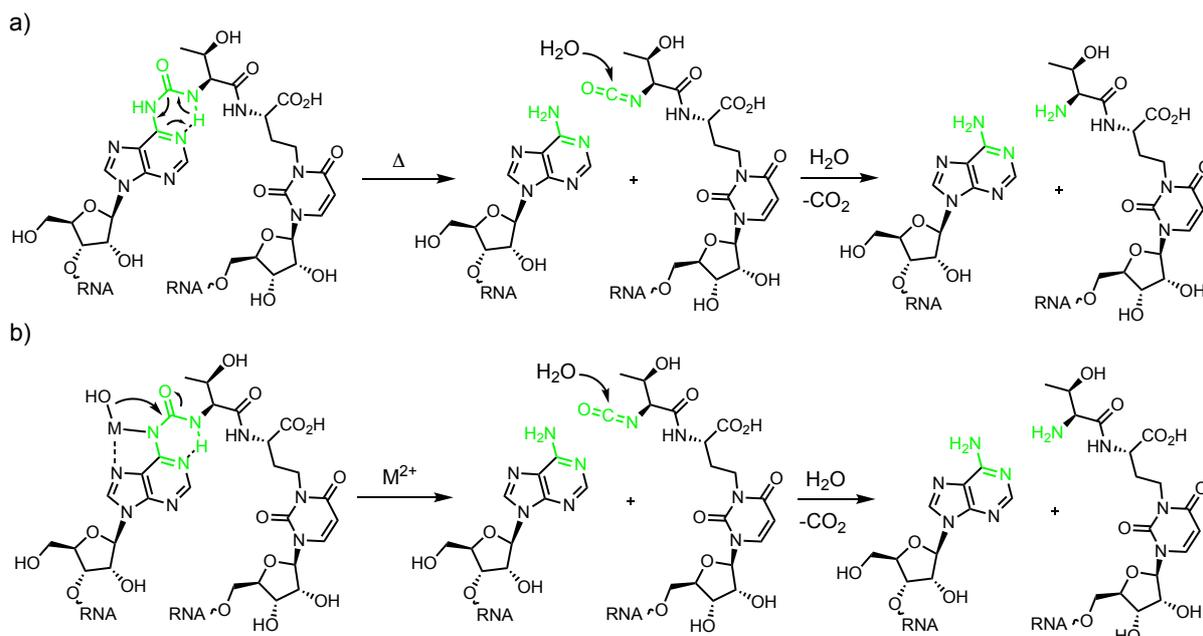
Figure 22. (a) Thermal hydrolysis of urea; (b) metal-assisted hydrolysis of urea.

Although the susceptibility of a phosphate bond to metal-ion attack causing its cleavage was heavily discussed in literature,[200-203] it was also reported that the nature of the nucleobase influences the rate of the phosphate bond cleavage[200]. Thus, metal-assisted urea hydrolysis (Figure 22b) was studied. However, the obtained data were not very promising. Cleavage in the presence of metal ions lead to increased degradation of the RNA backbone, providing just a small amount of target product. Performing the reaction with MCl_2 (M = Zn, Ni, Mg, Co) species or metal (Co, Ni) amine complexes at room temperature left the RNA strands intact, while heating of RNA oligonucleotides (60-95 °C) with metal ions lead to degradation of the RNA oligonucleotides. Thus, only traces of target products were detected.

Booker-Milburn and Lloyd-Jones reported that hindered urea derivatives can act as masked isocyanates due to their facile cleavage/solvolysis.[204] Remarkably, they reported that hindered trisubstituted ureas can easily decompose via isocyanate intermediates under neutral conditions. In some cases, the reaction reaches completion in less than an hour at 20 °C. m^6t^6A is another modified base in contemporary tRNA, whose methyl group would make the urea moiety significantly more hindered and vulnerable according to Booker-Milburn and Lloyd-Jones. Thus m^6aa^6A (e.g. m^6t^6A) derivatives could undergo faster thermal urea cleavage than the corresponding aa^6A analogues. Thermal cleavage would furnish an isocyanate attached to the amino acid, which is hydrolyzed and decarboxylated under aqueous conditions as shown in Figure 23.
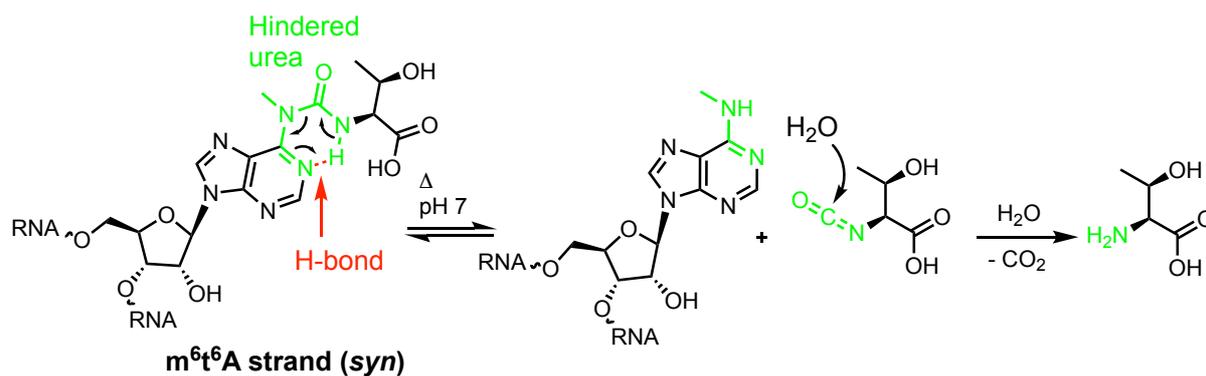
Figure 23. Thermal urea cleavage of an m^6t^6A containing RNA strand.

To test this hypothesis, the phosphoramidite building block of the methylated aa^6A derivative (m^6aa^6A) had to be synthesized. In order to expand the diversity of amino acids that could be attached during the proposed RNA-templated peptide formation, the phosphoramidites of Thr, Gly, Val and Leu were synthesized as shown in Scheme 4. First, the amino acid components for the coupling to the A-nucleoside were prepared. The synthesis of the protected threonine, glycine and valine derivatives was performed according to the procedures presented in chapter 4.3. The carboxyl group of leucine **10** was transformed into the *p*-nitrophenylethyl ester (npe, **11**).[205] Acidic Boc-deprotection using 4 M HCl in dioxane furnished the ready-to-couple amino acid **12**.[206] Next, the connection of the amino acid with the A-nucleoside via the urea moiety was carried out. Phenyl chloroformate was treated with *N*-methylimidazole to obtain the 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**13**).[207] Adenosine **14** was converted in parallel into the cyclic 3',5'-silyl protected nucleoside, followed by conversion of the 2'-OH group into the TBS-ether **15**.[208] The reaction of compound **15** with the activated carbonate **13** and of the corresponding amino acid, yielded in all cases the amino acid coupled products **16-19**, which were subsequently methylated on N^6-position using MeI under basic conditions. Compounds **20-23** were synthesized in good or excellent yields. Then the cleavage of the cyclic silylether was performed using HF-pyridine.[209,210] Subsequent protection of the 5'-OH with dimethoxytritylchloride (DMTCl)[211] provided the compounds **28-31**, which were converted into corresponding phosphoramidites **32-35**. The phosphoramidites were pre-dried over molecular sieves (4 Å) and incorporated into oligonucleotides (**ON3-ON6**) using standard solid phase RNA chemistry.[212-219]
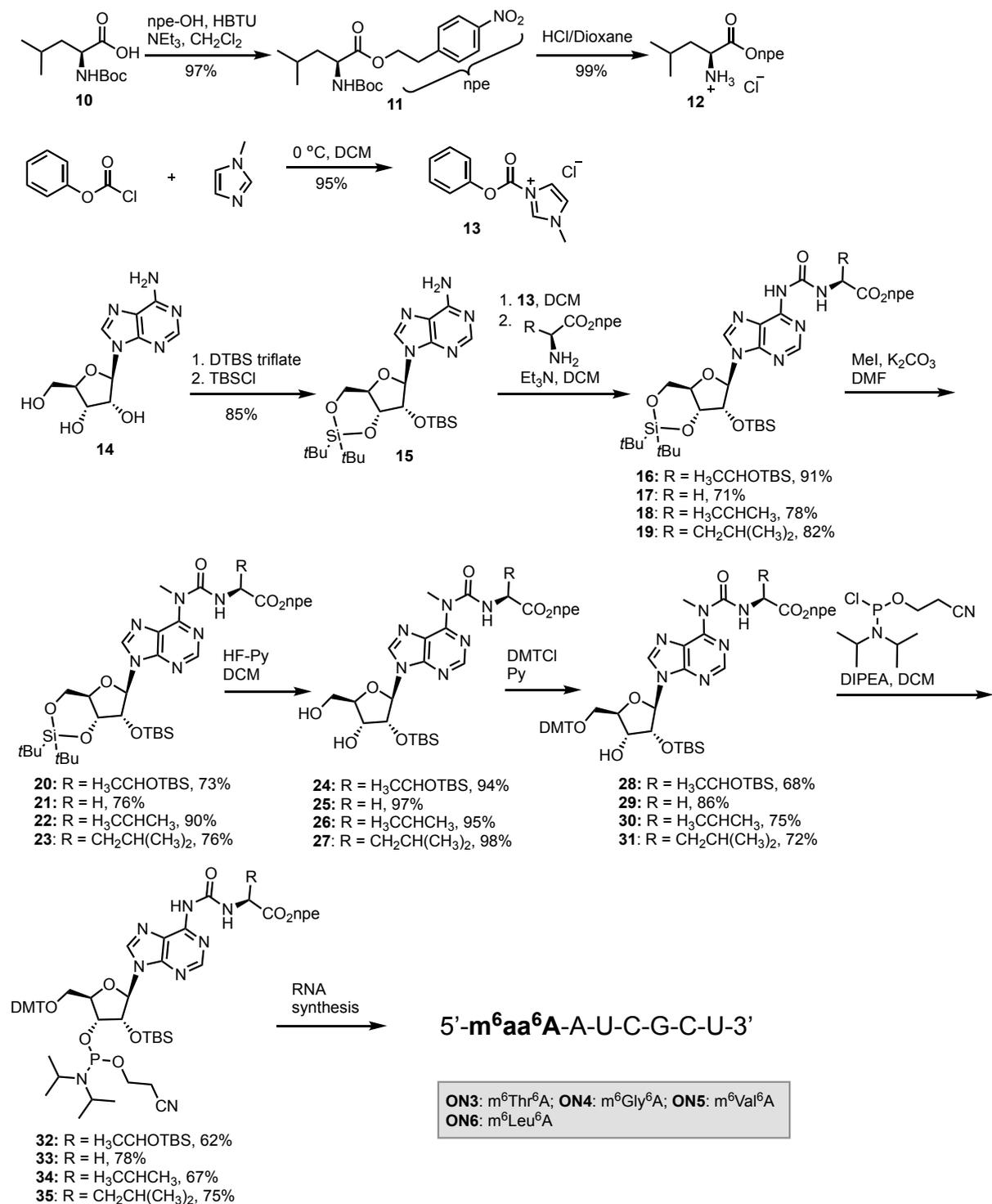
Scheme 4. General synthesis of phosphoramidite building blocks of m⁶aa⁶A and their incorporation into RNA oligonucleotides.

In order to evaluate differences regarding the thermal stability of methylated vs. unmethylated urea fragments, two RNA oligonucleotides (**ON7** and **ON8**, Figure 24) were heated at 95 °C in water for 24 h. To ensure homogenous heating, the reaction was performed in a PCR machine

with heated lid. The data (Figure 24) shows the difference regarding the cleavage of the methylated strand **ON7** versus the unmethylated **ON8**, which is consistent with the data published by Booker-Milburn and Lloyd-Jones about highly substituted ureas. After 24 h full conversion of **ON7** to give the m^{6}A strand was observed (Figure 24a,b; the sample was also coinjected with an independently synthesized 5'-GUC**m^{6}A**CCUGA-3' RNA strand). In contrast, the urea cleavage of the unmethylated strand **ON8** was significantly slower (Figure 24c,d). Thus, the strands with a methyl group at the N^6-atom (m^{6}aa^{6}A strands) were used for further investigations that are presented in the upcoming chapters.
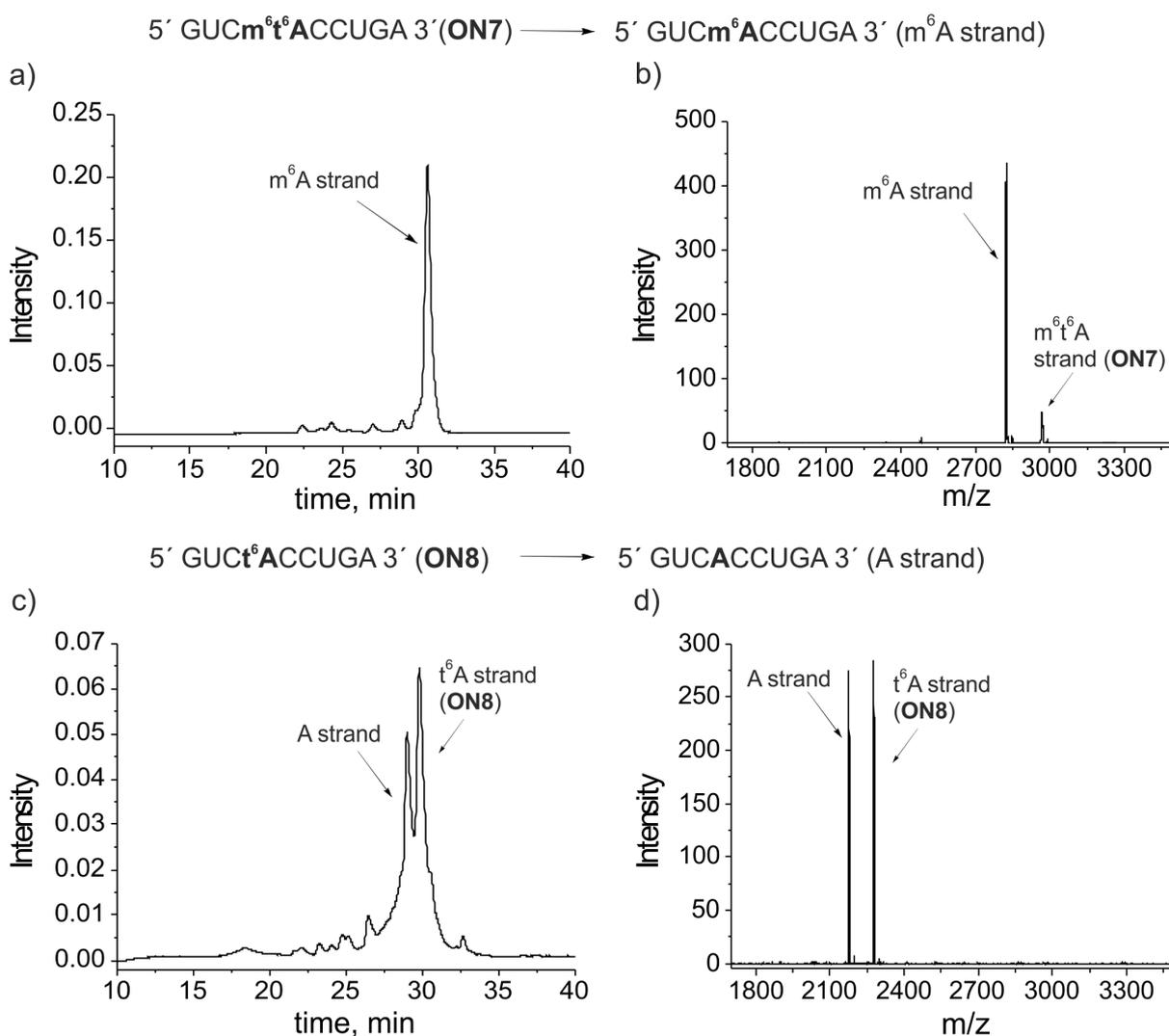
Figure 24. (a) HPL-chromatogram of the cleavage of **ON7** strand; (b) MALDI-TOF mass spectrum of the cleavage of **ON7** strand; (c) HPL-chromatogram of the cleavage of **ON8** strand; (d) MALDI-TOF mass spectrum of the cleavage of **ON8** strand.

5.1.3 acp³U as a primitive acceptor of amino acids

acp³U is another amino acid modified base that was chosen for the investigation of whether it can function as a primitive acceptor of amino acids. The phosphoramidite building block of acp³U (Figure 20) was synthesized according to procedures described in chapter 4.1 and incorporated into RNA oligonucleotide (5'-AGCGAU**acp³U**U-3', **ON9**). The peptide **P2** formation (Figure 25) between the acp³U (**ON9**) and m⁶t⁶A (**ON3**) modified strands was performed using DMTMM as an activator. The reaction was carried out at 4 ℃ for 48 h.
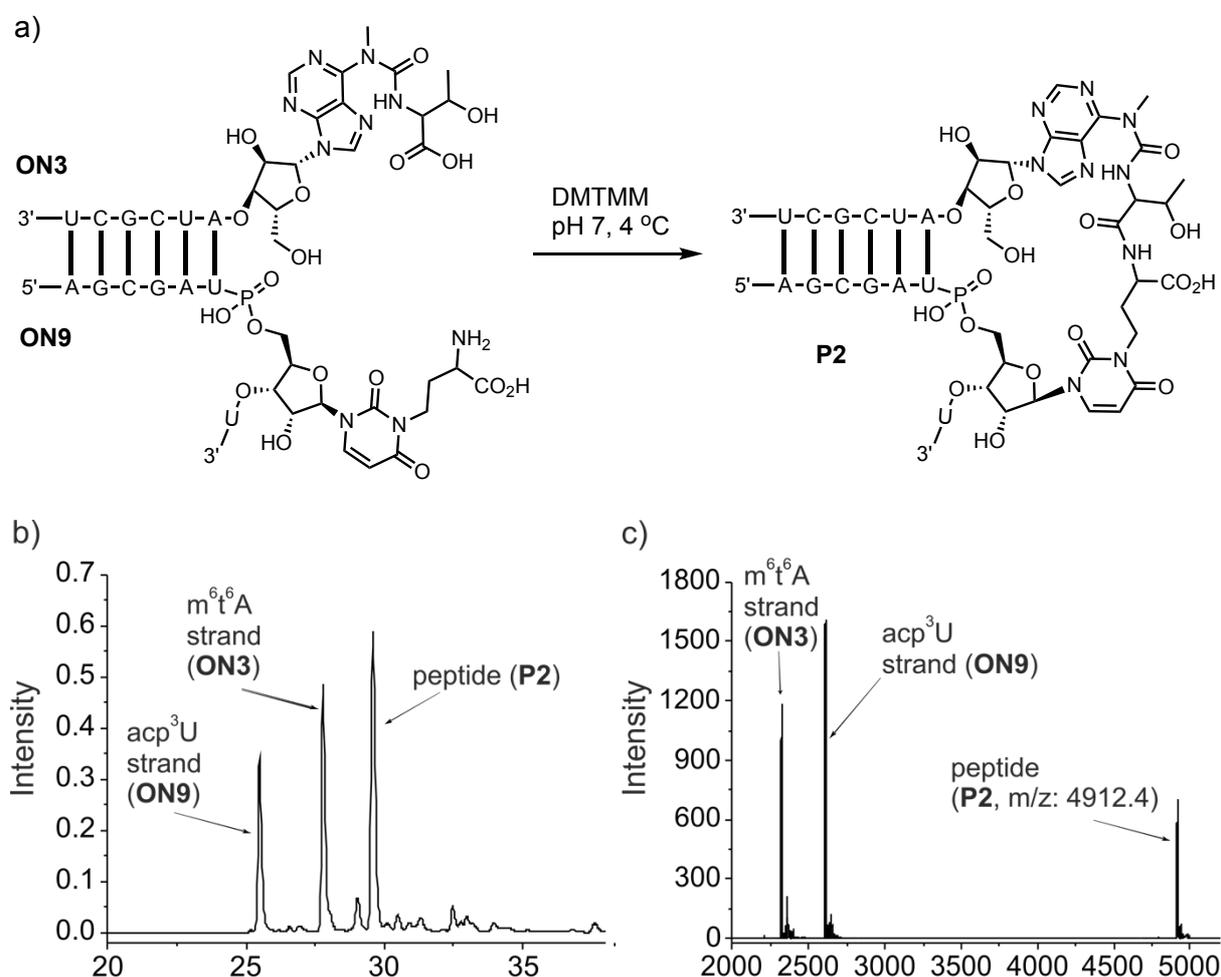
Figure 25. (a) Schematic depiction of peptide **P2** formation between **ON3** and **ON9**; (b), (c) HPL-chromatogram and MALDI-TOF mass spectrum of peptide **P2** formation between **ON3** and **ON9**.

Figure 25 shows the data obtained for the peptide **P2** formation between acp^{3}U and m^{6}t^{6}A modified RNA oligonucleotides, proving a clean and successful reaction. It is worth to mention that when keeping the reaction mixture for longer times or at higher temperatures, adducts of DMTMM start to form.

Next, the target peptide **P2** was purified by HPLC and subsequently its thermal urea cleavage was studied. The peptide was heated at 95 ºC (pH 7) and samples taken from the reaction mixture were analyzed by HPLC and MALDI-TOF after 4, 12 and 24 h. Although the peptide **P2** underwent urea cleavage, this did not lead to any observable amount of target product (path a, Figure 26). Instead, the isocyanate unexpectedly cyclized in an intramolecular fashion to give cyclic carbamate (green, path b, Figure 26).
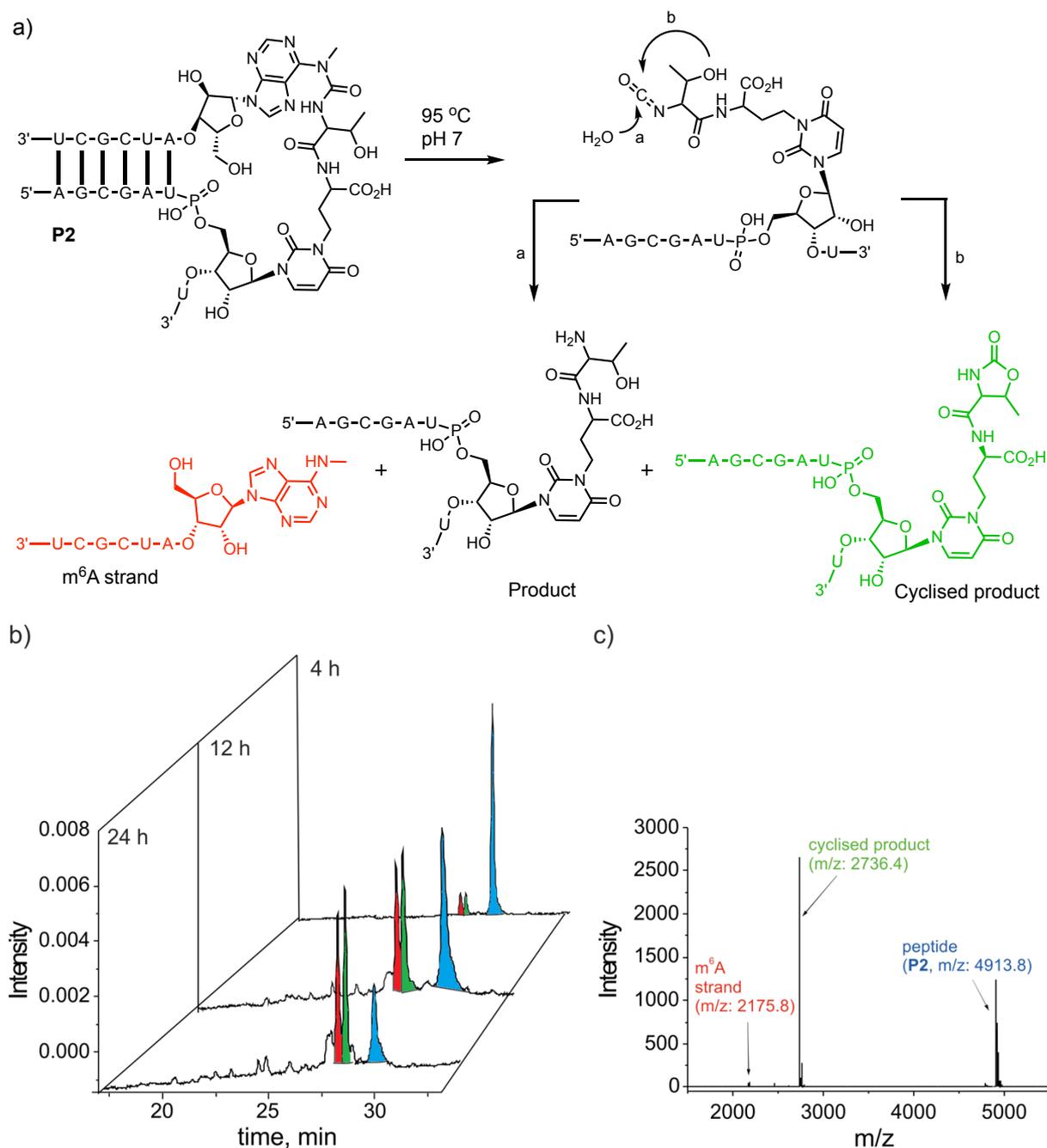
Figure 26. (a) Schematic depiction of the urea cleavage of **P2**; (b) HPL-chromatogram of the urea cleavage reaction, blue: starting material **P2**, red: m⁶A strand, green: cyclized product; (c) representative MALDI-TOF mass spectrum of the urea cleavage reaction.

The cyclisation is an unexpected reaction of the hydroxy group. Thus, the experiment was repeated with amino acids lacking any functionality in the side chains. We prepared a strand containing Val (**ON5**, m⁶v⁶A) instead of Thr (**ON3**, m⁶t⁶A) for further investigation. The data, summarized in Figure 27, show that the formation of the peptide **P3** was again successful as expected.
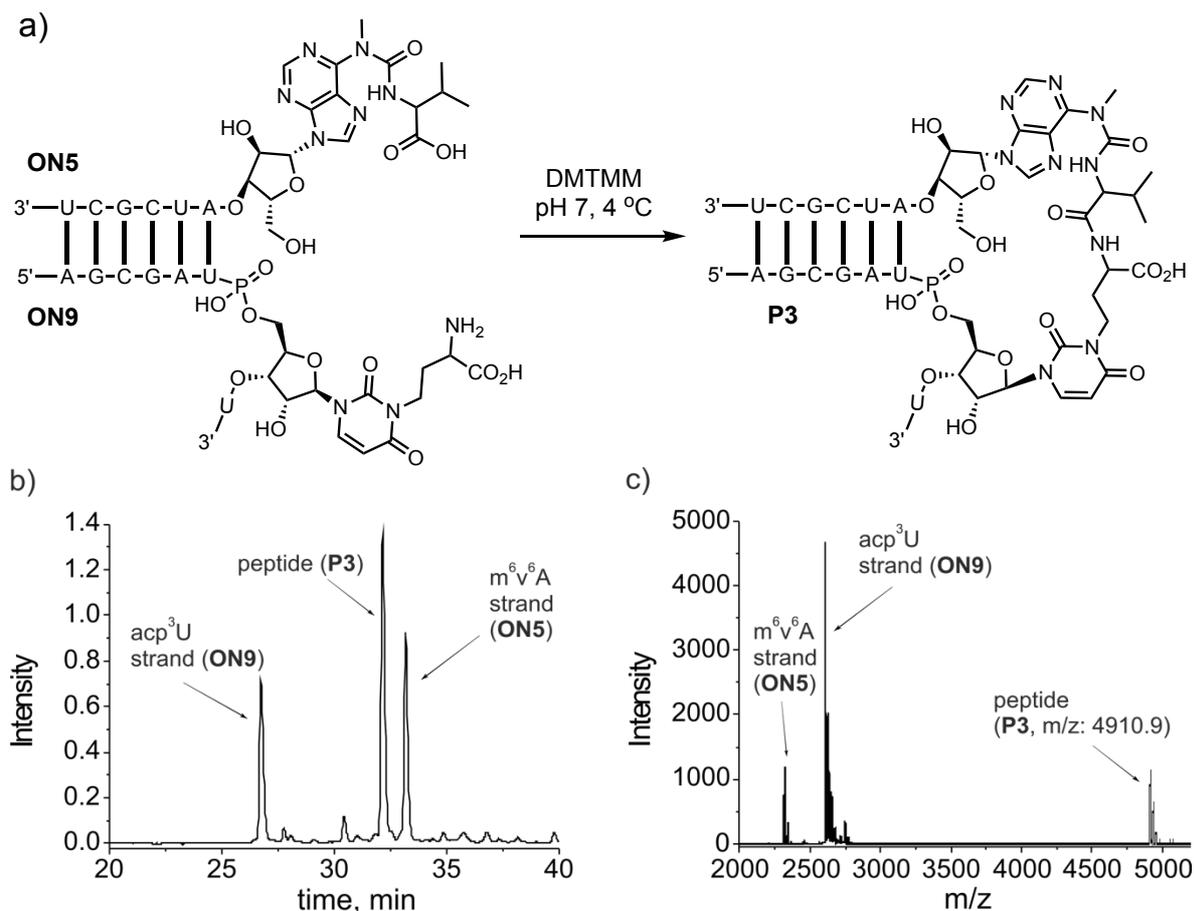
Figure 27. a) Schematic depiction of the peptide **P3** formation between **ON5** and **ON9**; (b), (c) HPLC and MALDI-TOF data of peptide **P3** growth reaction.

The cyclic peptide **P3** was again purified by HPLC and then subjected to urea cleavage as depicted in Figure 28. Now the dipeptide could indeed be detected (path a, Figure 28). However, we now also observed an intramolecular attack of the amide bond to give a cyclized product that unfortunately dominated (path b, Figure 28). Moreover, the HPLC analysis of the cleavage reaction was complex. Proper separation of the compounds could not be obtained (Figure 28c).
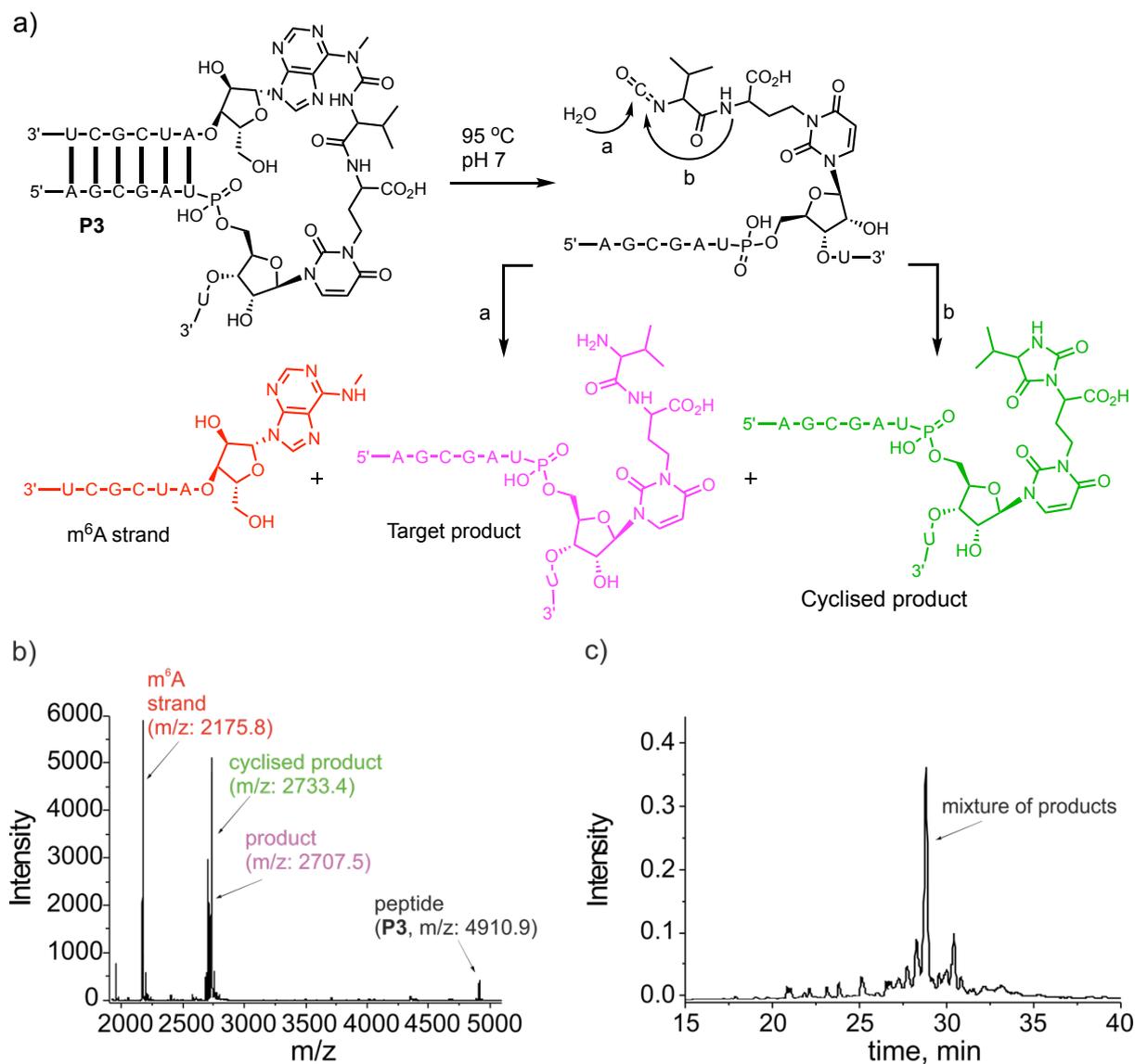
Figure 28. (a) Schematic depiction of the peptide **P3** urea cleavage; (b), (c) HPLC and MALDI-TOF data of urea cleavage experiment.

5.1.4 5-Aminomethyl modified uridines as candidates for an RNA-templated peptide formation

5.1.4.1 mnm^{5}U as a primitive acceptor tRNA

In order to also prevent the cyclisation reaction via an amide bond we investigated next aminomethyl modified uridines (nm^{5}U and mnm^{5}U, Figure 20) as acceptors as shown in Scheme 5.
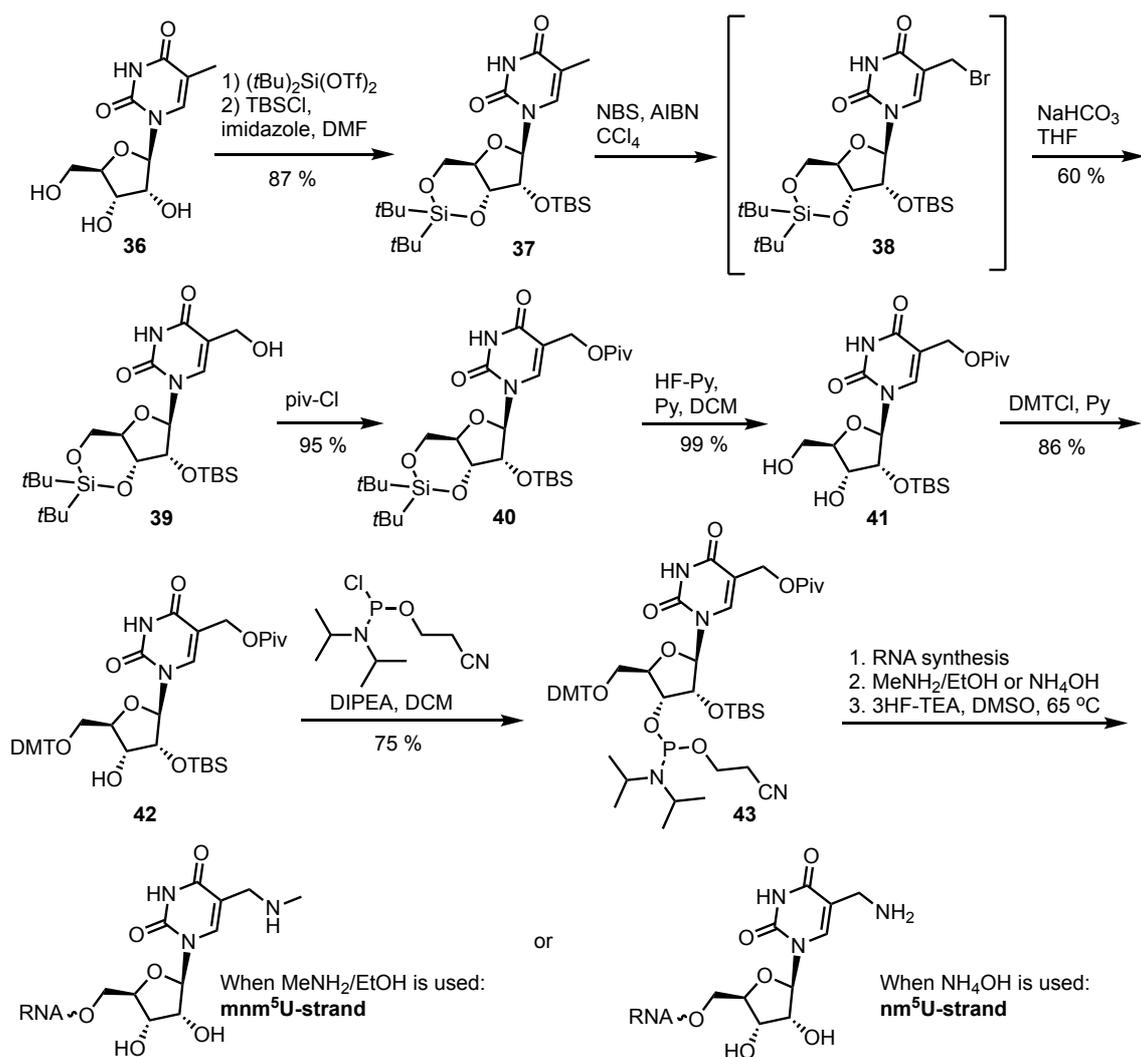
Scheme 5. Synthesis scheme of pivaloyloxymethyluridine (pivom⁵U) phosphoramidite building block **43** and its use for the synthesis of nm⁵U or mnm⁵U containing RNA oligonucleotides.

The synthesis started from methyluridine **36** with the protection of the 5'-OH and the 3'-OH followed by derivatization of the 2'-OH to the silyl ether **37**.[230] Then, a Wohl-Ziegler reaction was used to form the brominated compound **38**, which, due to its instability during column chromatography, was immediately converted into the 5'-hydroxymethyluridine derivative **39**.[220] The protection of a hydroxy group at the pseudobenzylic position in compound **39** was achieved by addition of piv-Cl in pyridine. The resultant nucleoside **40** was treated with HF·Py to remove the 5',3'-silyl protecting group. The subsequent protection of the primary 5'-OH group with dimethoxytritylchloride (DMTCl) allowed conversion of compound **42** into the corresponding pivom⁵U phosphoramidite **43**.[231] Standard solid phase chemistry was used to

incorporate the pivom^{5}U at the 3' terminus with *Universal Support III* resulting in *ca*. 96% coupling yield (as quantified by the DMT cation assay).
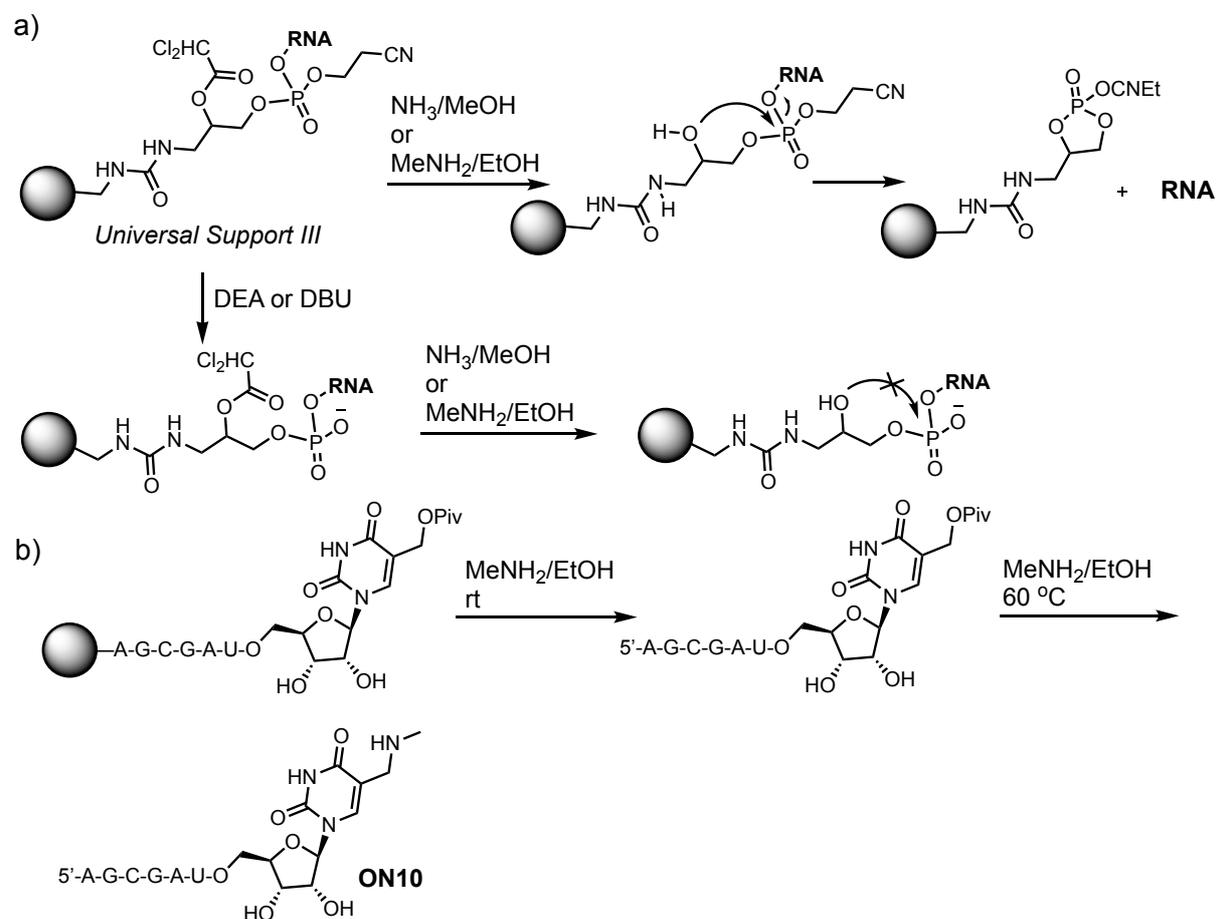
Scheme 6. (a) Amide assisted elimination of *Universal Support III*; (b) cleavage of pivom^{5}U containing strand from the solid support and subsequent conversion into the mnm^{5}U-modified strand **ON10**.

According to the manufacturer, the cleavage from such a resin is achieved using mild basic conditions (Scheme 6a). Also, it is important to mention that *Universal Support III* is incompatible with the potential preliminary elimination of the cyanoethyl deprotecting groups using hindered bases such as diethylamine (DEA) or 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU). This would cause the majority of the RNA to remain bound on the solid support as shown in Scheme 6a.

Thus, the pivom^{5}U modified oligonucleotide has to be first cleaved from the solid support by treating it with MeNH$_2$/EtOH at room temperature for 2 h (Scheme 6b). Then, in order to post-synthetically convert the pivom^{5}U containing oligonucleotide into a mnm^{5}U-modified strand

ON10, the strand needs to be treated with MeNH$_2$/EtOH at 60 °C for 1 h, as reported by Leszczynska.[221] Figure 29 shows the data obtained for the mnm^{5}U containing 7-mer RNA oligonucleotide (5'-AGCGAU**mnm^{5}U**-3', **ON10**) proving successful incorporation and nucleophilic substitution by MeNH$_2$/EtOH.

a)   b)
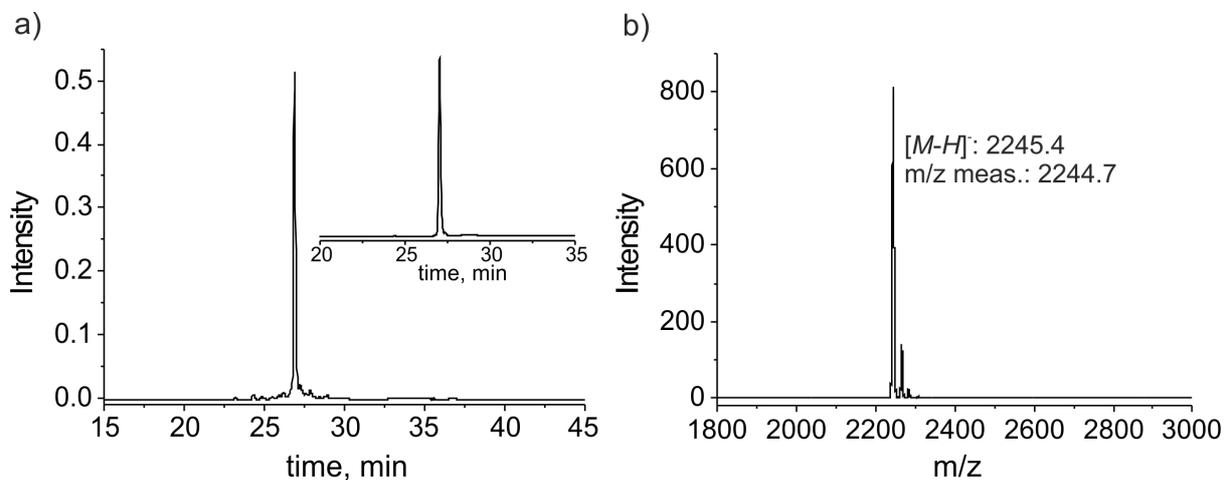
Figure 29. (a) HPL-chromatogram of unpurified **ON10** with the inset showing the HPL-chromatogram of purified **ON10**; (b) MALDI-TOF mass spectrum of **ON10** after purification.

The formation of peptide **P4** between the mnm^{5}U-modified strand **ON10** and the m^{6}v^{6}A strand **ON5** was again performed using DMTMM as an activator (Scheme 7). The reaction was kept for 48 h at 4 °C and monitored by MALDI-TOF and HPLC.
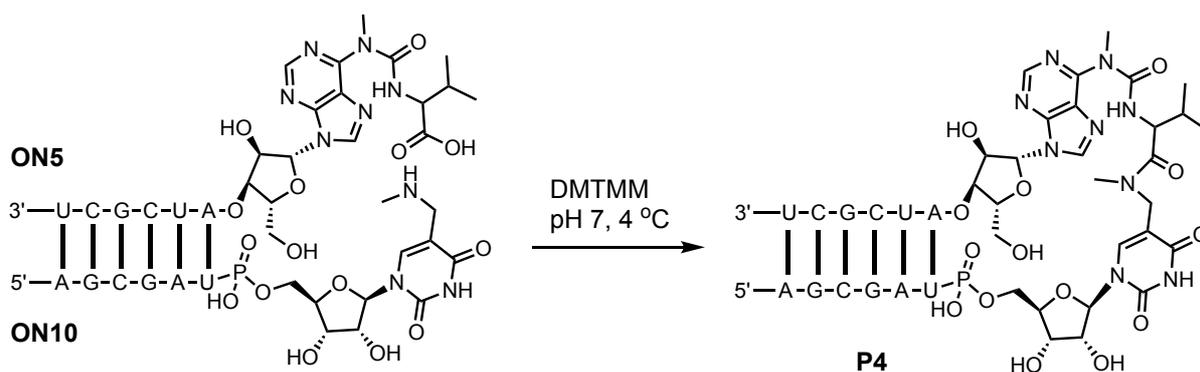
Scheme 7. Schematic depiction of peptide **P4** formation.

During the measurements it was noticed that rapid acetylation of the template strand **ON10** (Figure 30a-c) also occurs, which unfortunately limited peptide **P4** formation.
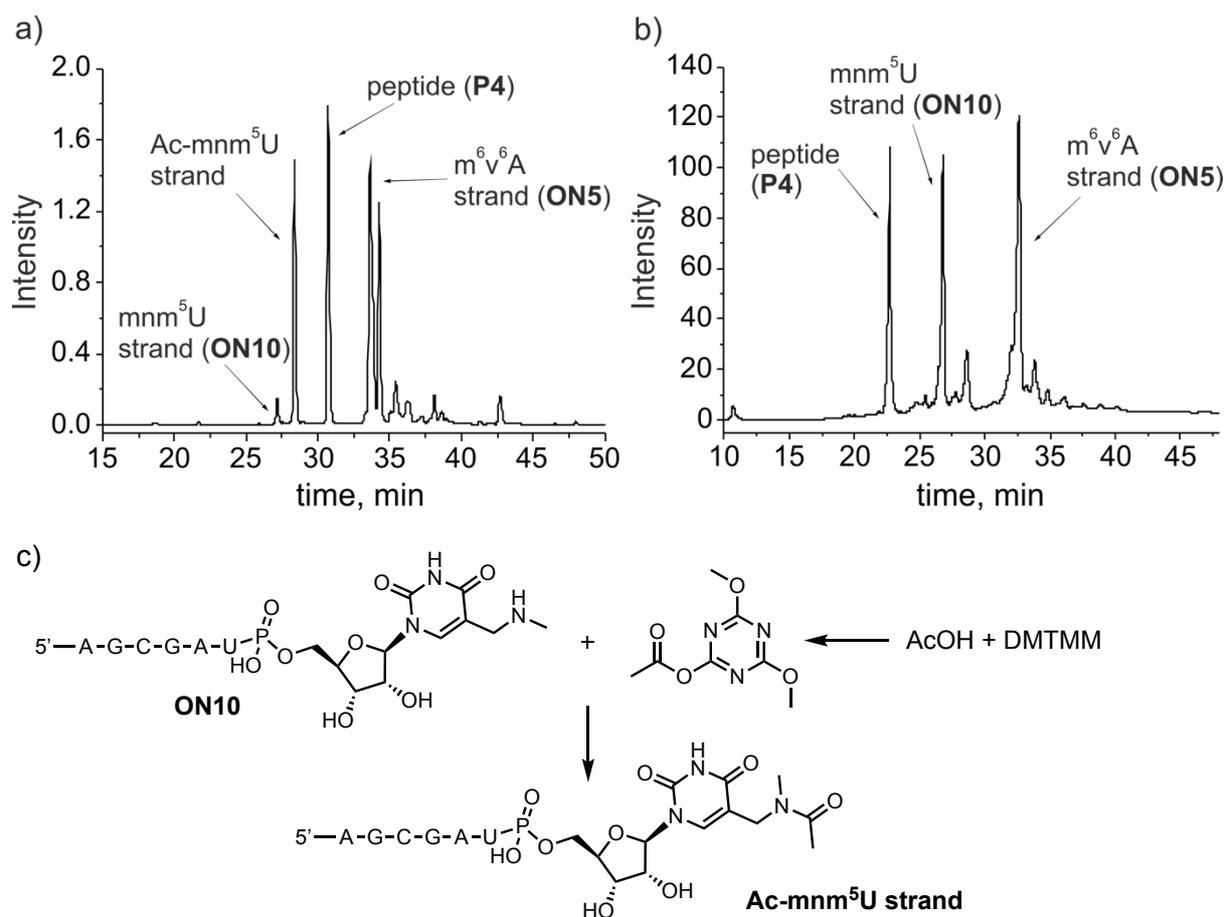
Figure 30. Representative HPL-chromatograms of peptide **P4** formation between **ON5** and **ON10** eluting with (a) TEA/AcOH buffer system; (b) eluting with phosphate buffer system; (c) schematic depiction of **ON10** acetylation.

There can be two sources of acetate ions: remaining acetate ions from the HPLC purification of the single strands **ON10** and **ON5**, and acetate from the buffer system, which is used during the injection of the sample into the HPLC system. The acetate gets obviously activated by DMTMM and such adduct reacts with the secondary amine of the **ON10** (Figure 30c). In order to suppress such reaction, a desalting procedure with *Sep-Pak® Plus Short C18* columns was applied. Additionally, TEA/AcOH buffers were replaced by a phosphate buffer system. Due to the limited solubility of sodium phosphate and its rapid precipitation in organic solvents, potassium phosphate was chosen for the buffer preparation. Desalting and the change of buffers indeed inhibited the formation of acetylated mnm^{5}U strand as shown in Figure 30b.

Next, the purified peptide **P4** was used for urea cleavage experiments (Figure 31). Since after the peptide growth a tertiary amide is formed and the side chain of valine contains not a nucleophilic group, cyclisation during the urea cleavage cannot take place.
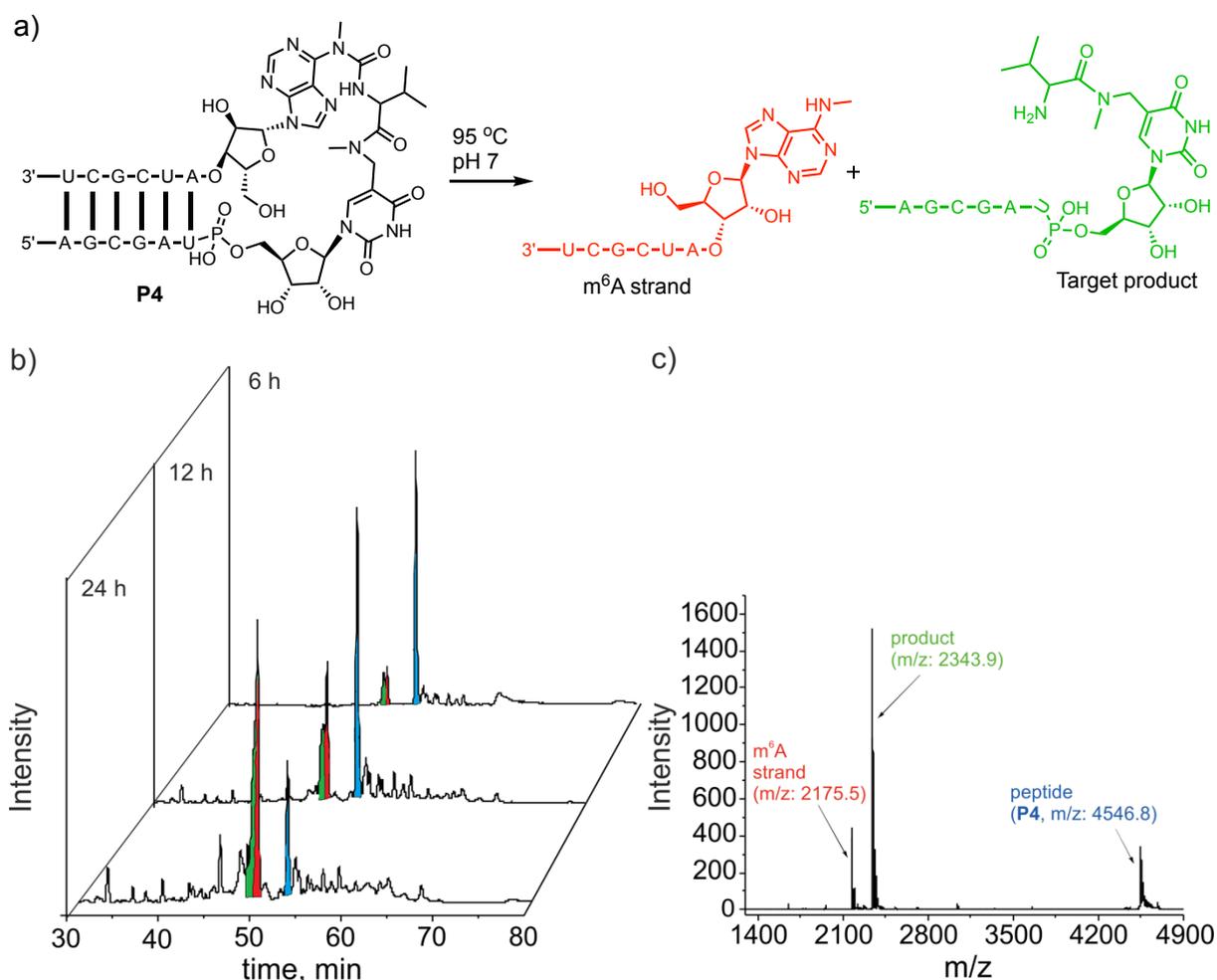
Figure 31. (a) Schematic depiction of urea cleavage of **P4**; (b), (c) HPLC and MALDI-TOF data of the urea cleavage reaction, respectively.

The cleavage reaction was performed at 95 °C in sodium phosphate buffer (pH 7) and monitored over time by HPLC and MALDI-TOF. Figure 31b shows that after 6 h, the target product (shown in green) and the m6A strand (shown in red) have already formed. Over time, the reaction conversion is increasing together with the formation of degradation products. After the separation of some degradation products and analysis with MALDI-TOF it seemed that the m6A strand degrades less than the target product. This observation indicates that the degradation pattern could be sequence-dependent. Unfortunately, the peaks of the product and of the m6A strand were overlapping in HPL chromatograms, irrespective of the gradient or buffer system that was applied. The change of C18 to C8 column for separation helped,

nevertheless, the strands had to be purified many times. Thus, it was decided to make a longer template strand in comparison to the donor strand. The differences in the length of strands should allow the separation by HPLC or gel electrophoresis. A longer strand containing the mnm^{5}U modification (**ON11**) at the 3' terminus was synthesized and used for peptide **P5** formation with the m^{6}v^{6}A modified strand **ON5** (Figure 32a).
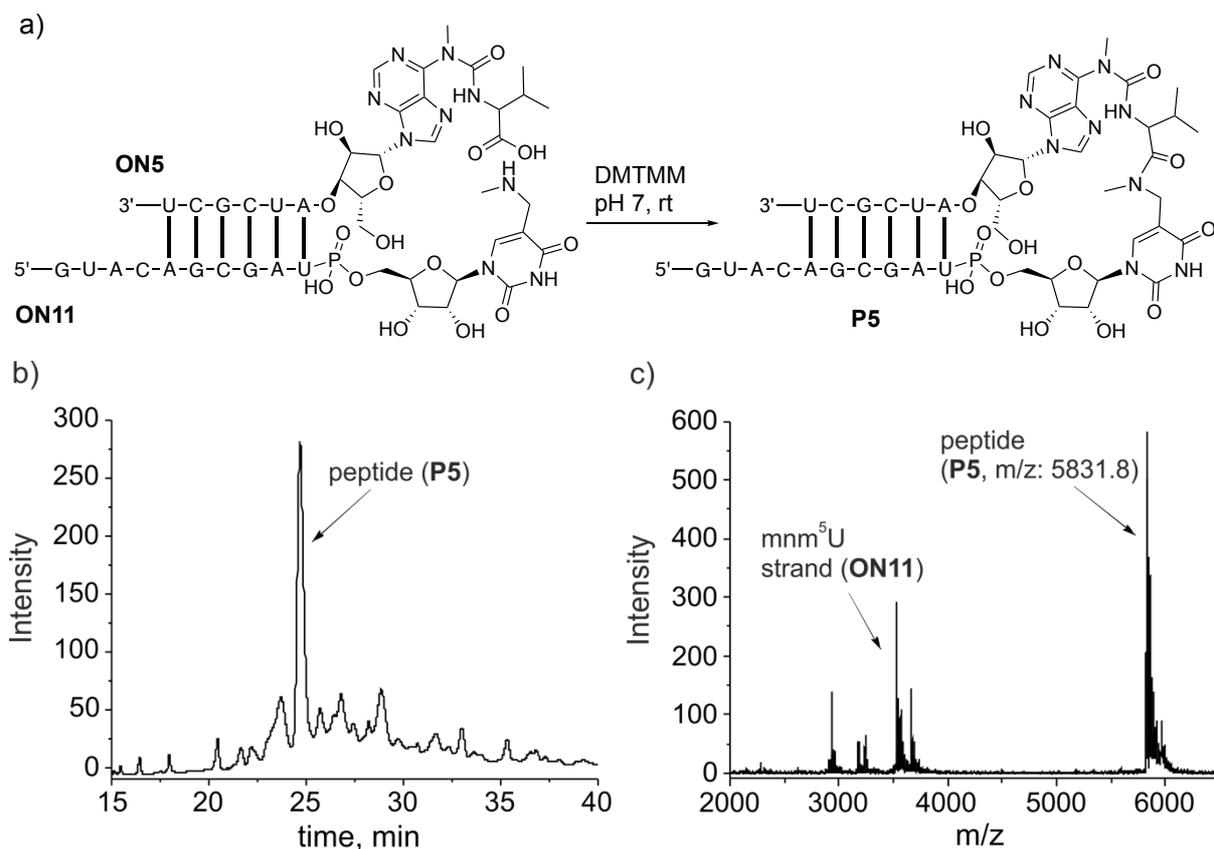
Figure 32. (a) Schematic depiction of peptide **P5** growth between **ON5** and **ON11**; (b) HPL-chromatogram of peptide **P5** formation; (c) MALDI-TOF data of peptide **P5** formation.

The HPLC data together with MALDI-TOF mass spectrum are shown in Figure 32b,c. This time phosphate buffers were used for HPLC analysis and no acetylation was observed. The reaction was initially performed at 4 °C and then at 20 °C, which is still below the melting point (T_m = 31 °C). In both cases the reaction proceeded well, however, it was faster at higher temperatures, so it was decided to keep the reaction mixture at 20 °C for further peptide formation experiments.

After the purification of peptide **P5**, it was heated to 95 °C for 9 h in a PCR machine. During the reaction, the target product indeed formed. However, some degradation products were also

observed as shown in Figure 33. The separation of the target product from the m^{6}A strand was indeed easier this time. Acetate buffers and a C8 column instead of C18 were used for HPLC purification.
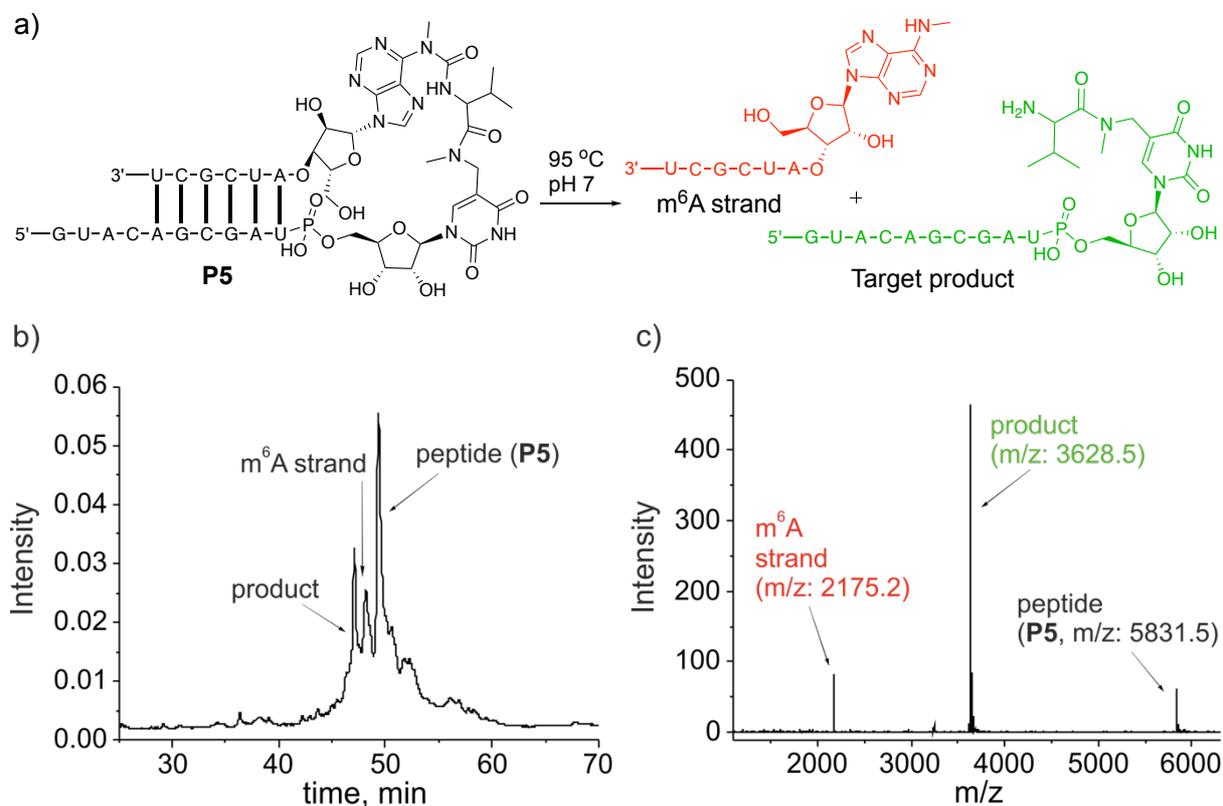
Figure 33. (a) Schematic depiction of urea cleavage of **P5**; (b) HPL-chromatogram of urea cleavage of **P5**, C8 column, TEA/AcOH buffer system; (c) MALDI-TOF mass spectrum of urea cleavage of **P5**.

Although in principle the urea cleavage experiment worked, the amount of target product after each of the purification step decreases. Thus, it was decided to independently synthesize the template strand carrying an amino acid (**ON12**, Scheme 8) and to use it for further peptide growth experiments to make a dipeptide.
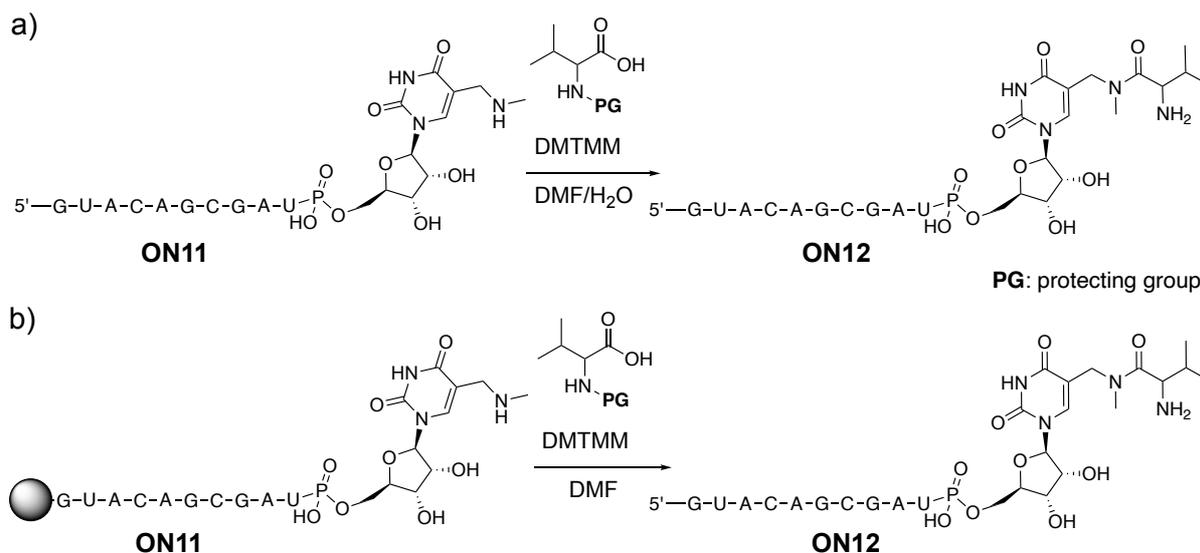
Scheme 8. (a) Amide **ON12** formation from **ON11** with an amino acid in solution; (b) amide **ON12** formation from solid support-bound **ON11** with an amino acid in suspension.

There are in principle two approaches of how to attach an amino acid on the strand without using templation chemistry. One can couple the amino acid, when the RNA is fully dissolved (Scheme 8a) or when it is still resin-bound (Scheme 8b). Attempts to form **ON12** with different activators in DMF/H$_2$O solution were not very promising. Although the product was forming, the conversion was not high (~20%). The amidation reaction of resin-bound RNA in DMF (investigated by *Felix Müller*) was more successful.

In order to attach the next amino acid on the template strand, **ON12** was then reacted with the glycine carrying strand **ON4** in an RNA-templated manner. In this case, the amino group is primary and more reactive. The peptide **P6** was nicely forming as indicated by HPLC and MALDI-TOF data (Figure 34).
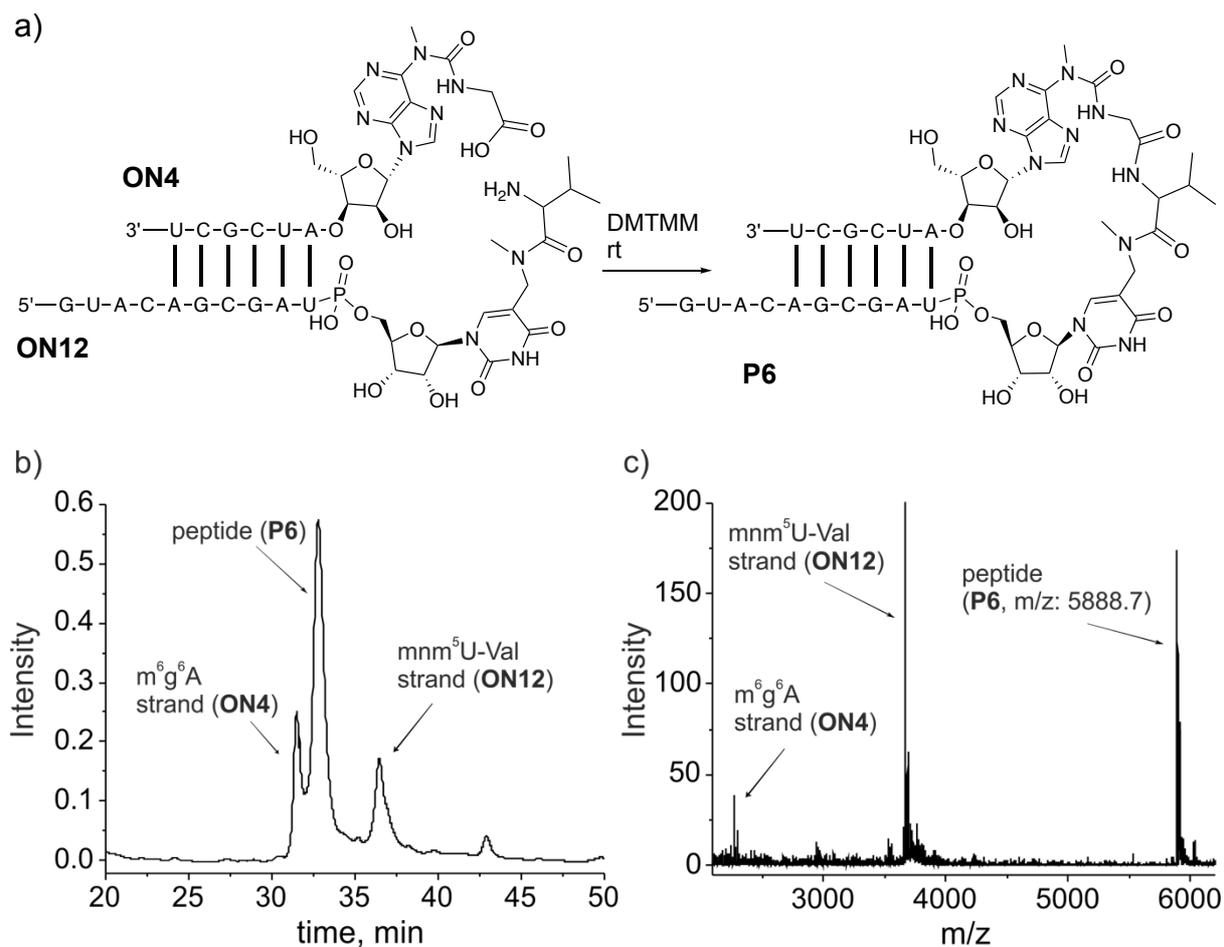
Figure 34. (a) Schematic depiction of peptide **P6** formation; (b), (c) HPL-chromatogram and MALDI-TOF mass spectrum of peptide **P6** formation.

After peptide **P6** was purified, it was subjected to urea cleavage. During the reaction, the intramolecular cyclisation can occur again to form hydantoin as indicated in Figure 35a (green). The peptide was heated to 95 °C in a PCR machine for 9 h at different pH (3.0-8.0) in order to study the influence of pH for the cyclisation reaction.
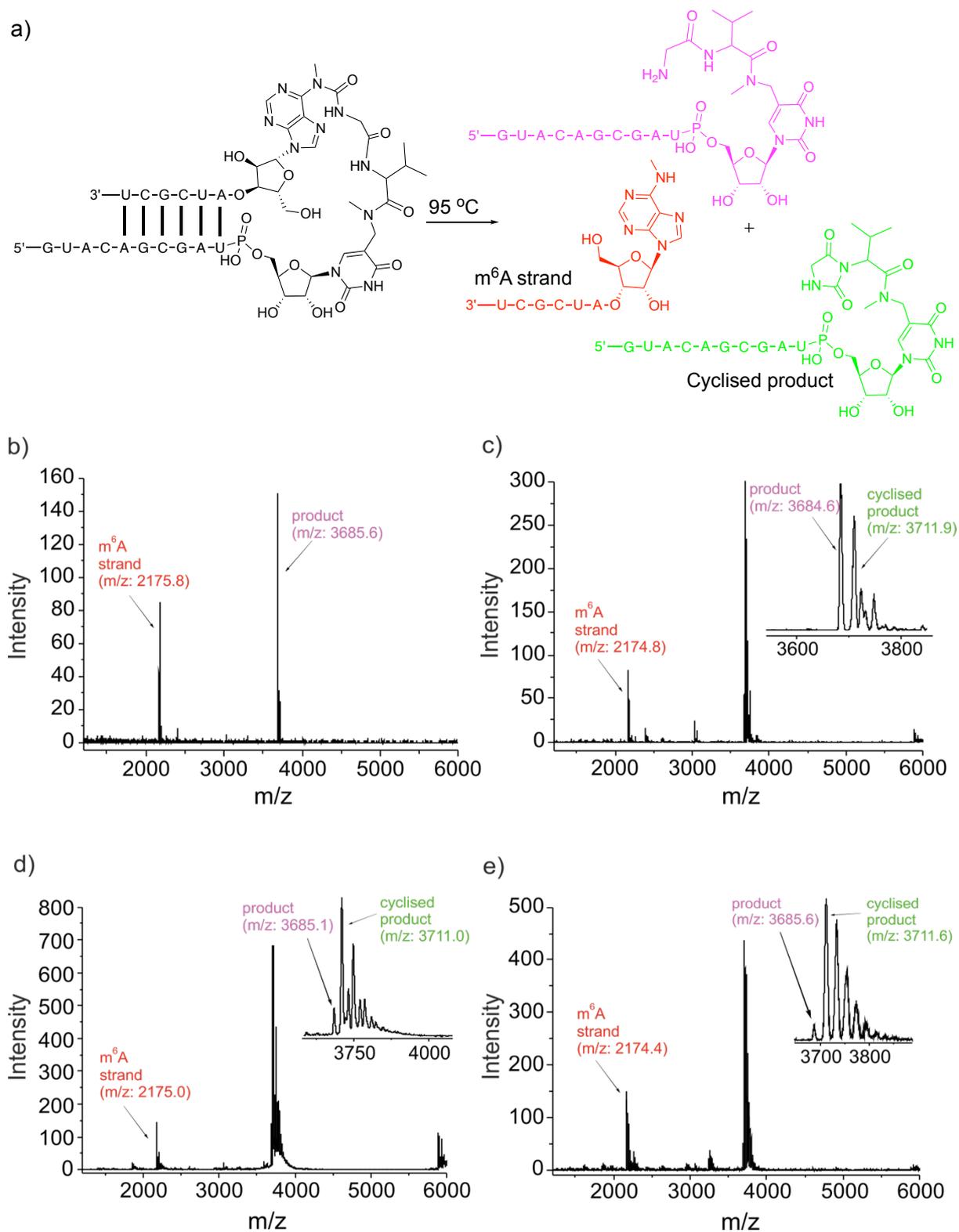
Figure 35. (a) Schematic depiction of urea cleavage of peptide **P6**; urea cleavage at (b) pH 3.3; (c) pH 3.5; (d) pH 4; (e) pH 7.

The reaction was monitored by MALDI-TOF and the data are shown in Figure 35b-e. At pH 3.3, basically only target product (shown in pink) and the m6A strand (shown in red) are formed, thus this pH was used for subsequent urea cleavage experiments. At more basic pH-values, the amount of cyclized product increased and at pH 7 and 8, the cyclized product dominated. At pH 3.0, only RNA degradation products were observed, which is in agreement with the literature, stating that RNA is stable up to pH 4-5. Changing of the pH to more acidic or more basic values induces the degradation of RNA, at 90 °C over prolonged times.[222]

After having two amino acids (Val, Gly) attached to mnm5U, experiments to grow the peptide further were performed. The strand **ON13** containing a Val-Gly moiety was reacted with Leu-carrying RNA oligonucleotide **ON6** (Figure 36a). The reaction was performed for 48 h at room temperature using DMTMM as an activator. Figure 36b and Figure 36c show HPLC data and MALDI-TOF mass spectrum, respectively, proving successful synthesis of **P7**.
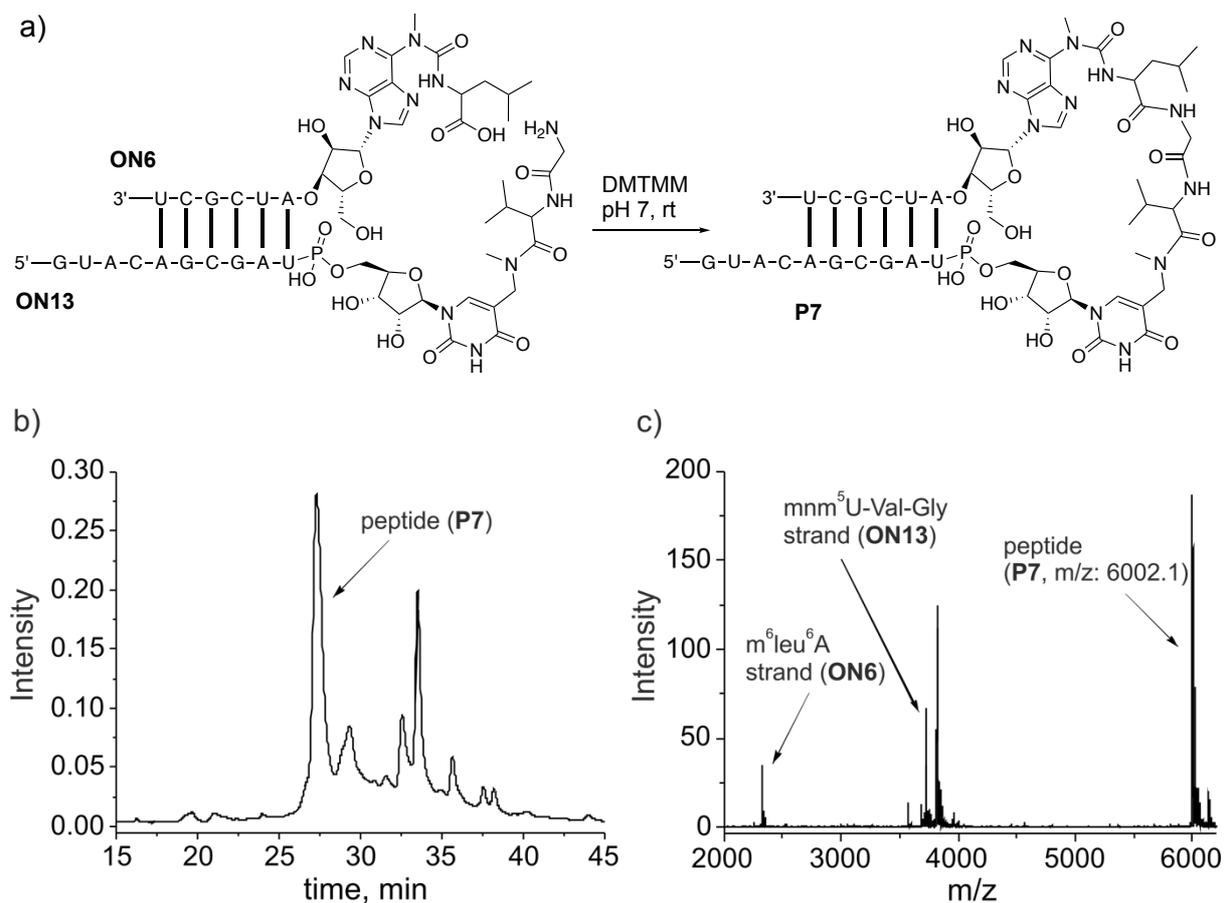
Figure 36. (a) Schematic depiction of peptide **P7** growth between **ON13** and **ON6**; (b) HPL-chromatogram of peptide **P7** formation; (c) MALDI-TOF mass spectrum of peptide **P7** formation.

The peptide **P7** was purified by HPLC, eluted with phosphate buffer and subsequently was heated at 95 °C for 9 h in a PCR machine at pH 3.3. The data show the formation of the target product and the absence of the cyclic hydantoin by-product (Figure 37).

a)

b)
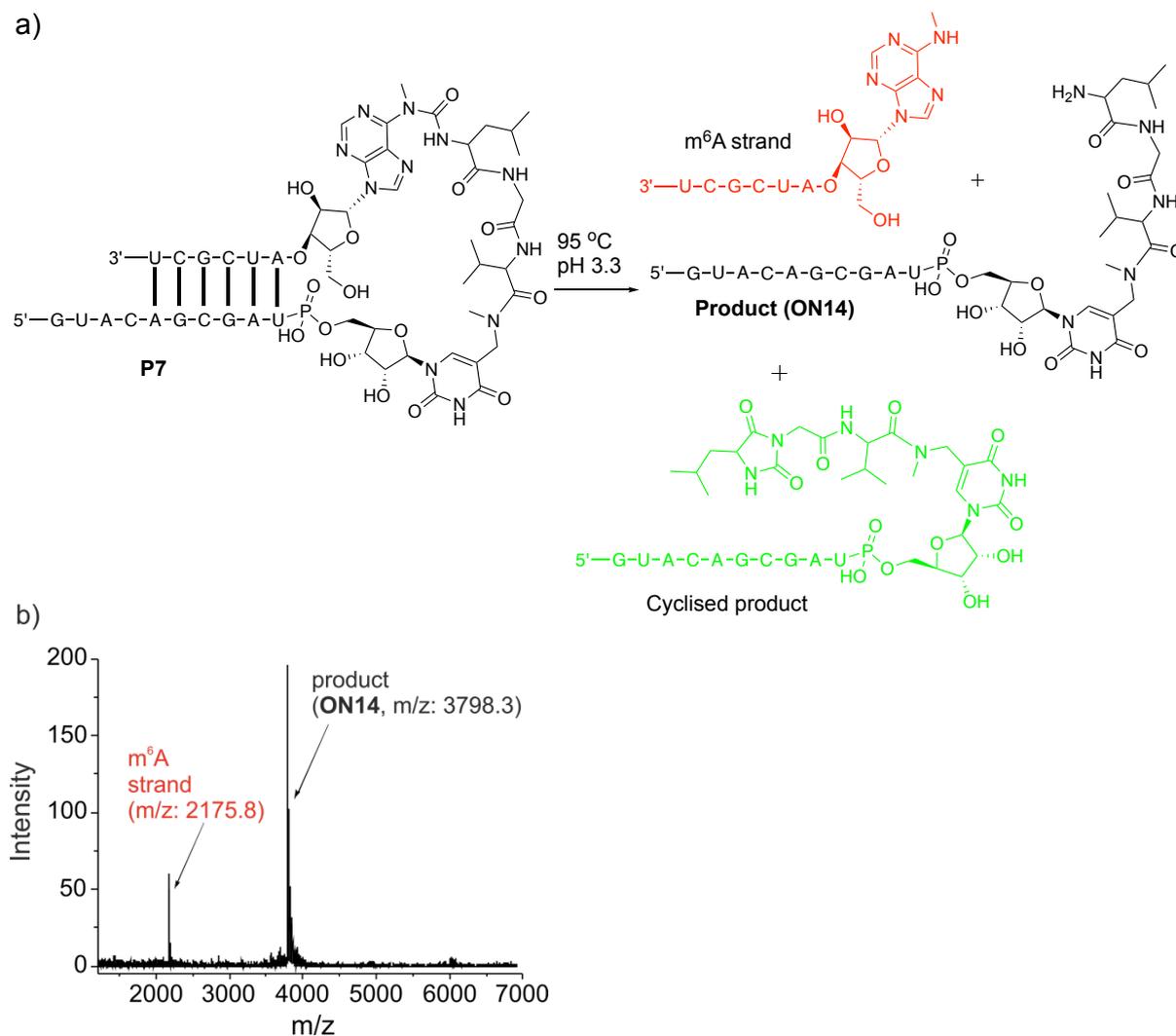
Figure 37. (a) Schematic depiction of urea cleavage of peptide **P7**; (b) MALDI-TOF mass spectrum of urea cleavage of peptide **P7**.

Following our RNA-templated peptide formation we succeeded to attach three amino acids (Val, Gly, Leu) onto the mnm^{5}U modification. Attempts to attach the next amino acid, namely glycine, were not successful, because limited stability of **ON14** was observed. MALDI-TOF and HPLC analyses showed that the RNA-amino acid conjugate encounters stability problems once 3 amino acids are covalently attached. One explanation could be that the additional amino groups from the amino acid induce changes in pK_a values due to the amino group of amino acid

or that peptide forms secondary structure and gains catalytic activity. The latter could be influenced by the change of ionic strength.[223]

5.1.4.2 nm⁵U modification as a candidate for an RNA-templated peptide formation

5-aminomethyluridine, a modification found in the wobble position of tRNA, was also investigated (initially studied by *Dr. Tynchtyk Amatov*) as a potential candidate for RNA-templated peptide formation. Pivom⁵U phosphoramidite building block **43** was used for the incorporation into an RNA strand as already shown in Scheme 5. The solid support-bound pivom⁵U carrying oligonucleotide was then cleaved from the resin under mild conditions using NH₃/MeOH at rt. Subsequently the strand was treated with NH₄OH at 60 °C for an hour to ensure the full substitution of pivaloyl group to an amino group and form **ON15** (Scheme 9).[221]
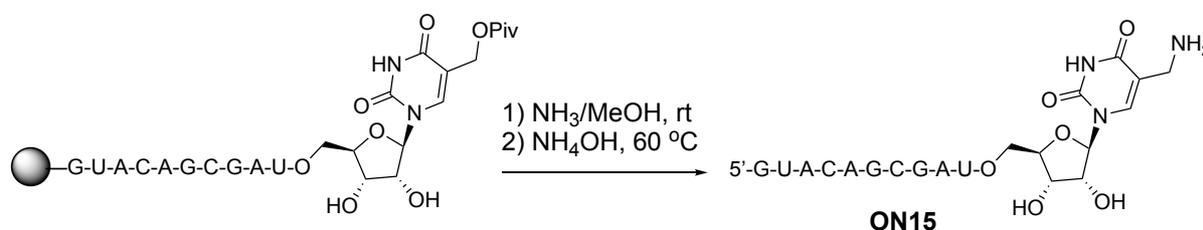
Scheme 9. Schematic depiction of pivom⁵U-modified strand conversion to nm⁵U containing oligonucleotide **ON15**.

Figure 38 shows the HPL-chromatogram of an unpurified and the chromatogram of a pure nm⁵U containing RNA oligonucleotide **ON15** in the inset together with its MALDI-TOF mass spectrum proving successful incorporation and subsequent derivatization of **ON15**.
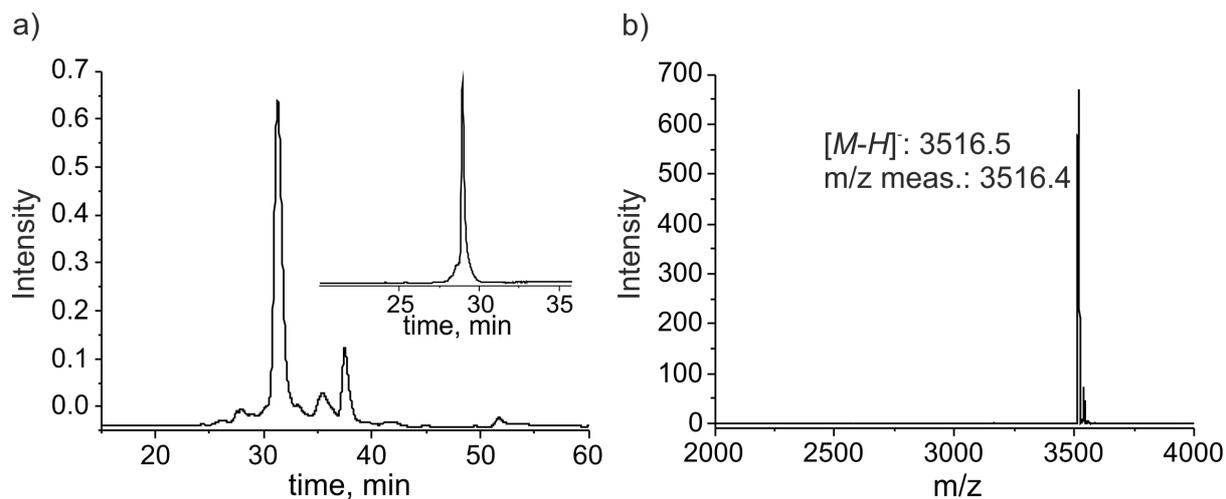
Figure 38. (a) HPL-chromatogram of the unpurified **ON15**, with the inset showing the HPL-chromatogram of purified **ON15**; (b) MALDI-TOF mass spectrum of **ON15** after purification.

The purified oligonucleotide **ON15** was reacted with valine carrying strand **ON5** in order to form the peptide bond (Figure 39a). The reaction was performed again by activating with DMTMM at room temperature. After 48 h, almost full conversion was reached. Figure 39b and Figure 39c show the data obtained for the formation of peptide **P8** and prove its successful synthesis.

a)

b)

c)
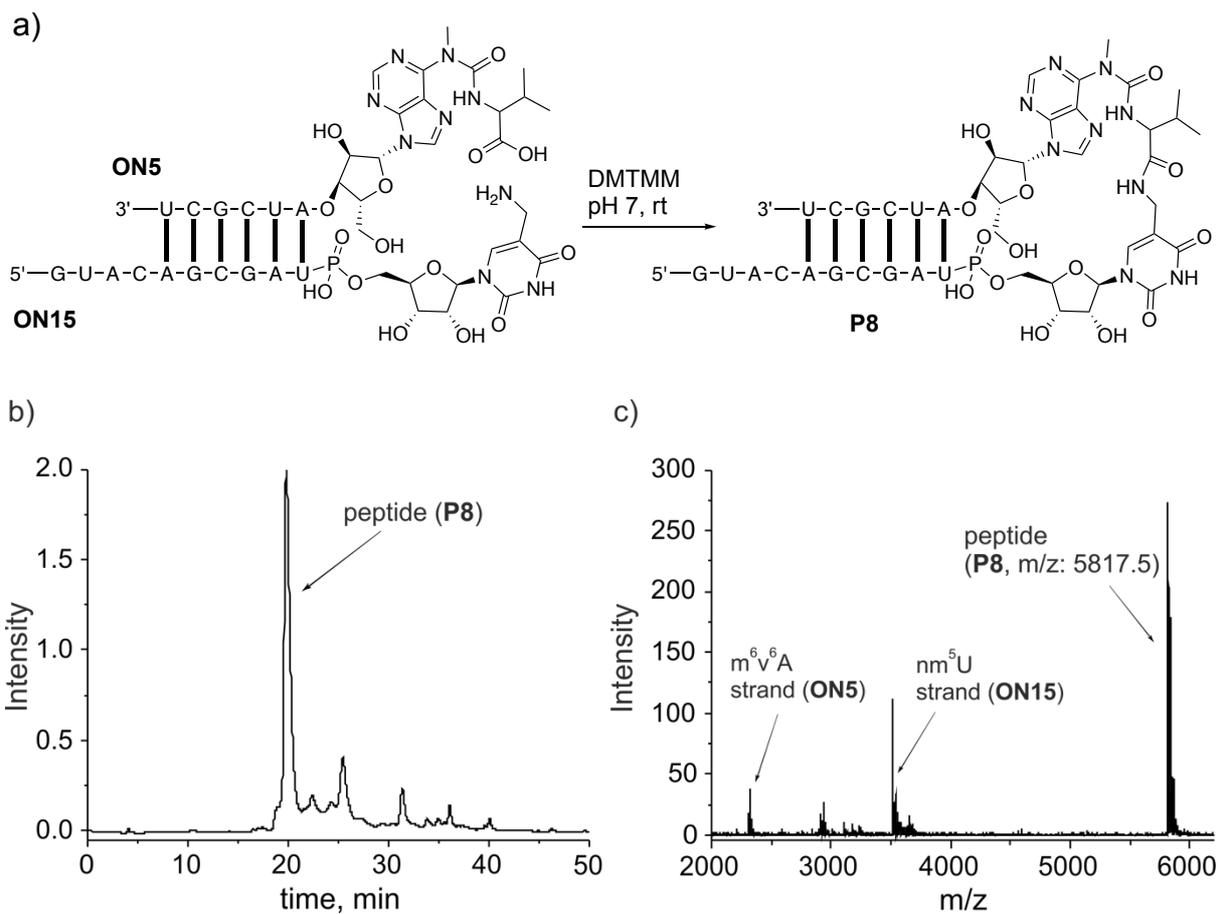
Figure 39. (a) Schematic depiction of **P8** formation; (b) HPL-chromatogram of peptide **P8** formation; (c) MALDI-TOF mass spectrum of peptide **P8** formation.

The purification of peptide **P8** was performed using HPLC, eluting with phosphate buffer. The purified peptide was heated at 95 °C for 9 h at pH 3.3 as depicted in Figure 40a.
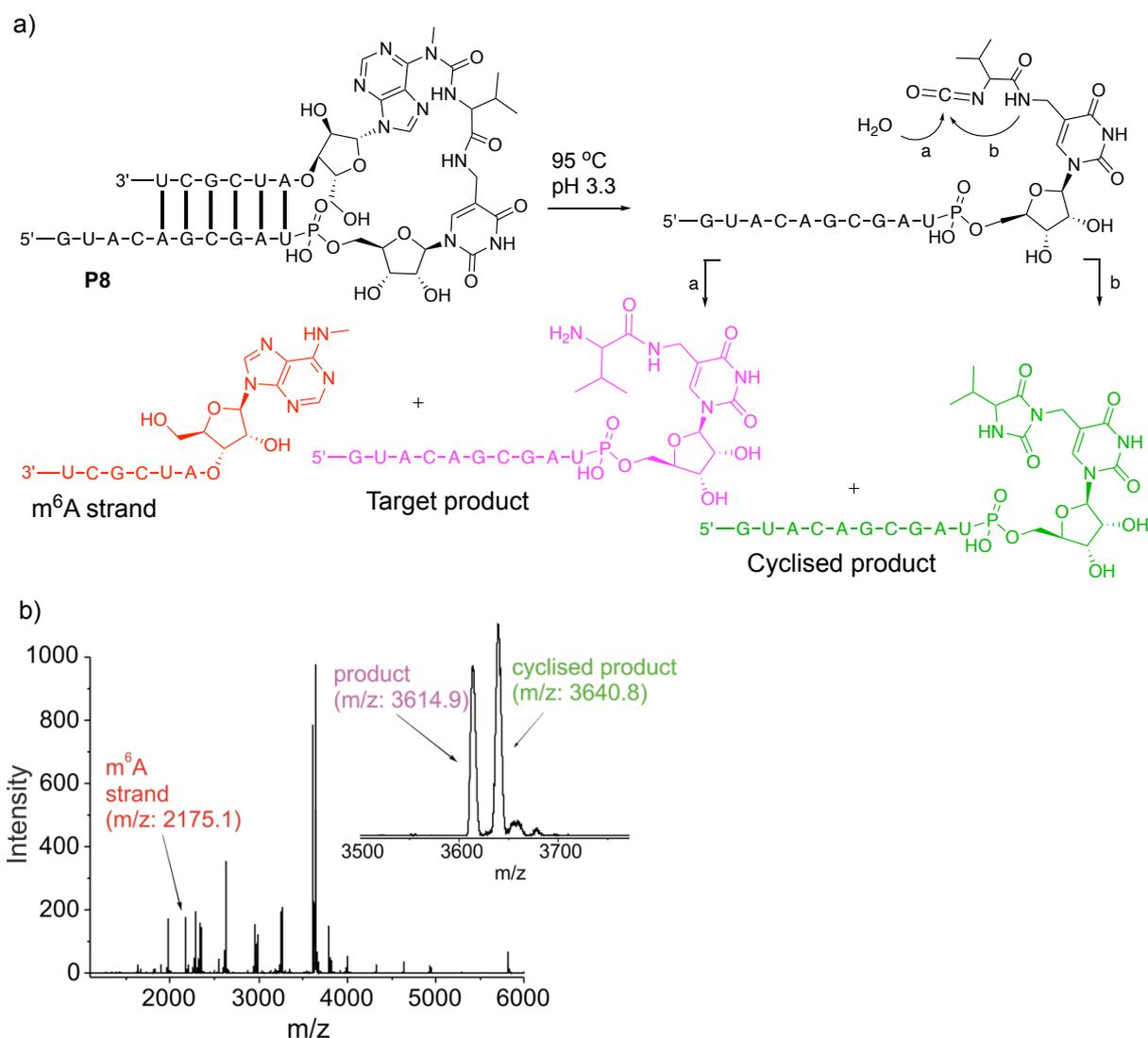
Figure 40. (a) Schematic depiction of urea cleavage of peptide **P8**; (b) MALDI-TOF mass spectrum of urea cleavage of peptide **P8**.

Although changing the pH to 3.3 for the mnm^5U modified strands helped to suppress the cyclisation (Figure 35), in the case of the nm^5U modified oligonucleotides, the cyclized product was still forming to a great extent. The MALDI-TOF mass spectrum (Figure 40b) shows the formation of the targeted product but only together with a lot of degradation products. Attempts to form more target product by changing the pH were not successful. More basic conditions provided only the cyclized version of the product, while more acidic conditions led to degradation of oligonucleotide. Thus, the modification was not investigated further.

5.2 Important factors for peptide growth

In principle various elements could influence the efficiency of an RNA-templated peptide growth. Factors like complementarity, secondary structure of RNA strands and distance

between the reactive centers should greatly impact the success of peptide formation. We used acp^{3}U modification to determine these elements. Thus, a mismatched strand containing m^{6}t^{6}A (Figure 41a) that does not base pair with the acp^{3}U containing RNA oligonucleotide was synthesized. Indeed, the peptide bond was not forming between these two strands proving that complementarity is important for peptide formation.

a) **Complementarity**

```
3′ C-G-U-A-C-G-m⁶t⁶A 5′
                          ⟶✕⟶
5′ A-G-C-G-A-U-acp³U-U 3′
```

b) **Distance**

```
3′ G-G-A-U-G-G-U-G-U-G-G-G-m⁶v⁶A 5′
   │ │ │ │ │ │ │ │ │ │ │ │
5′ C-C-U-A-C-C-A-C-A-C-C-A-U-A-G-C-G-A-U-acp³U-U 3′   ⟶✕⟶
```

c) **Secondary structure**

```
    C-C      A     C-C-C 5′
  /    \C-G/ \C-U-G/
 C      │ │  │ │ │        + 3′ U-C-G-C-U-A-m⁶t⁶A 5′ ⟶✕⟶
 │     G-C-G-A-U-acp³U-U 3′
 A
  \U-A/
```

Figure 41. Schematic depiction of (a) reaction of acp^{3}U strand with mismatched m^{6}t^{6}A strand, (b) reaction of acp^{3}U strand with m^{6}v^{6}A strand with big proximity between reaction centers, (c) reaction of acp^{3}U strand having a secondary structure and m^{6}t^{6}A strand.

The distance dependency was examined between an acp^{3}U containing strand and an m^{6}v^{6}A oligonucleotide bearing a remote carboxyl group (Figure 41b). No product formation was observed, although if a longer peptide would be already attached, the amidation could potentially occur due to increased flexibility of the attached chain. The influence of RNA secondary structure was demonstrated after the acp^{3}U-modified strand (Figure 41c) showed no product formation with the m^{6}t^{6}A containing strand while the same reaction took place when the strand was unable to form a secondary structure.

5.3 Conclusions and outlook

In the first part of the thesis, the properties of 5 different RNA modifications (k^{2}C, k^{2}U, acp^{3}U, mnm^{5}U and nm^{5}U) as amino acid acceptors and (m^6)aa^{6}A bases as amino acid donors were investigated together with the possibility to use them for an RNA-templated peptide formation system. The synthesis of a k^{2}C phosphoramidite was unsuccessful, however the synthetic pathways could be reinvestigated. The phosphoramidite building blocks of k^{2}U, acp^{3}U, mnm^{5}U

and nm^{5}U were successfully synthesized and then incorporation protocols into oligonucleotides were developed. k^{2}U modified RNA oligonucleotides proved to be successful in peptide bond formation, nonetheless, as k^{2}U being the deamination product of k^{2}C was not yet biologically detected, other potential acceptor nucleosides were tested. Then peptide bond formation under various conditions was investigated and successfully achieved between acp^{3}U, mnm^{5}U or nm^{5}U modified RNA oligonucleotides and several m^{6}aa^{6}A base containing RNA strands. The formed hairpins were subjected to urea cleavage conditions, providing the target or cyclised products, depending on the pH applied. Cyclised product dominated when the acp^{3}U modification was used. However, the conditions to provide just the target product could potentially be optimized by the change of pH. The target product was predominantly formed when the mnm^{5}U modification was used at pH 3.3. nm^{5}U modified oligonucleotides provided both the target and the cyclized product during the urea cleavage experiments, together with RNA degradation products. aa^{6}A bases in principle were suitable for an RNA-templated peptide formation and subsequent urea cleavage, however reaction mixtures using m^{6}aa^{6}A bases were cleaner.
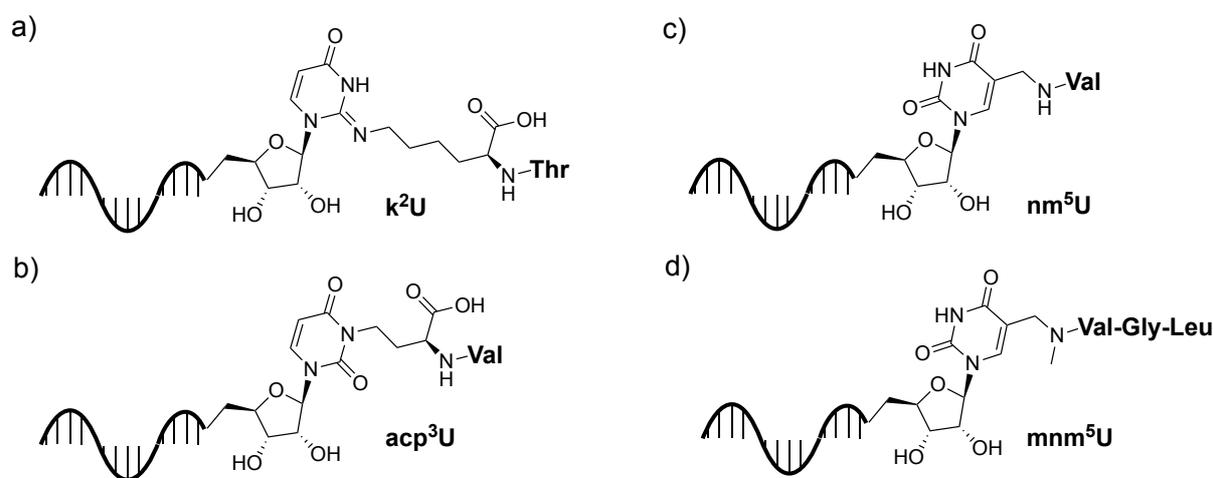
Figure 42. Schematic depiction of RNA-peptide conjugates synthesized via an RNA-templated peptide formation approach.

Further investigation of the RNA degradation pattern caused by pH and temperature need to be performed in the future. Following our proposed RNA-templated peptide formation, RNA-peptide conjugates were synthesized as represented in Figure 42, showing that the most promising RNA modification for this approach is mnm^{5}U. Further investigations about the formation of longer peptides and their potential catalytic activity are currently ongoing.

Part II

5.4 Synthesis of siRNAs against SARS-CoV-2 infection

In the second part of the doctoral thesis the knowledge about modified RNA bases was used for the design and synthesis of siRNA conjugates against the severe acute respiratory syndrome betacoronavirus 2 (SARS-CoV-2), first identified in December 2019 in China, which resulted in a worldwide exceptional pandemic.

In general, coronaviruses are positive-sense single-stranded RNA viruses with a genome size ranging from approximately 26 to 32 kb. The genome has a 5' cap structure with a modified 7-methylguanosine (m^7G) and m^6A as well as 2'-OMe-adenosine (A_m) modified bases along with a 3' poly(A) tail, enabling it to act as an mRNA for translation.[224] The replicase gene encodes the non-structural proteins (Nsps). It occupies two-thirds of the virus genome, with about 20 kb. Structural and accessory proteins make up about 10 kb of the viral genome. SARS-CoV-2 is an enveloped betacoronavirus that features a genome that encodes 16 Nsps (Nsp1-16), 4 structural proteins (spike, envelope, membrane, nucleocapsid), and 9 accessory factors as shown in Figure 43.
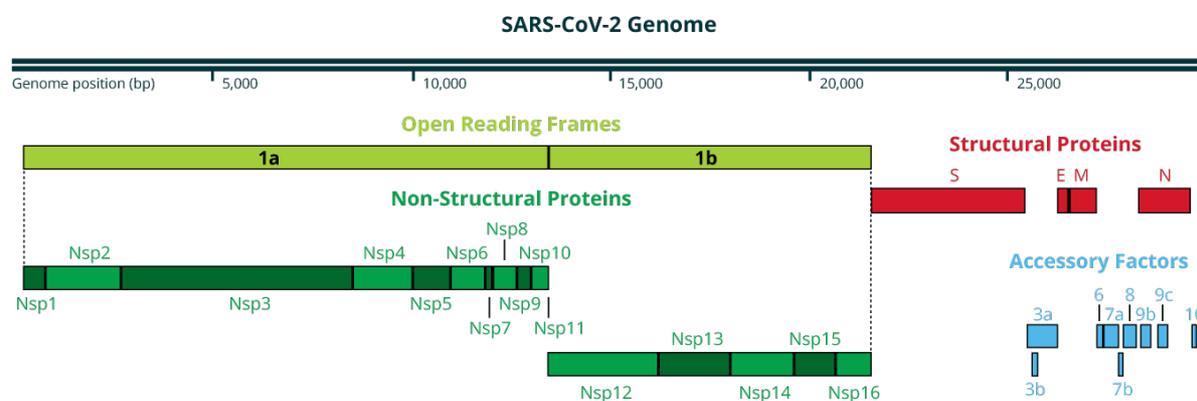
SARS-CoV-2 Genome

Figure 43. Genomic organization of SARS-CoV-2. S: spike, E: envelope, M: membrane, N: nucleocapsid proteins. Nsp: non-structural proteins. Adapted from *Bosterbio*.

The Nsp proteins assemble into the replicase-transcriptase complex (RTC) that creates an environment suitable for RNA synthesis. The proteins are responsible for RNA replication and transcription of sub-genomic RNAs. The spike protein (~150 kDa) is responsible for the cell entry. The protein binds to the receptor called ACE2 found on the surface of human cells.[224,225] In this doctoral thesis, several siRNA sequences (M1, M2 and M17, Figure 44) were designed in order to target the RNA-dependent RNA polymerase (Nsp12) and the spike protein.

Thymidine overhangs were introduced into the sequences in order to enhance exonuclease resistance of the siRNAs in the cell culture medium and within transfected cells. We furthermore introduced the modified bases 2'-O-methylcytidine (C_m) and 2'-O-methyluridine (U_m) to stabilize the siRNA regarding the action of endonucleases. The biological activity of the siRNAs was investigated by utilizing a commercially available dual-luciferase reporter assay. A plasmid containing the *Firefly* luciferase and the *Renilla* luciferase genes, the latter fused to CoV-2 target fragments was transfected into the cells. RNA interference was examined by targeting the expression of the *Renilla* luciferase, while the *Firefly* luciferase was used as an internal standard. Such reporter assays are established for the testing of siRNA sequences.[187,226] Biological experiments were performed by *Dr. Franziska Traube* and *Ammar Ahmedani*. As depicted in Figure 44, all siRNAs tested in A549 cells using the RNAiMAX transfection reagent showed strong silencing effects caused by specific binding of the siRNAs to the mRNA target, however the M17 sequence appeared to be the most potent and it was decided to use it for further investigations.
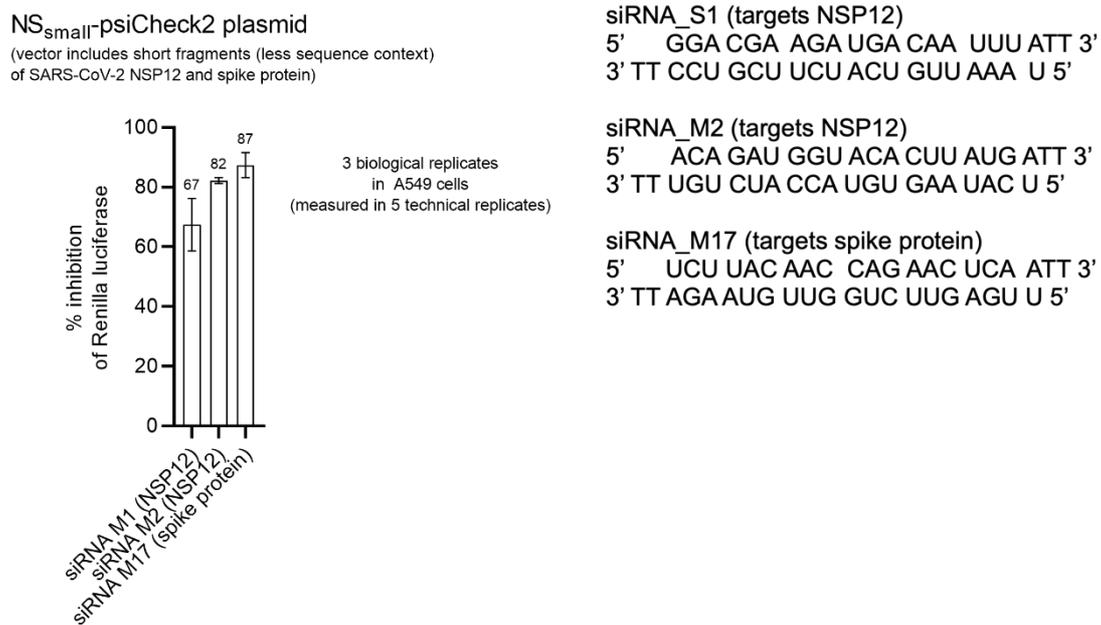
Figure 44. Silencing of the *Renilla* luciferase compared to the silencing of the *Firefly* luciferase mediated by siRNAs in A549 cells transfected with RNAiMAX transfection reagent.

As discussed in the introduction, the hurdles for siRNAs to become successful drugs are their stability and delivery into target cells. Conjugation with different molecules such as sugars or lipids has been used to improve pharmacokinetic properties of siRNA, without decreasing their

potency. Thus, we aimed to synthesize modified M17 sequences (**ON16** and **ON17**) containing several functionalities as shown in Figure 45a.
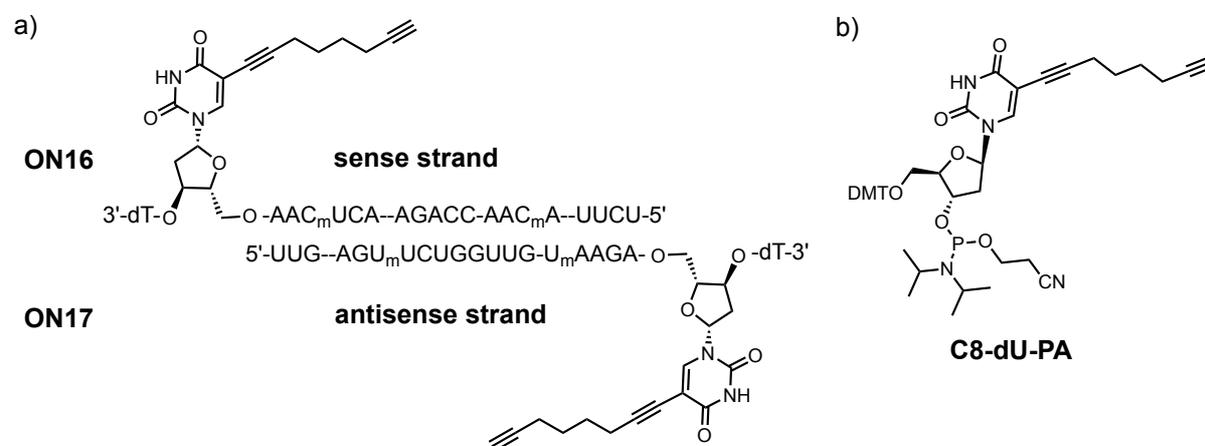
Figure 45. Schematic depiction of (a) modified M17 sequence with alkynyl- and 2'-OMe moieties. C_m: 2'-OMe-C, U_m: 2'-OMe-U; (b) alkyne moiety containing phosphoramidite building block C8-dU-PA.

Modified RNA bases with two 2'-OMe groups were introduced in each of the strands in order to increase the stability of siRNAs against nucleases. In addition, we modified the siRNA ends with a commercially available DNA phosphoramidite featuring in addition a C8 spacer between a reactive alkyne group and the base (C8-dU-PA, Figure 45b). The alkyne group was chosen in order to address the delivery issue by allowing to form bioconjugates with sugars or lipids using a "click" reaction. Such conjugates could potentially enter the cell via receptor-mediated uptake or natural transport mechanisms without using transfection reagents. Figure 46a and Figure 46b show the data for the **ON16** and **ON17** oligonucleotides, proving successful incorporation of U_m, C_m and of the alkynyl modified phosphoramidites.
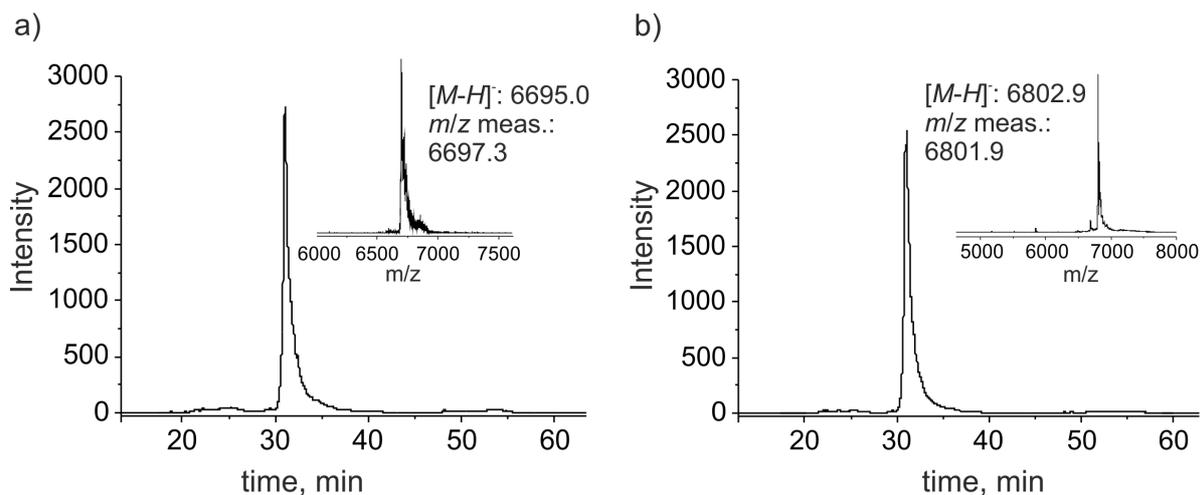
Figure 46. (a) HPL-chromatogram of unpurified **ON16**, with the inset showing the MALDI-TOF mass spectrum of purified **ON16**; (b) HPL-chromatogram of unpurified **ON17**, with the inset showing the MALDI-TOF mass spectrum of purified **ON17**.

Glucose, arachidonoyl- and cholesteryl azides were chosen as reactants in the 1,3-dipolar cycloaddition[227] with the alkynes present in the siRNA, for a so called "click" reaction (Scheme 10). Attempts to perform the "click" reaction in suspension with a $CuSO_4$/Na ascorbate system and resin-bound RNA led to no conversion to target product. Glucose was, however, successfully attached by using the CuBr/TBTA (tris(benzyltriazolylmethyl)amine) catalytic system. The RNA was not resin-bound but in solution. Full conversion was reached within 30 min. The glucose-siRNA conjugates **ON18** and **ON19** (modified via 6th position of glucose) were subsequently purified by HPLC. The same conditions were next applied for the reaction of M17 with arachidonoyl and cholesteryl azides. It is important to mention that in this case the reactions were carried in glass vials, because the formed lipid-siRNA conjugates (**ON20-ON23**, Scheme 10) were bounding to the walls of the plastic vials, resulting in low yields.

ON16 sense strand

3'-dT-O~~~O -AAC$_m$UCA--AGACC-AAC$_m$A--UUCU-5'

5'-UUG--AGU$_m$UCUGGUUG-U$_m$AAGA- O~~~O -dT-3'

ON17 antisense strand

1: arachidonoyl (AA)

N$_3$~~~O~~~O~~~O-NH-C(=O)-arachidonoyl

2: glucose (Glc)

3: cholesteryl (Chol)

N$_3$~~~O~~~O~~~O-NH-cholesteryl

click chemistry

1, 2 or 3 sense strand

3'-dTdT-AAC$_m$UCA--AGACCAAC$_m$A--UUCU-5'

5'-UUG-AGU$_m$UCUGGUUG-U$_m$AAGA-dTdT-3'

antisense strand **1, 2 or 3**
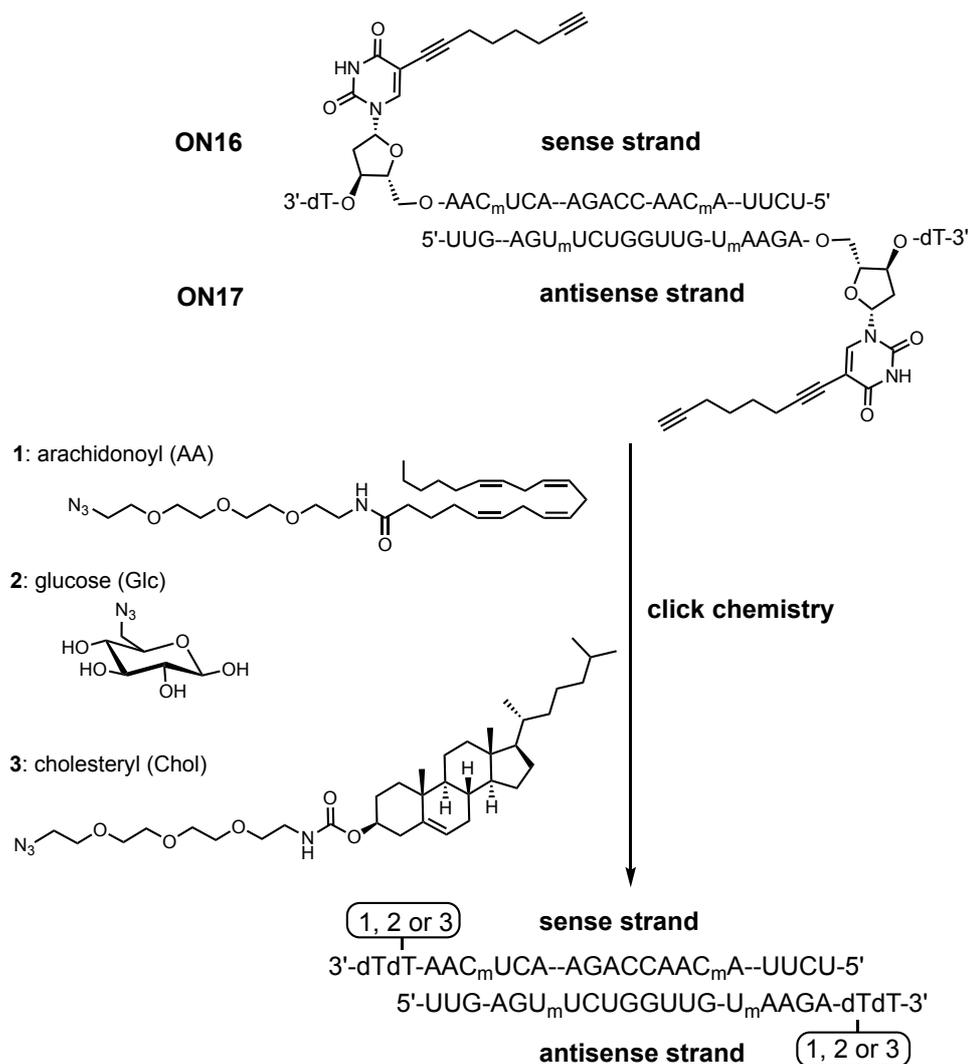
ON18: Glc on the sense strand; **ON19**: Glc on the antisense strand; **ON20**: AA on the sense strand
ON21: AA on the antisense strand; **ON22**:Chol on the sense strand; **ON23**: Chol on the antisense strand

Scheme 10. "Click" reaction of methylated siRNAs.

The lipid-modified siRNAs (**ON20-ON23**) were next purified by HPLC, with heating of the column to 65 °C. At lower temperatures, the RNA oligonucleotides did not elute properly. Figure 47 shows the data obtained for glucose- and lipid-modified siRNAs (modifications are attached on the sense strand). Figure 47a contains two overlapping peaks caused by the isomers of α/β glucose.

a)

b)

c)
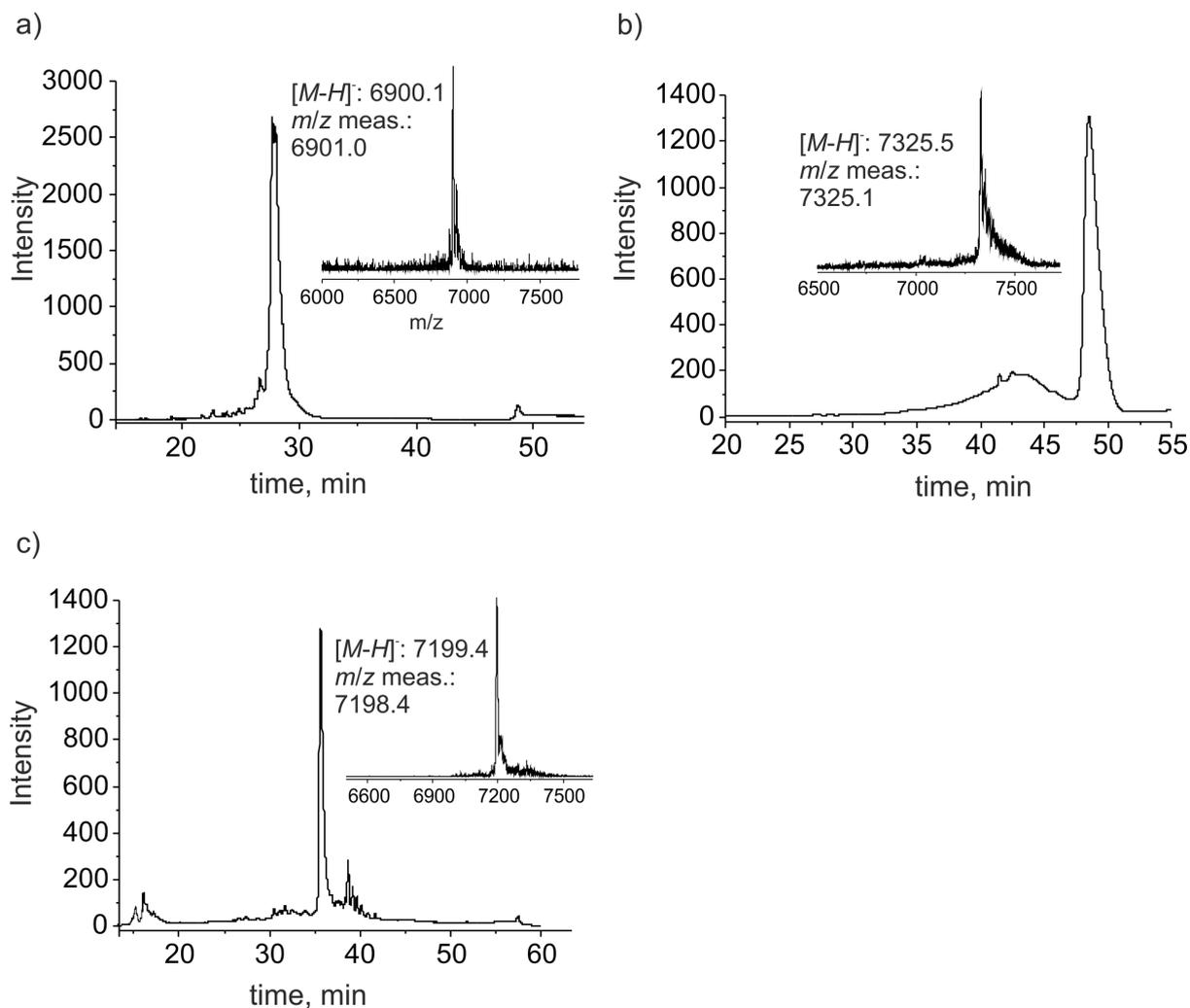
Figure 47. (a) HPL-chromatogram of unpurified **ON18**, with the inset of MALDI-TOF mass spectrum of purified **ON18**; (b) HPL-chromatogram of unpurified **ON20**, with the inset of MALDI-TOF mass spectrum of purified **ON20**; (c) HPL-chromatogram of unpurified **ON22**, with the inset of MALDI-TOF mass spectrum of purified **ON22**.

Initially, the duplexes composed of the sense strands (**ON18**, **ON20** or **ON22**) and the antisense strands containing just two 2'-OMe groups together with a fully unmodified M17 duplex as a control were examined for biological activity using the luciferase-based assay. For transfection, two different transfection reagents, lipofectamine RNAiMAX or non-liposomal, polymer-based JetPRIME were examined. Although in principle these siRNAs were designed to be used without any transfection agent, we tested with and without transfection agents to evaluate the best way of application. The experiments were performed by *Dr. Franziska Traube* and *Ammar Ahmedani*. The data are summarized in Figure 48. They show that administration of siRNAs resulted in robust RNAi-mediated gene silencing in two different cell lines (A549, Calu3) – the knockdown efficiencies were better with RNAiMAX transfectant and reached 83-93%

depending on the cell line. It is important to mention that the potency of siRNAs was not reduced by the sugar or lipid attached.
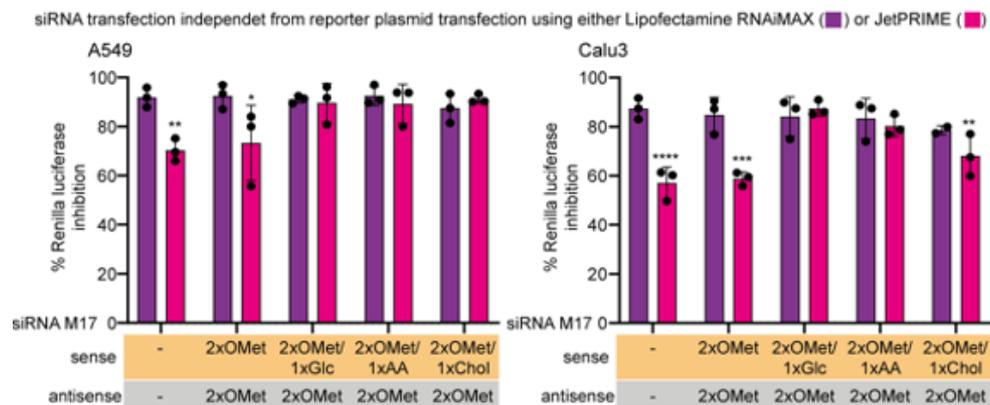
Figure 48. *Renilla* luciferase inhibition in A549 and Calu3 cells mediated by siRNAs containing glucose (Glc), arachidonoyl (AA) and cholesteryl (Chol) moieties on the sense strands. The cells were transfected with RNAiMAX or JetPrime transfection reagents.

Since the interaction between carbohydrates and carbohydrate-binding proteins is quite week, multiple carbohydrate moieties linked to oligonucleotides could increase the affinity. It was also reported that in order to ensure efficient glycotargeting, multivalent glycans are usually required.[228] Thus, we envisioned that the introduction of two or three glucoses could favor cellular uptake by receptor-mediated endocytosis. In parallel we wanted to investigate whether the positioning of the sugars on the siRNA had an influence on the knockdown efficiency. In general siRNA can be modified in three ways: 1) on both sense and antisense strands, 2) just on the sense strand, 3) just on the antisense strand. Thus, siRNAs containing two and three glucoses on the sense and antisense strands needed to be prepared.
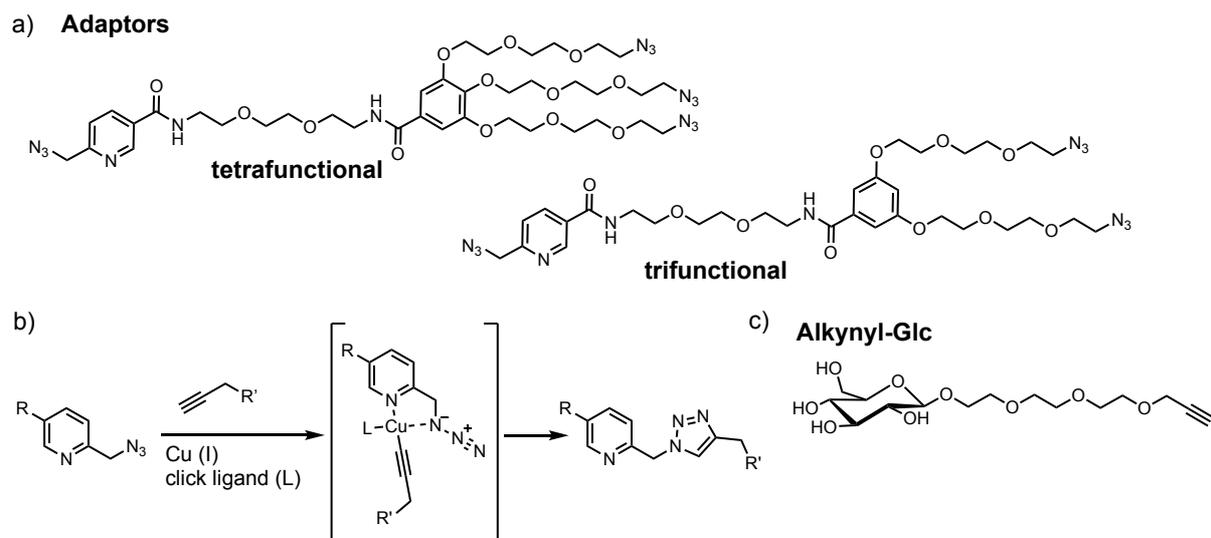
Figure 49. Schematic depiction of (a) adaptors containing picolyl and conventional azido groups; (b) chelation assisted Cu(I)-catalyzed "click" reaction with picolyl azides; (c) alkynyl-glucose modified via anomeric position.

We decided to synthesize the multifunctional adaptors shown in Figure 49a, which feature azido groups possessing different reactivities. It is known that Cu(I)-chelating picolyl azides are way more reactive in "click" reactions (Figure 49b) than conventional azides.[229] First, the more reactive picolyl azide could be employed to attach an oligonucleotide by the means of a first "click" reaction. Then the conventional azido groups could be used to connect them with an alkyne moiety containing glucose derivative (alkynyl-Glc, Figure 49c). The syntheses of adaptors and alkynyl-glucose modified via anomeric position were performed by *Philipp Streshnev* (synthesis procedures to be published in the PhD thesis of *Philipp Streshnev*).

The first "click" reaction between siRNA and picolyl azide (10 eq.) was accomplished (Scheme 11) while applying CuBr/TBTA catalytic system. The reaction was monitored by MALDI-TOF.

ON16   **sense strand**

3'-dT-O ⎯O -AAC$_m$UCA--AGACC-AAC$_m$A--UUCU-5'

5'-UUG--AGU$_m$UCUGGUUG-U$_m$AAGA- O⎯ O -dT-3'

ON17   **antisense strand**

**Trifunctional or
tetrafunctional adaptor**

N=N
N–**Adaptor**

sense strand

3'-dT-O ⎯O -AAC$_m$UCA--AGACC-AAC$_m$A--UUCU-5'

5'-UUG--AGU$_m$UCUGGUUG-U$_m$AAGA- O⎯ O -dT-3'

antisense strand

Adaptor–N
N=N

ON24: trifuncional of the sense strand
ON25: trifuncional of the antisense strand
ON26: tetrafuncional of the sense strand
ON27: tetrafuncional of the antisense strand

Scheme 11. Schematic depiction of the "click" reaction with trifunctional or tetrafunctional adaptors.

Figure 50 shows the data obtained for the attachment of trifunctional and tetrafunctional adaptors on the sense strand via "click" chemistry. After 30 min, full conversion was reached (monitored by MALDI-TOF and HPLC) and no cross reactions with other azido groups were observed.
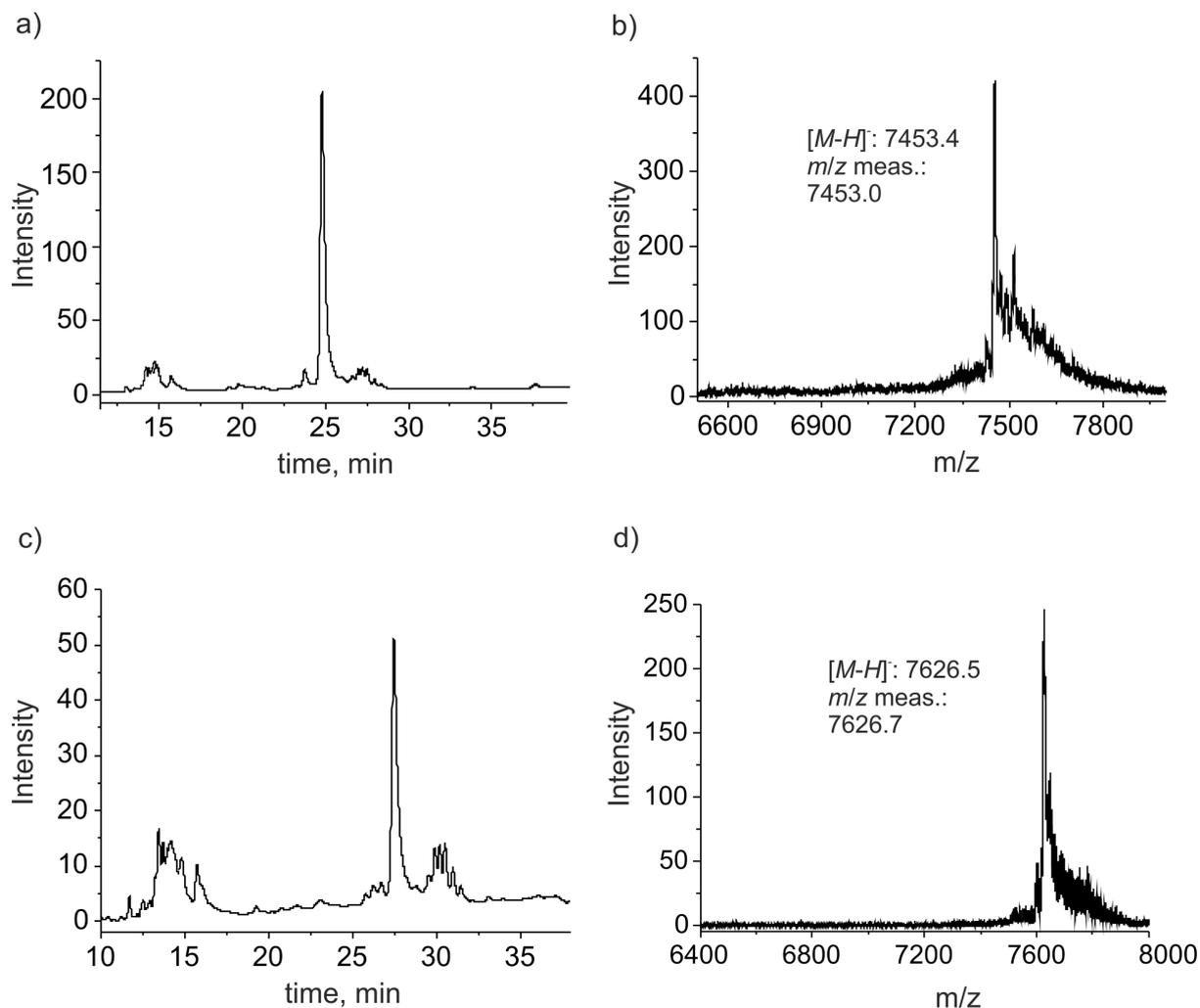
Figure 50. (a) HPL-chromatogram of unpurified trifunctional adaptor modified oligonucleotide **ON24**; (b) MALDI-TOF mass spectrum of unpurified oligonucleotide **ON24**; (c) HPL-chromatogram of unpurified tetrafunctional adaptor modified oligonucleotide **ON26**; (d) MALDI-TOF mass spectrum of unpurified oligonucleotide **ON26**.

The formed adaptor-siRNA conjugates **ON24**-**ON27** were directly used for the second "click" reaction with alkynyl-Glc (modified via anomeric position) without prior purification (Scheme 12). After 1 h, full conversion was reached, then the strands (**ON28**-**ON30**) were precipitated and subsequently purified by HPLC.

ON24-ON27

Alkynyl-Glc

sense strand

3'-dT-O ... O -AAC$_m$UCA--AGACC-AAC$_m$A--UUCU-5'

5'-UUG--AGU$_m$UCUGGUUG-U$_m$AAGA- O ... O -dT-3'

antisense strand

R$_{Glc}$ =

or

ON28: 2 Glc on the sense strand; **ON29**: 2 Glc on the antisense strand;
ON30: 3 Glc on the sense strand; **ON31**: 3 Glc on the antisense strand
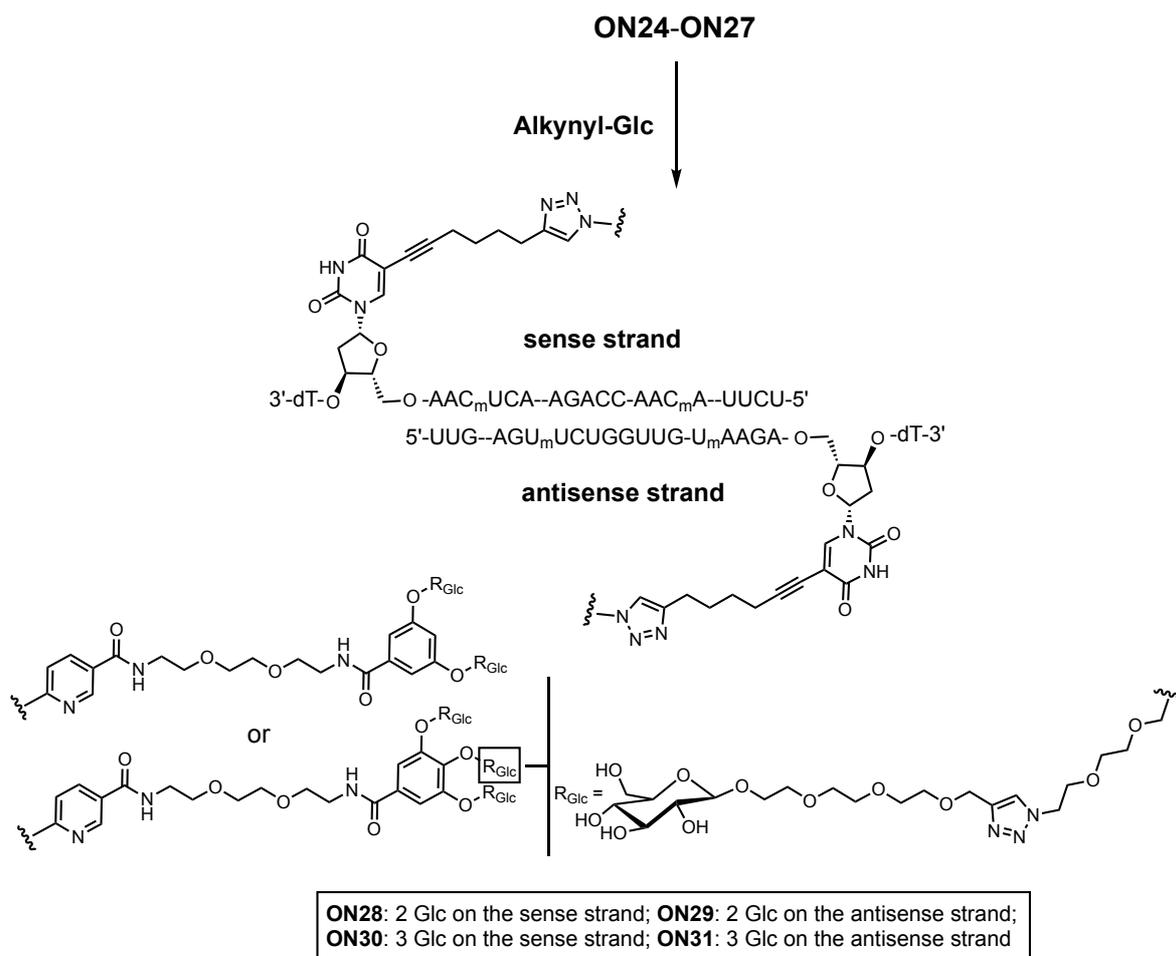
Scheme 12. Synthesis of sugar modified siRNAs **ON28-ON31**.

Figure 51 depicts the data obtained for sugar modified sense and antisense oligonucleotides **ON28-ON31** (modified via anomeric position of glucose). HPL-chromatograms show several overlapping peaks indicating the existence of several isomers of glucose (the peaks were separated to make sure that they correspond to the target product and not to impurities).
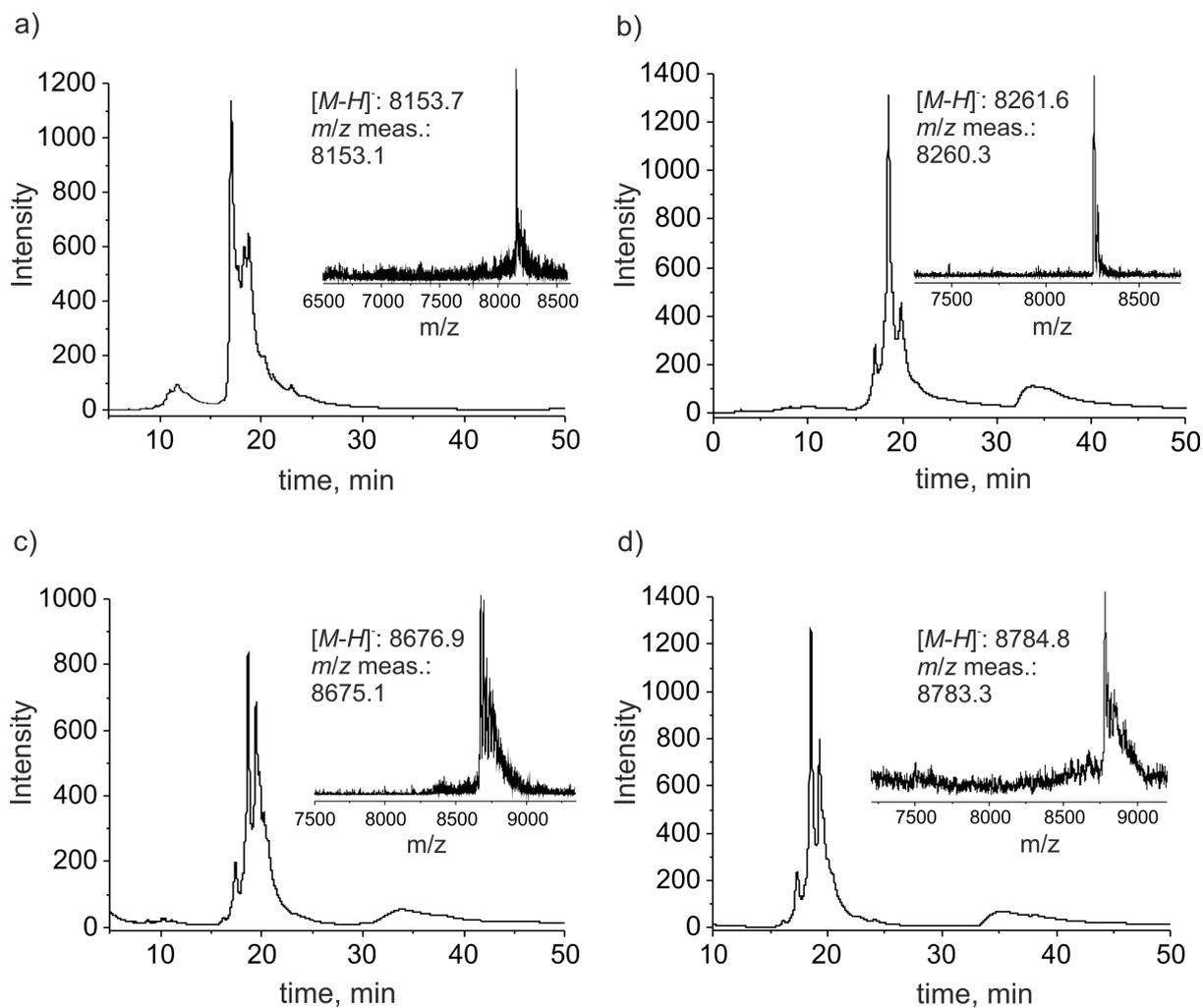
Figure 51. (a) HPL-chromatogram of unpurified **ON28** with the inset of MALDI-TOF mass spectrum of purified **ON28**; (b) HPL-chromatogram of unpurified **ON29** with the inset of MALDI-TOF mass spectrum of purified **ON29**; (c) HPL-chromatogram of unpurified **ON30** with the inset of MALDI-TOF mass spectrum of purified **ON30**; (d) HPL-chromatogram of unpurified **ON31** with the inset of MALDI-TOF mass spectrum of purified **ON31**.

Then the biological activity of the strands **ON28-ON31** together with single-glucose siRNAs **ON18** and **ON19** was tested in comparison with just 2'-OMe groups containing siRNA. Since it is known that transfection reagents are rather toxic for the cells, the investigation was performed without transfection reagents. Relative silencing of *Renilla* luciferase compared to the silencing of *Firefly* luciferase mediated by the glucose modified strands in A549 cells was measured. HBSS buffer or RPMI medium were used to deliver the glucose-siRNA conjugate into the cells. The data shown in Figure 52 left, right indicate, that all strands, even the unmodified duplex, showed a strong silencing effect. Potential explanation for that would be that the cells were under high stress and such conditions are not suitable for siRNA testing.
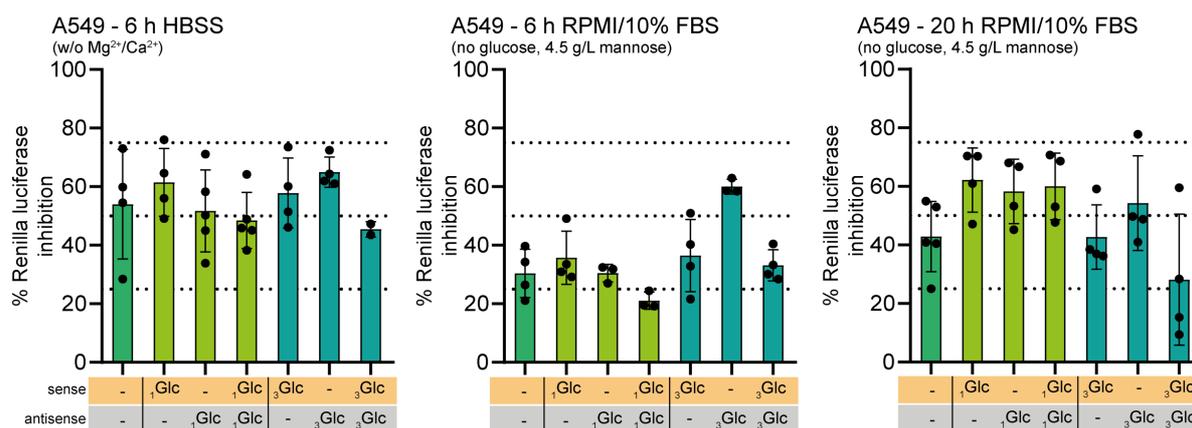
Figure 52. *Renilla* luciferase inhibition in cells mediated by siRNAs containing glucose (Glc) moieties. $_1$Glc-siRNA: modified via 6th position; $_3$Glc-siRNA: modified via anomeric position.

The experiments in RPMI medium (Figure 52, middle) showed more discriminating results. The positioning of the modification seems to be important. siRNA bearing one glucose on the sense strand is more potent than the glucose modified antisense strand. This is in contrast to siRNAs bearing multiple glucoses attached. In this case three glucoses on the antisense strand gives better results. Even ~60% knockdown is observed in the absence of transfection agents. Such increase in *Renilla* luciferase inhibition in comparison with unmodified siRNA could mean that sugars indeed improve the cellular uptake. In addition, they may also protect the siRNAs from degradation by nucleases. However, here presented biological data so far do not allow to draw a clear conclusion of whether glucose-siRNA conjugation can successfully replace the transfection agent and thus should be refined.

5.5 Conclusions and outlook

In an attempt to utilize modified RNA bases to create RNA based therapeutics, we could successfully identify siRNA sequence targeting the RNA-dependent RNA polymerase (Nsp12) and the spike protein of SARS-CoV-2. We then prepared C_m and U_m bearing siRNAs which contained in addition alkynyl moieties. This created siRNAs with greater stability. The ability to further functionalize the siRNAs via the alkyne groups allowed to obtain glucose- as well as lipid-siRNA conjugates (**ON18-ON23**). Biological assays proved the high potency of these bioconjugates. More than 90% knockdown was reached (with the use of transfectants).

Additionally, siRNAs (**ON28-ON31**) carrying multiple sugars were synthesized and their biological activity was investigated without the use of transfection reagents. At the same time, we evaluated the influence of the positioning of modification attached on siRNA.

6 Materials and methods

Chemicals were purchased from *Sigma-Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros Organics* and used without further purification. Reagent-grade dry solvents (*Sigma-Aldrich*, *Acros Organics*) were stored over molecular sieves and handled under inert gas atmosphere. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F254 TLC plates from *Merck KGaA*. Flash column chromatography was performed on *Geduran*® Si60 (40-63 μm) silica gel from *Merck KGaA*. NMR spectra were recorded on *Bruker AVIIIHD 400* spectrometers (400 MHz). ^{1}H NMR shifts were calibrated to the residual solvent resonances: DMSO-d_6 (2.50 ppm), CDCl$_3$ (7.26 ppm), Acetone-d_6 (2.05 ppm), CD$_2$Cl$_2$ (5.32 ppm). ^{13}C NMR shifts were calibrated to the residual solvent: DMSO-d_6 (39.52 ppm), CDCl$_3$ (77.16 ppm), Acetone-d_6 (29.84 ppm), CD$_2$Cl$_2$ (53.84 ppm). All NMR spectra were analysed using the program *MestreNova* 10.0.1 from *Mestrelab Research S. L.* High resolution electrospray ionization mass spectra (HRMS-ESI) were measured by the analytical section of the Department of Chemistry of the Ludwig-Maximilians-Universität München on a *Thermo Finnigan LTQ-FT GmbH*. IR spectra were recorded on a *PerkinElmer Spectrum BX II FT-IR* system. All substances were directly applied as solids on the ATR unit. Samples were filtered through *CHROMAFIL® PET 20/25* filters (0.20 μm pore size) from *Macherey-Nagel* prior to injection to HPLC systems. Analytical RP-HPLC was performed on an analytical HPLC *Waters Alliance* (2695 Separation Module, 2996 Photodiode Array Detector) or on an analytical HPLC *Agilent 1260 Infinity II* equipped with the column *Nucleosil 120-2 C18* from *Macherey Nagel* using a flow of 0.5 ml/min, a gradient of 0-30% of buffer B in 45 min was applied, unless otherwise specified. For peptide growth experiments: analytical RP-HPLC was performed on an analytical HPLC *Waters Alliance* (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column *Nucleosil 120-2 C18* from *Macherey Nagel* using a flow of 0.5 ml/min, a gradient of 0-15% of buffer B in 45 min was applied. Preparative RP-HPLC was performed on a HPLC *Waters Breeze* (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column *VP 250/32 C18* from *Macherey Nagel* using a flow of 5 ml/min, a gradient of 0-25% of buffer B in 45 min was applied for the purifications. For peptide growth experiments: preparative RP-HPLC was performed on a HPLC *Waters Breeze* (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column *VP 250/32 C18* from *Macherey Nagel* using a flow of 5 ml/min, a gradient of 0-15% of buffer B in 45 min was applied for the purifications. Oligonucleotides were purified using the following buffer system: buffer A: 100 mM

NEt$_3$/HOAc (pH 7.0) in H$_2$O and buffer B: 100 mM NEt$_3$/HOAc in 80% (v/v) acetonitrile. For peptide growth experiments: oligonucleotides were purified using the following buffer system: buffer A: 10 mM K$_2$HPO$_4$/KH$_2$PO$_4$ (pH 7.0) in H$_2$O and buffer B: 10 mM K$_2$HPO$_4$/KH$_2$PO$_4$ in MeCN/MeOH (1/7/2, v/v/v). The pH values of buffers were adjusted using a MP 220 pH-meter (*Metter Toledo*). All of the oligonucleotides used in this study were synthesized on a 1 μmol scale using a DNA automated synthesizer (*Applied Biosystems* 394 DNA/RNA Synthesizer) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleosides were purchased from *Glen Research* and *Sigma-Aldrich*. Oligonucleotides containing non-canonical nucleosides were synthesized in DMT-OFF mode using phosphoramidites (Bz-A, Dmf-G, Ac-C, U) with BTT in CH$_3$CN as an activator, DCA in CH$_2$Cl$_2$ as a deblocking solution and Ac$_2$O in pyridine/THF as a capping reagent. Oligonucleotides were detected at wavelength: 260 nm. Melting profiles were measured on a *JASCO V-650* spectrometer. Calculation of concentrations was assisted using the software *OligoAnalyzer 3.0* (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). For strands containing artificial bases, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a *Bruker Autoflex II*. For MALDI-TOF measurements, the samples were desalted on a 0.025 μm VSWP filter (*Millipore*) against ddH$_2$O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

Synthesis of the Phosphoramidite Building-Blocks

4-Nitrophenethyl (*tert*-butoxycarbonyl)-L-leucinate (11)
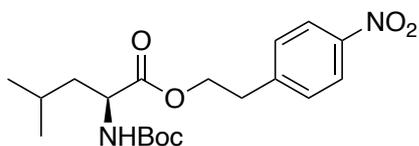
Boc-leucine (**10**, 1 g, 4.32 mmol, 1 eq.) was dissolved in CH$_2$Cl$_2$. Then npeOH (2 g, 11.9 mmol, 2.4 eq.), HBTU (2.13 g, 5.62 mmol, 1.3 eq.), DMAP (0.1 g, 0.86 mmol, 0.2 eq.) were added followed by addition of TEA (2.4 ml, 17.3 mmol, 4 eq.). The reaction mixture was left to stir overnight at room temperature. Then the solution was washed with water, organic phase was dried over Na$_2$SO$_4$ and concentrated *in vacuo*. The crude product was purified by flash chromatography eluting with Hex/EtOAc (2/1, v/v) to afford the target product.

Yield: 97%; **IR**: $\tilde{v}$ = 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**: 8.17 (d, J = 8.7 Hz, 2H), 7.39 (d, J = 8.7 Hz, 2H), 4.81 (d, J = 8.7 Hz, 1H), 4.38 (tt, J = 7.0, 3.4 Hz, 2H), 4.28 – 4.22 (m, 1H), 3.07 (t, J = 6.7 Hz, 2H), 1.58 (s, 2H), 1.43 (s, 9H), 0.88 (t, J = 7.0 Hz, 6H); ^{13}**C NMR (101 MHz, CDCl$_3$) δ**: 173.5, 155.5, 147.1, 145.5, 129.9, 123.9, 80.1, 64.6, 52.2, 41.8, 35.0, 28.4, 24.9, 22.9, 22.0; **HRMS (ESI)**: calculated for C$_{19}$H$_{29}$N$_2$O$_6^+$: m/z = 381.2020 [M+H]$^+$; found: m/z = 381.2015 [M+H]$^+$.

(S)-4-methyl-1-(4-nitrophenethoxy)-1-oxopentan-2-aminium chloride (12)
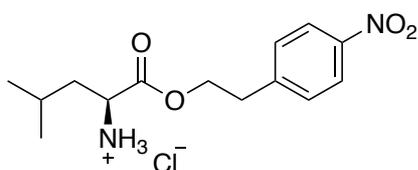
Npe-protected leucine **11** was dissolved in 4 M HCl/Dioxane mixture at 0 °C. The reaction mixture was left to stir for 2 h and afterwards was evaporated to dryness. The resulting product was used for further steps without additional purification.

Yield: 99%; **IR**: $\tilde{v}$ = 3663 (w), 2871 (s), 1737 (vs), 1589 (s), 1516 (vs), 1503 (vs), 1380 (vs), 1260 (w), 1207 (s), 1109 (w), 959 (w), 856 (s), 812 (s), 735 (vs) 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), 4.52 – 4.36 (m, 2H), 3.82 (t, J = 6.5 Hz, 1H), 3.11 (t, J = 6.5 Hz, 2H), 1.57 – 1.40 (m, 3H), 0.75 (t, J = 5.3 Hz, 6H); ^{13}**C NMR (101 MHz, DMSO-d_6) δ**: 169.8, 146.4, 130.4, 123.4, 65.3, 50.4, 33.8, 23.6, 22.2, 21.8; **HRMS (ESI)**: calculated for C$_{14}$H$_{21}$N$_2$O$_4^+$: m/z = 281.1496 [M+H]$^+$; found: m/z = 281.1495 [M+H]$^+$.

4-Nitrophenethyl ((9-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-9*H*-purin-6-yl)carbamoyl)-L-leucinate (19)
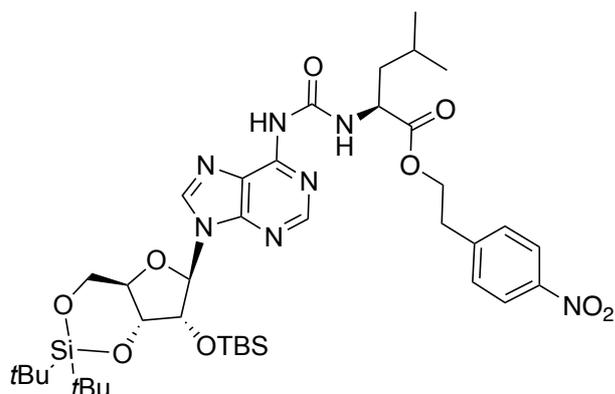
The 3',5'-silyl-protected adenosine derivative (**15**, 0.7 g, 1.34 mmol, 1 eq.) was dissolved in dry CH$_2$Cl$_2$ under nitrogen atmosphere. 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**13**, 0.64 g, 2.68 mmol, 2 eq.) was added to the reaction mixture and the resulting suspension was stirred at room temperature for 2 hours (the solution in time becomes clear). Afterwards the npe-protected leucine derivative (**12**, 0.85 g, 2.68 mmol, 2 eq.) was added together with TEA (750 µl, 2.68 mmol, 2 eq.) as a solution in CH$_2$Cl$_2$ and the resulting solution was stirred overnight at room temperature. The reaction was quenched by addition of saturated aqueous NaHCO$_3$ solution. The solution was extracted three times with CH$_2$Cl$_2$, and the organic phase was dried, filtered and concentrated *in vacuo*. The residue was purified by flash chromatography eluting with Hex/EtOAc (4/3, v/v) to give product as a white foam.

Yield: 82%; **IR**: $\tilde{v}$ = 3237 (w), 2168 (w), 1666 (s), 1572 (w), 1511 (s), 1429 (w), 1335 (s), 1271 (s), 1227 (s), 1178 (w), 1151 (w), 1119 (s), 1090 (vs), 1019 (w), 908 (s), 843 (s), 781 (s) cm^{-1};
^{1}H NMR (400 MHz, CDCl$_3$) δ: 9.87 (d, *J* = 7.7 Hz, 1H), 8.50 (s, 1H), 8.49 (s, 1H), 8.18 (s, 1H), 8.06 (d, *J* = 8.6 Hz, 2H), 7.38 (d, *J* = 8.6 Hz, 2H), 5.98 (s, 1H), 4.64 – 4.56 (m, 2H), 4.54 – 4.38 (m, 4H), 4.24 (td, *J* = 10.0, 5.1 Hz, 1H), 4.07 (t, *J* = 9.0 Hz, 1H), 3.09 (t, *J* = 6.5 Hz, 2H), 1.71 – 1.65 (m, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.97 – 0.91 (m, 15H), 0.18 (s, 3H), 0.16 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 172.9, 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$_7$O$_9$Si$_2$$^+$: m/z = 828.4142 [M+H]$^+$; found: m/z = 828.4149 [M+H]$^+$.

General procedure for N^6 methylation

The modified adenosine derivative (1 eq.) was dissolved in DMF. K_2CO_3 (3 eq.) was added at 0 °C and the solution was stirred for 5 min. Iodomethane (1.20 eq.) was added dropwise at 0 °C and the reaction mixture was stirred at room temperature for 1 h. Then the reaction mixture was diluted with EtOAc. The solution was then washed with water, dried over Na_2SO_4 and concentrated *in vacuo*. The crude product was purified by flash chromatography on silica gel (5% hexanes in EtOAc).

4-Nitrophenethyl *O*-(*tert*-butyldimethylsilyl)-*N*-((9-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)- L-threoninate (20)

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

Yield: 73%; **IR**: $\tilde{v}$ = 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**: 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, 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 (101 MHz, CDCl$_3$) δ**: 171.3, 156.4, 153.4, 151.6, 150.1, 146.8, 145.7, 139.4, 129.8, 123.7, 122.8, 92.5, 75.9, 75.6, 74.8, 68.9, 67.9, 64.6, 60.6, 34.9, 27.6, 27.1, 26.0, 25.7, 22.9, 21.3, 20.5, 18.5, 17.9, -4.15, -4.21, -4.8, -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]$^+$.

4-Nitrophenethyl ((9-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)- L-valinate (22)

Eluent: Hex/EtOAc (5/3, v/v).

Yield: 90%; **IR**: $\tilde{v}$ = 3239 (w), 2933 (w), 1741 (s), 1690 (s), 1569 (s), 1520 (vs), 1469 (s), 1345 (vs), 1259 (s), 1134 (s), 1057 (s), 1013 (s), 896 (w), 826 (vs), 780 (s), 750 (s) 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), 6.00 (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 (101 MHz, CDCl$_3$) δ**: 172.4, 156.0, 153.3, 151.6, 150.0, 146.9, 145.7, 139.3, 129.9, 126.4, 123.8, 122.7, 121.0, 92.5, 76.0, 75.6, 74.8, 67.9, 64.4, 59.8, 35.2, 34.7, 30.8, 27.6, 27.1, 26.0, 22.9, 20.5, 19.6, 18.4, 18.2, -4.1, -4.9; **HRMS (ESI)**: calculated for C$_{39}$H$_{62}$N$_7$O$_9$Si$_2$$^+$: m/z = 828.4142 [M+H]$^+$; found: m/z = 828.4132 [M+H]$^+$.
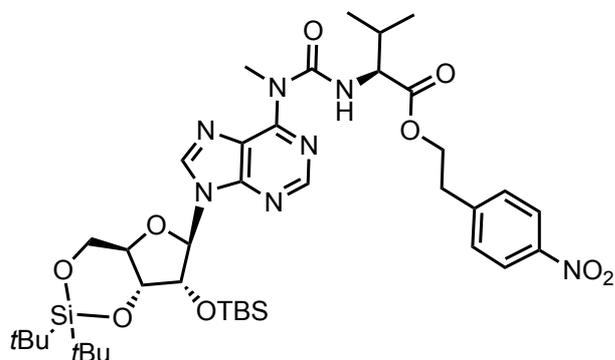
4-Nitrophenethyl ((9-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)glycinate (21)
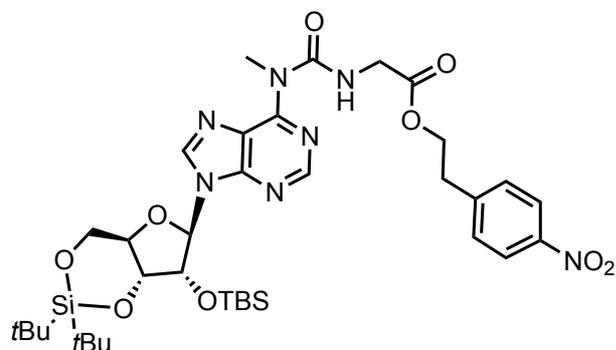
Eluent: Hex/EtOAc (5/3, v/v).

Yield: 76%; **IR**: $\tilde{v}$ = 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**: 10.98 (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* = 5.4, 1.7 Hz, 2H), 4.05 – 4.00 (m, 1H), 3.98 (s, 3H), 3.08 (t, *J* = 6.6 Hz, 2H), 1.07 (s, 9H), 1.04 (s, 9H) 0.94 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 170.4, 156.2, 153.2, 151.6, 150.2, 147.0, 145.5, 139.3, 129.9, 123.8, 122.7, 92.4, 76.0, 75.7, 74.8, 67.9, 64.6, 43.0, 34.8, 27.6, 27.1, 26.0, 22.9, 20.5, 18.5, -4.2, -4.9; **HRMS (ESI)**: calculated for C$_{36}$H$_{56}$N$_7$O$_9$Si$_2$$^+$: m/z = 786.3673 [M+H]$^+$; found: m/z = 786.3661 [M+H]$^+$.

4-Nitrophenethyl ((9-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)- L-leucinate (23)

Eluent: Hex/EtOAc (5/3, v/v).

Yield: 76%; **IR**: $\tilde{v}$ = 3230 (w), 2933 (w), 1740 (s), 1690 (s), 1580 (s), 1520 (vs), 1469 (s), 1345 (vs), 1259 (s), 1134 (s), 1057 (s), 1013 (s), 900 (w), 826 (vs), 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 – 4.38 (m, 5H), 4.25 (td, J = 10.1, 5.1 Hz, 1H), 4.06 – 3.99 (m, 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 (101 MHz, CDCl$_3$) δ**: 173.3, 155.6, 149.9, 145.6, 139.2, 129.8, 123.6, 92.40, 75.9, 75.5, 74.7, 67.8, 64.3, 52.9, 41.1, 34.9, 34.6, 27.5, 27.0, 25.9, 25.2, 22.9, 22.8, 22.0, 20.4, 18.3, -4.3, -5.0; **HRMS (ESI)**: calculated for C$_{40}$H$_{64}$N$_7$O$_9$Si$_2^+$: m/z = 842.4299 [M+H]$^+$; found: m/z = 842.4296 [M+H]$^+$.

General procedure for deprotection of 3',5'-silyl protecting group

The modified adenosine (0.86 mmol) was dissolved in CH$_2$Cl$_2$ under N$_2$ atmosphere and transferred into a plastic flask. Pyridine (1 ml) was added and the solution was cooled in an ice-bath. Then HF-Py (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$_2$Cl$_2$. Organic phase was washed with water and dried over Na$_2$SO$_4$. The solvents were removed *in vacuo*. The crude product was purified by flash chromatography eluting with CH$_2$Cl$_2$/MeOH (9/1, v/v) to afford the product as a colourless foam.

4-Nitrophenethyl *O*-(*tert*-butyldimethylsilyl)-*N*-((9-((2*R*,3*R*,4*R*,5*R*)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)- L-threoninate (24)

Eluent: CH$_2$Cl$_2$/MeOH (9/1, v/v).

Yield: 94%; **IR**: $\tilde{v}$ = 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**: 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), 1.09 – 1.03 (m, 1H), 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 (101 MHz, CDCl$_3$) δ**: 171.2, 156.1, 154.1, 151.2, 149.7, 147.0, 145.6, 141.5, 130.0, 123.8, 91.5, 87.7, 74.2, 72.8, 68.8, 65.0, 63.5, 60.6, 35.1, 25.6, 21.3, 18.0, 17.9, -4.1, -5.1, -5.2, -5.3; **HRMS (ESI)**: calculated for C$_{36}$H$_{58}$N$_7$O$_{10}$Si$_2$$^+$: m/z = 804.3778 [M+H]$^+$; found: m/z = 804.3768 [M+H]$^+$.
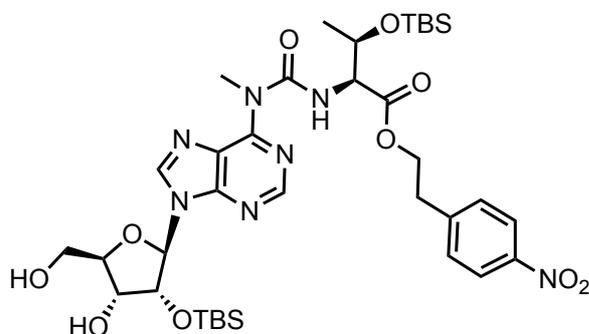
4-Nitrophenethyl ((9-((2*R*,3*R*,4*R*,5*R*)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-L-valinate (26)

Eluent: CH$_2$Cl$_2$/MeOH (9/1, v/v).

Yield: 95%; **IR**: $\tilde{v}$ = 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**: 11.03 (d, J = 7.4 Hz, 1H), 8.53 (s, 1H), 8.14 (d, J = 8.6 Hz, 2H), 7.95 (s, 1H), 7.40 (d, J = 8.6 Hz, 2H), 5.82 (d, J = 7.4 Hz, 1H), 5.14 (dd, J = 7.3, 4.8 Hz, 1H), 4.56 (dd, J = 7.4, 4.8 Hz, 1H), 4.51 – 4.39 (m, 3H), 4.39 – 4.35 (m, 2H), 3.99 (s, 3H), 3.93 (dd, J = 13.0, 1.8 Hz, 1H), 3.77 (dd, J = 13.0, 1.8 Hz, 1H), 3.10 (t, J = 6.7 Hz, 2H), 2.23 – 2.17 (m, 1H), 1.09 – 1.03 (m, 1H), 0.98 (s, J = 6.9 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 (101 MHz, CDCl$_3$) δ**: 172.4, 155.7, 154.0, 151.2, 149.6, 147.0, 145.6, 141.5, 130.0, 123.8, 121.0, 91.4, 87.7, 74.2, 72.8, 64.6, 63.4, 59.8,

35.0, 34.9, 30.8, 25.6, 19.6, 18.0, 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 = 688.3117 [M+H]$^+$.

4-Nitrophenethyl ((9-((2R,3R,4R,5R)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)glycinate (25)

Eluent: CH_2Cl_2/MeOH (9/1, v/v).

Yield: 97%; **IR**: $\tilde{v}$ = 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**: 11.03 (d, *J* = 7.4 Hz, 1H), 8.53 (s, 1H), 8.14 (d, *J* = 8.6 Hz, 2H), 7.95 (s, 1H), 7.40 (d, *J* = 8.6 Hz, 2H), 5.82 (d, *J* = 7.4 Hz, 1H), 5.14 (dd, *J* = 7.3, 4.8 Hz, 1H), 4.56 (dd, *J* = 7.4, 4.8 Hz, 1H), 4.51 – 4.39 (m, 3H), 4.39 – 4.35 (m, 2H), 3.99 (s, 3H), 3.93 (dd, *J* = 13.0, 1.8 Hz, 1H), 3.77 (dd, *J* = 13.0, 1.8 Hz, 1H), 3.10 (t, *J* = 6.7 Hz, 2H), 2.23 – 2.17 (m, 1H), 1.09 – 1.03 (m, 1H), 0.98 (s, *J* = 6.9 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 (101 MHz, CDCl$_3$) δ**: 172.4, 155.7, 154.0, 151.2, 149.6, 147.0, 145.6, 141.5, 130.0, 123.8, 121.0, 91.4, 87.7, 74.2, 72.8, 64.6, 63.4, 59.8, 35.0, 34.9, 30.8, 25.6, 19.6, 18.0, 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 = 646.2644 [M+H]$^+$.

4-Nitrophenethyl ((9-((2*R*,3*R*,4*R*,5*R*)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-L-leucinate (27)
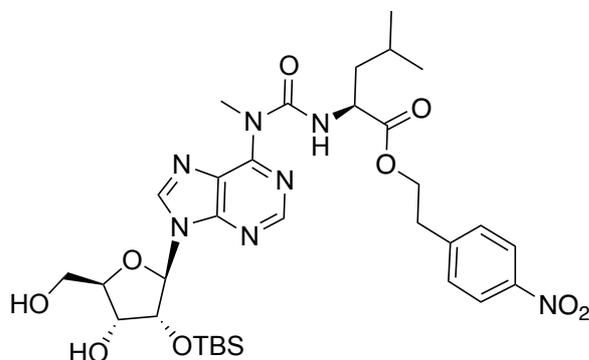
Eluent: CH$_2$Cl$_2$/MeOH (9/1, v/v).

Yield: 98%; **IR**: $\tilde{v}$ = 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, 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 (q, *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 (101 MHz, Acetone-*d$_6$*) δ**: 173.4, 155.9, 154.0, 152.7, 150.5, 147.5, 142.5, 131.1, 124.1, 90.5, 87.5, 76.6, 72.4, 65.1, 62.8, 53.7, 41.7, 35.3, 34.9, 26.0, 25.8, 23.1, 22.1, 18.0, -4.9, -5.1; **HRMS (ESI)**: calculated for C$_{32}$H$_{47}$N$_7$O$_9$Si$^+$: m/z = 702.3277 [M+H]$^+$; found: m/z = 702.3279 [M+H]$^+$.

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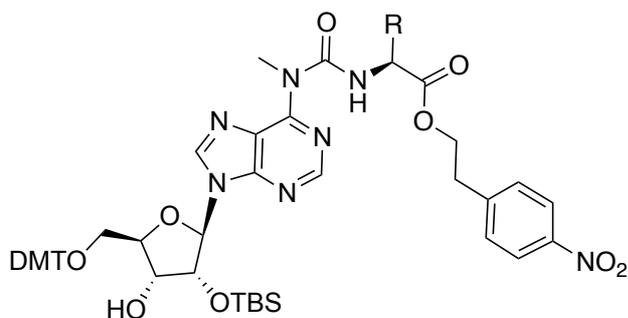
The 3',5'-deprotected adenosine derivative (1 eq.) was dissolved in pyridine under N_2 atmosphere. DMT chloride (1.5 eq.) was added in two portions (first half of the portion was added and after 10 min the second half was added) and the mixture was stirred at room temperature overnight. Then the volatiles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

4-Nitrophenethyl *N*-((9-((2*R*,3*R*,4*R*,5*R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxytetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-*O*-(*tert*-butyldimethylsilyl)-L-threoninate (28)
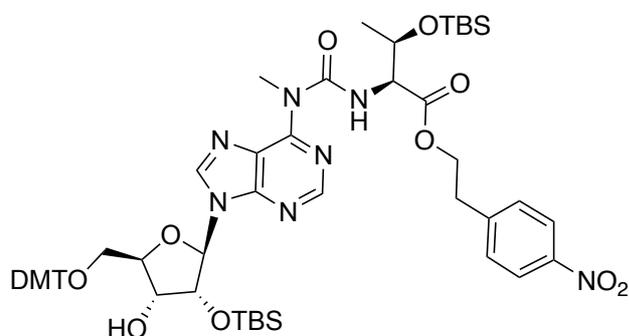
Yield: 68%; **IR**: $\tilde{v}$ = 2908 (w), 1757 (w), 1718 (w), 1670 (w), 1608 (w), 1507 (vs), 1441 (w), 1294 (w), 1248 (vs), 1177 (s), 1090 (s), 1034 (vs), 975 (s), 913 (s), 869 (s), 776 (s), 703 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-*d*$_6$) δ**: 10.89 (d, *J* = 8.7 Hz, 1H), 8.50 (s, 1H), 8.42 (s, 1H), 8.03 (d, *J* = 8.7 Hz, 2H), 7.52 (d, *J* = 8.8 Hz, 4H), 7.40 – 7.30 (m, 7H), 6.85 (dd, *J* = 8.8, 1.8, 4H), 6.20 (d, *J* = 4.5 Hz, 1H), 5.11 (t, *J* = 4.7 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 (101 MHz, Acetone-*d*$_6$) δ**: 171.7, 159.6, 153.1, 150.6, 150.4, 142.0, 136.7, 131.0, 131.0, 129.0, 128.6, 127.6, 124.6, 124.1, 113.8, 90.0, 87.1, 84.8, 76.4, 71.9, 69.7, 65.5, 64.4, 61.2, 55.5, 35.2, 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]$^+$.

4-Nitrophenethyl ((9-((2R,3R,4R,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxytetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-L-valinate (30)
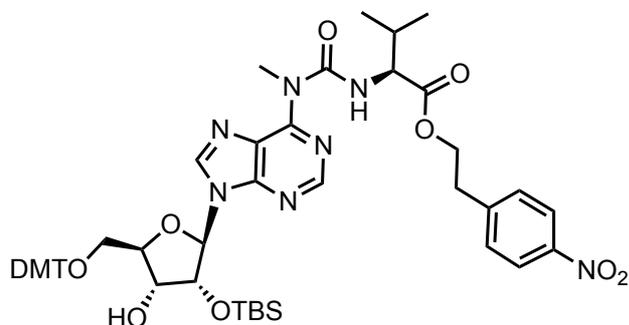
Yield: 75%; **IR**: $\tilde{v}$ = 2950 (w), 2850 (w), 1730 (w), 1670 (w), 1607 (s), 1577 (s), 1508 (vs), 1464 (w), 1347 (s), 1250 (vs), 1177 (s), 1150 (w), 1090 (s), 1035 (s), 981 (w), 913 (s), 866 (s), 839 (s), 701 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-*d*$_6$) δ**: 11.01 (d, *J* = 7.7 Hz, 1H), 8.52 (s, 1H), 8.49 (s, 1H), 8.10 (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 (t, *J* = 7.5 Hz, 3H), 6.86 (dd, *J* = 9.0, 2.7 Hz, 4H), 6.19 (d, *J* = 4.3 Hz, 1H), 5.05 (t, *J* = 4.6 Hz, 1H), 4.54 – 4.41 (m, 3H), 4.39 – 4.35 (m, 1H), 4.30 – 4.28 (m, 1H), 3.92 (s, 3H), 3.78 (s, 6H), 3.48 – 3.44 (m, 2H), 3.15 (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), 0.86 (s, 9H), 0.07 (s, 3H), -0.02 (s, 3H); **^{13}C NMR (101 MHz, Acetone-*d*$_6$) δ**: 172.4, 159.6, 150.6, 141.8, 136.7, 131.0, 130.0, 129.0, 128.6, 127.6, 124.1, 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 = 990.4415 [M+H]$^+$.

4-Nitrophenethyl ((9-((2R,3R,4R,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxytetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)glycinate (29)
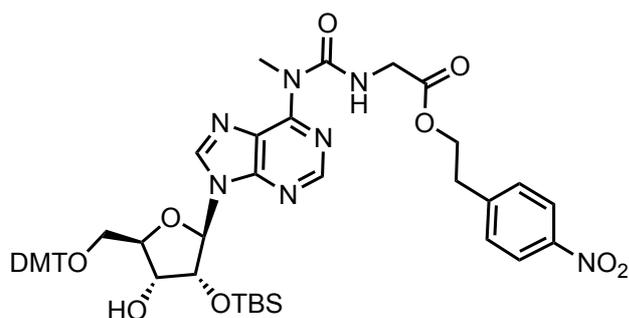
Yield: 86%; **IR**: $\tilde{v}$ = 3320 (w), 2912 (w), 2856 (w), 1755 (w), 1663 (s), 1607 (w), 1520 (vs), 1443 (w), 1340 (w), 1294 (w), 1248 (vs), 1176 (vs), 1090 (s), 1033 (vs), 981 (s), 914 (s), 867 (s), 701 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.86 (t, J = 5.6 Hz, 1H), 8.49 (s, 1H), 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 – 7.22 (m, 3H), 6.86 (dd, J = 9.0, 2.3 Hz, 4H), 6.18 (d, J = 4.4 Hz, 1H), 5.06 (t, J = 4.7 Hz, 1H), 4.52 (q, J = 5.2 Hz, 1H), 4.43 (t, J = 6.4 Hz, 2H), 4.31 – 4.26 (m, 1H), 4.11 (d, J = 5.8 Hz, 2H), 4.00 (d, J = 5.8 Hz, 1H), 3.93 (s, 3H), 3.77 (s, 6H), 3.47 (t, J = 4.4 Hz, 2H), 3.13 (t, J = 6.4 Hz, 2H), 0.85 (s, 9H), 0.06 (s, 3H), -0.04 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 170.6, 159.6, 156.5, 153.7, 153.1, 150.7, 147.4, 146.1, 141.8, 136.7, 131.1, 129.0, 128.6, 127.6, 124.2, 113.9, 90.0, 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$_7$O$_{11}$Si$^+$: m/z = 948.3958 [M+H]$^+$; found: m/z = 948.3953 [M+H]$^+$.

4-Nitrophenethyl ((9-((2*R*,3*R*,4*R*,5*R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxytetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-L-leucinate (31)
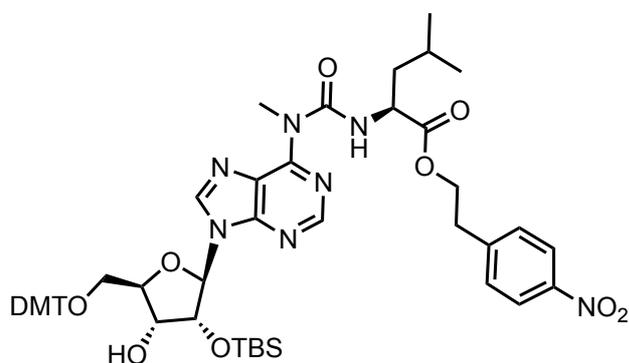
Yield: 72%; **IR**: $\tilde{v}$ = 2950 (w), 2852 (w), 1729 (w), 1670 (w), 1607 (s), 1577 (s), 1508 (vs), 1464 (w), 1347 (s), 1250 (vs), 1177 (s), 1152 (w), 1091 (s), 1035 (s), 981 (w), 913 (s), 866 (s), 839 (s), 699 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.89 (d, J = 7.1 Hz, 1H), 8.50 (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 – 7.20 (m, 3H), 6.86 (dd, J = 9.0, 2.5 Hz, 4H), 6.18 (d, J = 4.3 Hz, 1H), 5.05 (t, J = 4.6 Hz, 1H), 4.53 – 4.46 (m, 1H), 4.38 – 4.45 (m, 3H), 4.29 (q, J = 4.6 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 (101 MHz, Acetone-d_6) δ**: 173.4, 159.6, 156.0, 153.8, 153.1, 150.6, 147.5, 146.1, 141.8, 136.7, 131.1, 129.1, 128.9, 128.6, 127.6, 124.1, 113.9,

90.0, 87.1, 84.7, 76.5, 71.9, 65.1, 64.3, 55.5, 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 $[M+H]^+$.

General procedure for phosphoramidite formation
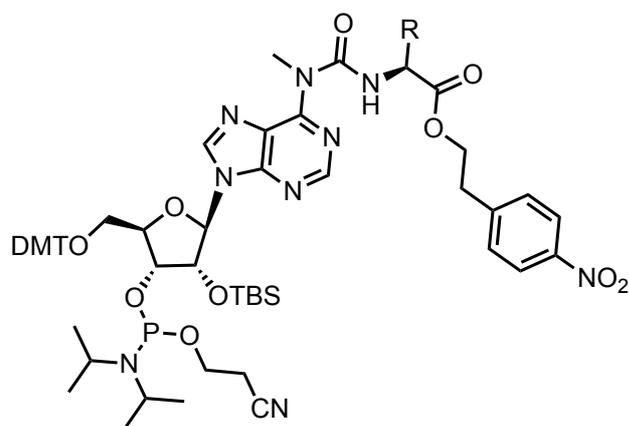
In an oven-dried flask under argon atmosphere, 5'-DMT protected compound (1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 °C. Hünig's base (4 eq.) was added dropwise followed by the addition of 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (2.5 eq.). The solution was stirred at room temperature for 2.5 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 (2/1, v/v) containing 0.1% of pyridine. After the lyophilization from benzene the phosphoramidite was obtained as a mixture of diastereomers as white foam.

4-Nitrophenethyl *N*-((9-((2*R*,3*R*,4*R*,5*R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-*O*-(*tert*-butyldimethylsilyl)-L-threoninate (32)
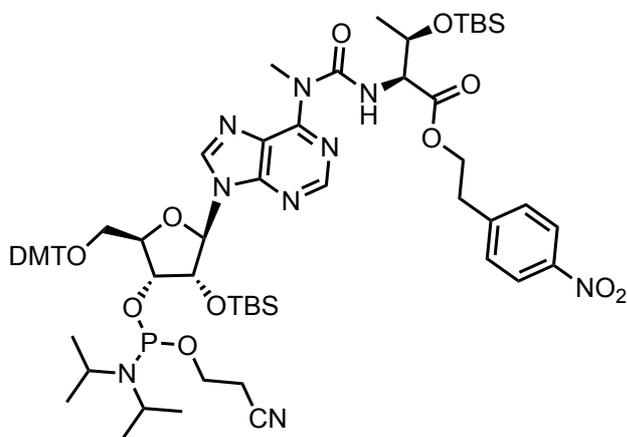
Yield: 62%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.18, 148.41; **HRMS (ESI)**: calculated for $C_{66}H_{93}N_9O_{13}PSi_2^+$: m/z = 1306.6164 [M+H]$^+$; found: m/z = 1306.6189 [M+H]$^+$.

4-Nitrophenethyl ((9-((2*R*,3*R*,4*R*,5*R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-L-valinate (34)
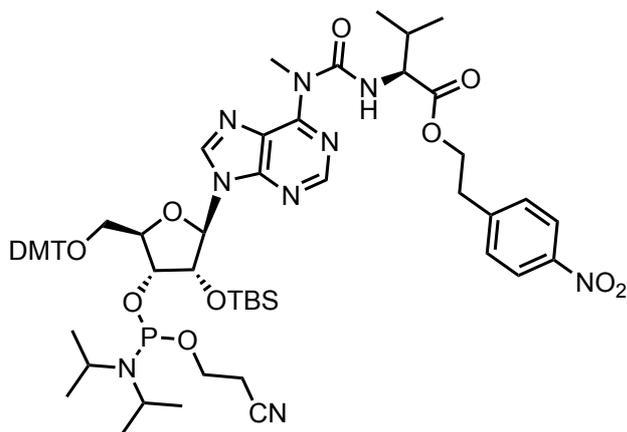
Yield: 67%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.09, 148.68; **HRMS (ESI)**: calculated for $C_{61}H_{81}N_9O_{12}PSi^+$: m/z = 1190.5506 [M+H]$^+$; found: m/z = 1190.5492 [M+H]$^+$.

4-Nitrophenethyl ((9-((2R,3R,4R,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)glycinate (33)
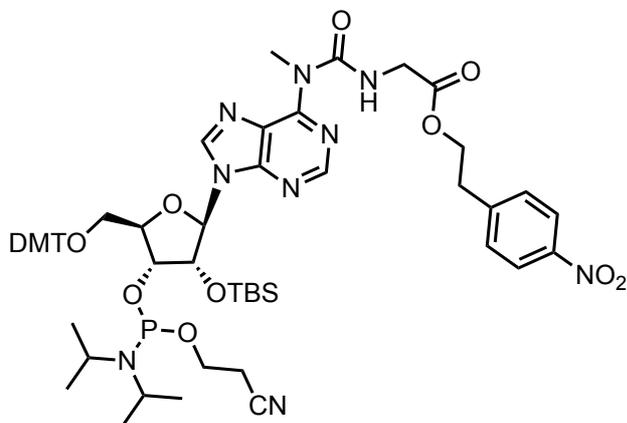
Yield: 78%; **³¹P NMR (162 MHz, Acetone-*d*₆) δ**: 150.21, 148.62; **HRMS (ESI)**: calculated for $C_{61}H_{81}N_9O_{12}PSi^+$: m/z = 1148.5037 [M+H]⁺; found: m/z = 1148.5052 [M+H]⁺.

4-Nitrophenethyl ((9-((2R,3R,4R,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)tetrahydrofuran-2-yl)-9*H*-purin-6-yl)(methyl)carbamoyl)-L-leucinate (35)
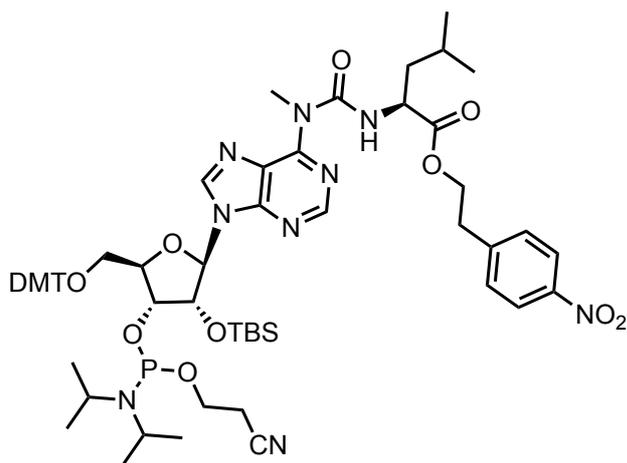
Yield: 75%; **³¹P NMR (162 MHz, Acetone-*d*₆) δ**: 150.14, 148.68; **HRMS (ESI)**: calculated for $C_{62}H_{83}N_9O_{12}PSi^+$: m/z = 1204.5663 [M+H]⁺; found: m/z = 1204.5682 [M+H]⁺.

1-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-5-methylpyrimidine-2,4(1*H*,3*H*)-dione (37)[230]
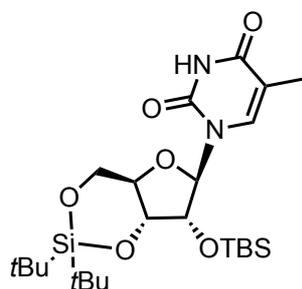
The reaction was performed according to the procedure published before.[230]

5-Methyluridine (**36**, 1.35 g, 5.16 mmol, 1 eq.) was suspended in DMF and di-*tert*-butylsilyl ditriflate (1.85 ml, 5.68 mmol, 1.1 eq.) was added dropwise under stirring at 0 °C. The resulting solution was stirred at 0 °C for 45 min. Then imidazole (1.8 g, 26.4 mmol, 5 eq.) was added and the reaction was warmed to room temperature over a period of 30 min. Then TBSCl (1 g, 6.64 mmol, 1.3 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, 5/1, v/v).

Yield: 87%; **IR**: $\tilde{v}$ = 3663 (w), 2932 (w), 2859 (w), 1691 (s), 1610 (s), 1513 (w), 1469 (s), 1387 (w), 1251 (w), 1118 (w), 1060 (vs), 999 (s), 835 (vs), 790 (vs) cm^{-1}; **¹H NMR (400 MHz, CDCl₃) δ**: 9.11 (s, 1H), 7.01 (d, *J* = 1.3 Hz, 1H), 5.65 (s, 1H), 4.49 (dd, *J* = 9.1, 5.0 Hz, 1H), 4.29 (d, *J* = 5.0 Hz, 1H), 4.12 (ddd, *J* = 10.4, 9.1, 5.0 Hz, 1H), 4.01 – 3.89 (m, 2H), 1.92 (d, *J* = 1.2 Hz, 3H), 1.05 (s, 9H), 1.02 (s, 9H), 0.92 (s, 9H), 0.16 (s, 3H), 0.13 (s, 3H); **¹³C NMR (101 MHz, CDCl₃) δ**: 163.9, 149.9, 135.6, 111.1, 92.3, 76.2, 75.3, 74.5, 67.7, 27.6, 27.1, 25.9, 22.9, 20.5, 18.4, 12.8, -4.2, -4.9; **HRMS (ESI)**: calculated for $C_{24}H_{45}N_2O_6Si_2^+$: m/z = 513.2811 [M+H]$^+$; found: m/z = 513.2794 [M+H]$^+$.

The analytical data is in agreement with the literature.[230]

1-((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-5-(hydroxymethyl)pyrimidine-2,4(1*H*,3*H*)-dione (39)[219]
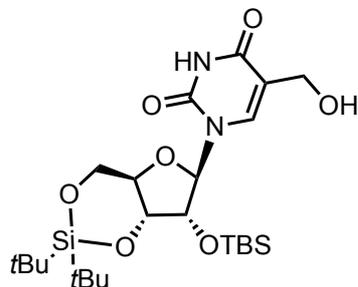
The reaction was performed according to the modified procedure published before.[219]

The silyl-protected 5-methyluridine (**37**, 0.5 g, 0.98 mmol, 1 eq.) was dissolved in CCl$_4$. To the solution NBS (0.26 g, 1.45 mmol, 1.5 eq.) and AIBN (0.019 g, 0.12 mmol, 0.12 eq.) were added. The reaction mixture was heated at 80 °C under reflux for 2 hours until a floating solid could be observed in the flask. The solvents were removed *in vacuo* to afford the brominated intermediate **38**. Then a mixture of saturated NaHCO$_3$ and THF (2/1, v/v) were added to the compound **38**. The reaction mixture was left to stir at room temperature overnight. Afterwards the organic phase was extracted with EtOAc, washed with water, brine and dried over Na$_2$SO$_4$. The solvents were removed *in vacuo*. The crude product was purified by flash chromatography eluting first with Hex/EtOAc (3/1, v/v), later with Hex/EtOAc (1/1, v/v) to afford the 5-methyluridine derivative as a light brown foam.

Yield: 60%; **IR**: $\tilde{v}$ = 3461 (w), 2958 (w), 1725 (vs), 1699 (vs), 1679 (vs), 1518 (w), 1462 (s), 1286 (s), 1163 (s), 1129 (s), 1094 (s), 1075 (w), 1032 (s), 918 (w), 858 (vs), 840 (vs), 760 (vs), 678 (w) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 9.02 (s, 1H), 7.28 (s, 1H), 5.70 (s, 1H), 4.50 (dd, *J* = 9.2, 5.1 Hz, 1H), 4.47 – 4.35 (m, 2H), 4.28 (d, *J* = 4.7 Hz, 1H), 4.20 – 4.09 (m, 2H), 4.03 – 3.93 (m, 1H), 3.89 (dd, *J* = 9.6, 4.7 Hz, 1H), 1.05 (s, 9H), 1.02 (s, 9H), 0.93 (s, 9H), 0.17 (s, 3H), 0.13 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 163.4, 149.6, 136.8, 113.9, 94.0, 76.2, 75.5, 74.6, 67.7, 58.8, 27.6, 27.4, 27.1, 22.9, 20.5, 18.4, -4.2, -4.9; **HRMS (ESI)**: calculated for C$_{24}$H$_{45}$N$_2$O$_7$Si$_2^+$: m/z = 529.2765 [M+H]$^+$; found: m/z = 529.2759 [M+H]$^+$.

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

(1-(((4a*R*,6*R*,7*R*,7a*R*)-2,2-di-*tert*-butyl-7-((*tert*-butyldimethylsilyl)oxy)tetrahydro-4*H*-furo[3,2-*d*][1,3,2]dioxasilin-6-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)methyl pivalate (40)
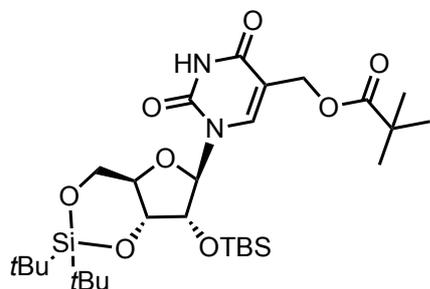
Compound **39** (0.24 g, 0.45 mmol, 1 eq.) was dissolved in pyridine and cooled to 0 °C. Then DMAP (0.011 g, 0.09 mmol, 0.2 eq.) and pivaloyl chloride (0.071 g, 0.59 mmol, 1.3 eq.) were slowly added to the reaction mixture. The ice-bath was removed, the mixture was allowed to warm to room temperature and stirred overnight. Afterwards the reaction was quenched by the addition of saturated NaHCO$_3$ solution. The organic phase was extracted with EtOAc, washed with water and dried over Na$_2$SO$_4$. The solvents were removed *in vacuo*. The crude product was purified by flash chromatography eluting with Hex/EtOAc (3/1, v/v) to afford the pivaloylated 5-methyluridine derivative **40**.

Yield: 95%; **IR**: $\tilde{v}$ = 3466 (w), 2955 (w), 1725 (vs), 1699 (vs), 1679 (vs), 1518 (w), 1464 (s), 1286 (s), 1163 (s), 1130 (s), 1094 (s), 1075 (w), 1030 (s), 918 (w), 858 (vs), 840 (vs), 760 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 8.95 (s, 1H), 7.52 (s, 1H), 5.66 (s, 1H), 4.90 – 4.78 (m, 2H), 4.51 (dd, *J* = 9.1, 5.0 Hz, 1H), 4.29 (d, *J* = 4.6 Hz, 1H), 4.24 – 4.13 (m, 1H), 4.09 – 4.00 (m, 1H), 3.86 (dd, *J* = 9.6, 4.6 Hz, 1H), 1.18 (s, 9H), 1.05 (s, 9H), 1.02 (s, 9H), 0.93 (s, 9H), 0.18 (s, 3H), 0.14 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 178.6, 162.4, 149.5, 140.0, 109.6, 94.0, 76.1, 75.5, 74.8, 67.6, 58.7, 39.0, 27.6, 27.3, 27.1, 26.0, 22.8, 20.5, 18.4, -4.2, -4.9; **HRMS (ESI)**: calculated for C$_{29}$H$_{53}$N$_2$O$_8$Si$_2^+$: m/z = 613.3340 [M+H]$^+$; found: m/z = 613.3320 [M+H]$^+$.

(1-((2R,3R,4R,5R)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)methyl pivalate (41)
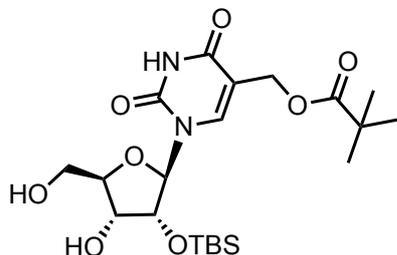
The modified 5-methyluridine (**40**, 0.18 g, 0.29 mmol, 1 eq.) was dissolved in CH_2Cl_2 under nitrogen atmosphere and transferred into a plastic flask. Pyridine (0.5 ml) was added, and the solution was cooled in an ice-bath. Then HF-Py (50 µl) was added, and the mixture was stirred at 0 °C for 2 hours. The reaction was quenched by the addition of saturated $NaHCO_3$ solution 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 (10/1, v/v) to afford the target compound as a colorless foam.

Yield: 98%; **IR**: $\tilde{v}$ = 3466 (w), 2955 (w), 1725 (vs), 1699 (vs), 1679 (vs), 1642 (w), 1518 (w), 1464 (s), 1286 (s), 1262 (w), 1163 (s), 1130 (s), 1094 (s), 1075 (w), 1030 (s), 918 (w), 858 (vs), 840 (vs), 760 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.23 (s, 1H), 8.27 (s, 1H), 5.96 (d, J = 5.1 Hz, 1H), 4.76 (s, 2H), 4.49 – 4.39 (m, 2H), 4.26 – 4.18 (m, 1H), 4.10 – 4.03 (m, 1H), 3.92 – 3.84 (m, 1H), 3.81 (d, J = 4.7 Hz, 2H), 1.15 (s, 9H), 0.90 (s, 9H), 0.11 (s, 3H), 0.09 (s, 3H); **^{13}C NMR (100 MHz, Acetone-d_6) δ**: 178.3, 162.8, 151.4, 141.8, 110.2, 89.6, 86.1, 78.6, 72.2, 62.2, 59.9, 39.2, 27.4, 26.1, 18.7, -4.70, -4.72; **HRMS (ESI)**: calculated for $C_{21}H_{36}N_2O_8Si^+$: m/z = 473.2319 [M+H]$^+$; found: m/z = 473.2311 [M+H]$^+$.

(1-((2*R*,3*R*,4*R*,5*R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-hydroxytetrahydrofuran-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)methyl pivalate (42)[231]
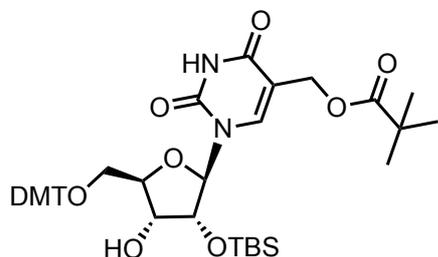
The 3',5'-deprotected adenosine derivative (**41**, 0.53 g, 1.12 mmol, 1 eq.) was dissolved in pyridine under N_2 atmosphere. DMT chloride (0.49 g, 1.68 mmol, 1.5 eq.) was added in two portions (first half of the portion was added and after 10 min the other half was added) and the mixture was stirred at room temperature overnight. Then the volatiles were evaporated, and crude product was purified by flash chromatography eluting with CH_2Cl_2/MeOH (10/1, v/v) with an addition of 0.1% of pyridine to afford the DMT protected derivative as white foam.

Yield: 86%; **IR**: $\tilde{v}$ = 3639 (w), 2944 (w), 2929 (w), 2856 (w), 1738 (w), 1689 (s), 1610 (s), 1580 (s), 1512 (vs), 1462 (s), 1344 (vs), 1250 (vs), 1177 (s), 1108 (w), 1014 (s), 913 (w), 830 (vs), 780 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-*d*$_6$) δ**: 10.26 (s, 1H), 7.89 (s, 1H), 7.50 (d, *J* = 8.7 Hz, 2H), 7.37 (d, *J* = 8.7 Hz, 4H), 7.32 (d, *J* = 7.9 Hz, 2H), 7.29 – 7.21 (m, 1H), 6.90 (d, *J* = 7.6 Hz, 4H), 5.98 (d, *J* = 4.7 Hz, 1H), 4.50 (t, *J* = 4.9 Hz, 1H), 4.38 (d, *J* = 11.9 Hz, 1H), 4.31 (q, *J* = 4.9 Hz, 1H), 4.24 – 4.14 (m, 2H), 3.93 (d, *J* = 5.6 Hz, 1H), 3.79 (s, 6H), 3.47 (dd, *J* = 10.8, 4.0 Hz, 1H), 3.40 (dd, *J* = 10.8, 2.7 Hz, 1H), 1.10 (s, 9H), 0.93 (s, 9H), 0.16 (s, 3H), 0.15 (s, 3H); **^{13}C NMR (101 MHz, Acetone-*d*$_6$) δ**: 177.9, 162.7, 159.7, 151.2, 145.8, 141.2, 136.6, 136.4, 130.9, 129.0, 128.8, 127.8, 114.0, 89.6, 87.5, 84.4, 76.7, 71.7, 64.3, 59.9, 55.5, 39.2, 27.4, 26.2, 18.7, -4.6; **HRMS (ESI)**: calculated for $C_{42}H_{53}N_2O_{10}Si^+$: m/z = 773.3469 [M-H]$^-$; found: m/z = 773.3478 [M-H]$^-$.

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

(1-((2R,3R,4R,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((*tert*-butyldimethylsilyl)oxy)-4-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)tetrahydrofuran-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)methyl pivalate (43)[231]
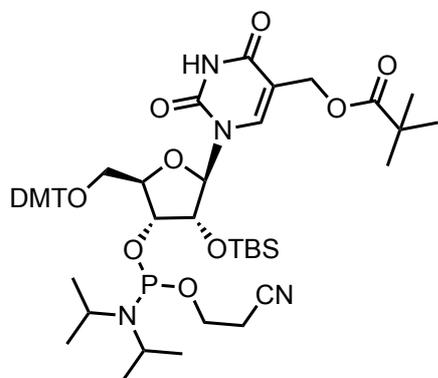
Yield: 75%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.14, 148.94; **HRMS (ESI)**: calculated for $C_{51}H_{72}N_4O_{11}PSi^+$: m/z = 975.4699 [M+H]$^+$; found: m/z = 975.4716 [M+H]$^+$.

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

General protocol for deprotection and purification of m^{6}aa^{6}A-strands 5'-m^{6}aa^{6}AAUCGCU-3'

RNA synthesis was performed on a 1 μmol scale using standard RNA synthesis protocol. The cartridge containing the RNA strand was dried under high vacuum for an hour. The solid support was suspended in DBU solution (1% in THF, 1 ml), vortexed and incubated for 2 h at room temperature. After the supernatant had been removed and the beads had been washed with THF, a mixture of aqueous ammonia and methylamine (1/1, 1 ml) was added, and the suspension was incubated for 15 min at 65 °C. Subsequently the supernatant was removed, and the beads were washed with water. The supernatant and washings were combined, and the solvents were removed by speed vacuum overnight. RNA grade DMSO (50 μl) was added to the dry residue followed by triethylamine trihydrofluoride (NEt$_3$·3HF 98%, 125 μl). The mixture was incubated at 60 °C for 1 h and subsequently cooled in an ice bath. The RNA was precipitated by adding aqueous NaOAc solution (3 M, 25 μL) and *n*-BuOH (1 ml). To ensure complete precipitation the sample was incubated at -80 °C for 1 h. After centrifugation (4 °C, 4000 rpm, 15 min), the supernatant was removed, and the precipitated RNA was lyophilized. The RNA was then redissolved in water and filtered. The crude RNA was purified by reversed-phase HPLC, the structural integrity was analyzed by MALDI-TOF.

General protocol for deprotection and purification of mnm⁵U-modified RNA positioned at the 3'-terminal position using the *Universal Support III*

After RNA synthesis the solid support carrying the *pivom⁵U*-modified RNA was transferred to an *Eppendorf* tube. A solution of 8 M MeNH₂ in absolute ethanol (0.5 ml) was added and incubated at rt for 2 h. Afterwards, the supernatant was transferred to another *Eppendorf* tube, the residual solid support was washed additionally with ddH₂O/EtOH mixture (3:1) three times. The combined supernatants were dried by speed vacuum. The substitution of pivaloyl to methylamino group was performed by treating the RNA with solution of 8 M MeNH₂ in absolute ethanol (0.5 ml) at 60 °C for 1 h. Then the reaction mixture was concentrated *in vacuo*. The 2′-OTBS protecting groups were deprotected by adding DMSO (100 µl) and triethylamine trihydrofluoride (NEt₃·3HF 98%, 125 µl), the reaction tube was shaken at 65 °C for 1 h 30 min. Then the solution was cooled to 0 °C. Aqueous NaOAc (3 M, 25µl) and *n*-BuOH (1 ml) were added. The suspension was vortexed and left at -80 °C for 1 h. After centrifugation (4000 rpm, 4 °C, 15 min), the supernatant was carefully removed. The pellet was frozen with liquid nitrogen and lyophilized. The crude RNA was purified by reversed-phase HPLC, the structural integrity was analyzed by MALDI-TOF.

General protocol for deprotection and purification of nm⁵U-modified RNA positioned at the 3'-terminal position using the *Universal Support III*

After RNA synthesis the solid support carrying the *pivom⁵U*-modified RNA was transferred to an *Eppendorf* tube. A solution of NH₃/MeOH (0.5 ml) was added and incubated at rt for 2 h. After this, the supernatant was transferred to another *Eppendorf* tube, the residual solid support was washed additionally with ddH₂O/EtOH mixture (3:1) three times. The combined supernatants were dried by speed vacuum. The substitution of pivaloyl to amino group was performed by treating the RNA with a solution of 30% aq. NH₃ (0.5 ml). The mixture was incubated at 60 °C for 1 h. After this, the supernatant was transferred to another *Eppendorf* tube, the residual solid support was washed additionally with ddH₂O/EtOH mixture (3:1) three times. The combined supernatants were dried by speed vacuum. The 2′-OTBS protecting groups were deprotected by adding DMSO (100 µl) and triethylamine trihydrofluoride (NEt₃·3HF 98%, 125 µl), the reaction tube was shaken at 65 °C for 1 h 30 min. Then the solution was cooled to 0 °C. Aqueous NaOAc (3 M, 25µl) and *n*-BuOH (1 ml) were added. The suspension was vortexed and left at -80 °C for 1 h. After centrifugation (4000 rpm, 4 °C, 15 min), the supernatant was carefully removed. The pellet was frozen with liquid nitrogen and

lyophilized. The crude RNA was purified by reversed-phase HPLC, the structural integrity was analyzed by MALDI-TOF.

General protocol for desalting of oligonucleotides using *Sep-Pak® Plus Short C18* column

The cartridge was first equilibrated with 10 ml of MeCN, then with MeCN/100 mM NaCl (1/1, v/v, 10 ml) and finally with 100 mM NaCl (10 ml). The sample was loaded in 5 ml of 100 mM NaCl. The cartridge was washed with 10 ml of H_2O. The oligonucleotide was eluted with 5 ml of MeCN/MeOH/H_2O (7/7/6) and evaporated.

General protocol for peptide bond formation

Equimolar amounts of the desalted $(m^6)aa^6A$ donor strand and a desalted complementary acceptor were mixed in an *Eppendorf* tube. Next, 100-120 equivalents of a 144 mM solution of DMTMM in aqueous 1 M NaCl, 0.1 M MOPS buffer (pH 7) were added. Typically, the peptide coupling experiments were conducted at 40-60 μM concentration of RNA. The mixture was incubated at room temperature for a total of 2-5 days while frequently being monitored by HPLC and MALDI-TOF analysis. Every day the new portion of DMTMM (100 eq.) is added to the reaction mixture, if necessary. Product was purified by reversed-phase HPLC eluting with phosphate buffers (C18 column; buffer A: 10 mM K_2HPO_4/KH_2PO_4 (pH 7.0) in H_2O and buffer B: 10 mM K_2HPO_4/KH_2PO_4 in MeCN/MeOH (1/7/2, v/v/v); a flow of 0.5 ml/min was used, a gradient of 0-15% of buffer B in 45 min was applied).

General protocol for urea cleavage

The peptide was dissolved in citrate buffer (0.1 M, pH 3.3) and was heated for 9 h at 95 °C in PCR machine. The reaction progress was checked by MALDI-TOF and HPLC.

RP-HPLC was performed on an analytical HPLC *Waters Alliance* (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column *Nucleosil 120-3 C8* from *Macherey Nagel* using a flow of 0.5 ml/min, a gradient of 0-30% of buffer B in 90 min was applied. The buffer system: buffer A: 100 mM NEt_3/HOAc (pH 7.0) in H_2O and buffer B: 100 mM NEt_3/HOAc in 80% (v/v) acetonitrile.

Synthesis and Purification of siRNAs

All of the oligonucleotides used in this study were synthesized on a 1 μmol scale using a DNA automated synthesizer (*Applied Biosystems 394 DNA/RNA Synthesizer*) with standard phosphoramidite chemistry. The phosphoramidites of nucleosides were purchased from *Glen Research*, *Sigma-Aldrich*, *Baseclick* or *Linktech*. Oligonucleotides were synthesized in DMT-OFF mode using phosphoramidites (Bz-A, Dmf-G, Ac-C, U, dT, 2'-OMe-Ac-C, 2'-OMe-U, C8-Alkyne-dU) with BTT in CH_3CN as an activator, DCA in CH_2Cl_2 as a deblocking solution and Ac_2O in pyridine/THF as a capping reagent. The solid support was suspended in a mixture of aqueous ammonia and methylamine (1:1, 1 ml) and heated at 65 °C for 15 min. The supernatant was removed, and the beads were washed with water. The supernatant and washings were combined, and the solvents were evaporated under reduced pressure. The residue was subsequently heated with a solution of triethylamine trihydrofluoride ($NEt_3\cdot3HF$ 98%, 125 μl) in DMSO (50 μl) at 65 °C for 1.5 h. The RNA was precipitated by addition of aqueous NaOAc solution (3 M, 25 μl) and *n*-butanol (1 ml). To ensure complete precipitation, the sample was incubated at -80 °C for 1 h. After centrifugation (4 °C, 4000 rpm, 15 min), the supernatant was removed, and the precipitated RNA was lyophilized. Oligonucleotides were purified using reverse-phase HPLC (*Agilent 1260 Infinity II*) equipped with the column *VP 250/32 C18* from *Macherey Nagel* using a flow rate of 5 ml/min with a gradient of 0-30% of buffer B in 45 min. The buffer system applied: buffer A: 100 mM NEt_3/HOAc (pH 7.0) in H_2O and buffer B: 100 mM NEt_3/HOAc in 80% (v/v) acetonitrile. Calculation of concentrations was assisted using the software *OligoAnalyzer 3.0* (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). The structural integrity of the synthesized oligonucleotides was analyzed by MALDI-TOF mass measurements.

Table S1. Summary of the synthesized RNA oligonucleotides as well as their measured and calculated masses. X: C8-alkyne-dU, C_m: 2'-OMe-C, U_m: 2'-OMe-U.

Click product		Sequence	[M-H]⁻ (calc.)	[M-H]⁻ (meas.)
ON16	ON16_M17_sense	UCUUAC$_m$AACCAGAACUC$_m$AA**X**T	6695.0	6697.3
ON17	ON17_M17_antisense	UUGAGU$_m$UCUGGUUGU$_m$AAGA**X**T	6802.9	6801.9

"Click" reaction of RNA oligonucleotides with 6-azido-6-deoxy-D-glucose or cholesteryl-TEG-azide or arachidonoyl-modified azide (AA)[226]

To a solution of the oligonucleotide (2.0 mM, 5.0 µl) in DMSO/H₂O 1:1 an azide solution (50 mM, 10 eq.) in DMSO/*t*-BuOH 3:1 was added. A fresh solution of CuBr (50 mM) was prepared by quickly dissolving CuBr in a 100 mM TBTA solution in DMSO/*t*-BuOH 3:1. 5.0 µl of this stabilized CuBr solution was added to the oligonucleotide/azide mixture and thoroughly shaken at room temperature for 1 h. RNA was precipitated with 100 µl of 3 M NaOAc and 1 ml of *n*-BuOH. To ensure the precipitation the mixture was kept at -80 °C for 1 h. After centrifugation the supernatant was removed, and the RNA was lyophilized. The crude mixture was analyzed by HPLC and MALDI-TOF.

For glucose-modified strands: the strands were purified on analytical RP-HPLC (*Agilent 1260 Infinity II*) equipped with the column *Nucleosil 120-2 C18* from *Macherey Nagel* using a flow of 1 ml/min, a gradient of 0-50% of buffer B in 45 min was applied. The buffer system: buffer A: 100 mM NEt₃/HOAc (pH 7.0) in H₂O and buffer B: 100 mM NEt₃/HOAc in 80% (v/v) acetonitrile.

For cholesteryl or arachidonoyl-modified strands: the strands were purified on analytical RP-HPLC (*Agilent 1260 Infinity II*) equipped with the column *Nucleosil 120-2 C18* from *Macherey Nagel* at 65 °C using a flow of 1 ml/min, a gradient of 0-85% of buffer B in 45 min was applied.

The buffer system: buffer A: 100 mM NEt$_3$/HOAc (pH 7.0) in H$_2$O and buffer B: 100 mM NEt$_3$/HOAc in 80% (v/v) acetonitrile.

"Click" reaction of RNA oligonucleotides with a linker and with sugars

To a solution of the oligonucleotide (2.0 mM, 5.0 μl) in DMSO/H$_2$O 1:1 an azide linker solution (50 mM, 10 eq.) in DMSO/*t*-BuOH 3:1 was added. A fresh solution of CuBr (50 mM) was prepared by quickly dissolving CuBr in a 100 mM TBTA solution in DMSO/*t*-BuOH 3:1. 5.0 μl of this stabilized CuBr solution was added to the oligonucleotide/azide mixture and thoroughly shaken at room temperature for 30 min (the reaction progress was monitored by MALDI-TOF). Then a alkynyl-glucose solution (50 mM, 20 eq.) in DMSO/*t*-BuOH 3:1 was added. A fresh solution of CuBr (50 mM) was prepared by quickly dissolving CuBr in a 100 mM TBTA solution in DMSO/*t*-BuOH 3:1. 5.0 μl of this stabilized CuBr solution was added to the oligonucleotide/azide mixture and thoroughly shaken at room temperature for 1 h (the reaction progress was monitored by MALDI-TOF). Then the RNA was precipitated with 100 μl of 3 M NaOAc and 1 ml of *n*-BuOH. To ensure the precipitation the mixture was kept at -80 °C for 1 h. After centrifugation the supernatant was removed, and the RNA was lyophilized. The crude mixture was analyzed by HPLC and MALDI-TOF. The strands were purified on analytical RP-HPLC (*Agilent 1260 Infinity II*) equipped with the column *Nucleosil 120-2 C18* from *Macherey Nagel* using a flow of 1 ml/min, a gradient of 0-85% of buffer B in 45 min was applied. The buffer system: buffer A: 100 mM NEt$_3$/HOAc (pH 7.0) in H$_2$O and buffer B: 100 mM NEt$_3$/HOAc in 80% (v/v) acetonitrile.

Table S2. Summary of the synthesized RNA oligonucleotides as well as their measured and calculated masses. X: C8-alkyne-dU, C$_m$: 2'-OMe-C, U$_m$: 2'-OMe-U.

Click product		Sequence	[M-H]$^-$ (calc.)	[M-H]$^-$ (meas.)
ON18	ON18_sense + Glu	Glu + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T	6900.1	6901.0
ON19	OM19_antisense + Glu	Glu + UUGAGU$_m$UCUGGUUGU$_m$AAGA**X**T	7008.0	7007.2

ON28	ON28_sense + 2Glu	2Glu + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T		8153.7	8153.1
ON29	ON29_antisense + 2Glu	2Glu + UUGAGU$_m$UCUGGUUGU$_m$AAGA**X**T		8261.6	8260.3
ON30	ON30_sense + 3Glu	3Glu + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T		8676.9	8675.1
ON31	ON31_antisense + 3Glu	3Glu + UUGAGU$_m$UCUGGUUGU$_m$AAGA**X**T		8784.8	8783.3
ON22	ON22_sense + chol	Chol + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T		7325.5	7325.1
ON23	ON23_antisense + chol	Chol + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T		7433.4	7432.1
ON20	ON20_sense + AA[226]	AA + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T	7199.4	7198.4
ON21	ON21_antisense + AA[226]	AA + UCUUAC$_m$AACCAGAACUC$_m$AA**X**T	7307.3	7306.4

7 Appendix I

Supporting information of the publication "Synthesis of an acp³U phosphoramidite and incorporation of the hypermodified base into RNA"

Supporting information

Synthesis of an acp³U phosphoramidite and incorporation of the hypermodified base into RNA

Milda Nainytė, Tynchtyk Amatov, and Thomas Carell[*]

Department of Chemistry
Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany
E-mail: thomas.carell@lmu.de

1. General Experimental Methods

Chemicals were purchased from Sigma-Aldrich, TCI, Fluka, ABCR, Carbosynth or Acros organics and used without further purification. Strands containing canonical bases were purchased from Metabion. The solvents were of reagent grade or purified by distillation, unless otherwise specified. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F254 TLC plates from Merck KGaA. Flash column chromatography was performed on Geduran® Si60 (40-63 μm) silica gel from Merck KGaA. NMR spectra were recorded on Bruker AVIIIHD 400 spectrometers (400 MHz). ^{1}H NMR shifts were calibrated to the residual solvent resonances: DMSO-d_6 (2.50 ppm), CDCl$_3$ (7.26 ppm), Acetone-d_6 (2.05 ppm). ^{13}C NMR shifts were calibrated to the residual solvent: DMSO-d_6 (39.52 ppm), CDCl$_3$ (77.16 ppm), Acetone-d_6 (29.84 ppm). All NMR spectra were analysed using the program MestRENOVA 10.0.1 from Mestrelab Research S. L. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of the Ludwigs-Maximilians-Universität München on the spectrometer MAT 90 (ESI) from Thermo Finnigan GmbH. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system. All substances were directly applied as solids or on the ATR unit. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow of 0.5 mL/min, a gradient of 0-30% of buffer B in 45 min was applied. Preparative RP-HPLC was performed on a HPLC Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel using a flow of 5 mL/min, a gradient of 0-25% of buffer B in 45 min was applied for the purifications. Oligonucleotides were purified using the following buffer system: buffer A: 100 mM NEt$_3$/HOAc (pH 7.0) in H$_2$O and buffer B: 100 mM NEt$_3$/HOAc in 80% (v/v) acetonitrile. The pH values of buffers were adjusted using a MP 220 pH-meter (Metter Toledo). Oligonucleotides were detected at wavelength: 260 nm. Melting profiles were measured on a JASCO V-650 spectrometer. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). For strands containing artificial bases, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a Bruker Autoflex II. For MALDI-TOF measurements, the samples were desalted on a 0.025 μm VSWP filter (Millipore) against ddH$_2$O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

2. Synthesis of the Phosphoramidite Building-Block

Compound 3b

The reaction was conducted according to a published procedure with minor modifications.[1] L-homoserine (**2**) (0.371 g, 3.12 mmol) was suspended in methanol (25 mL) and heated under reflux until the mixture became clear. Then, a solution of 9-BBN, 9-borabicyclo(3.3.1)nonane (6.7 mL, 3.35 mmol) in tetrahydrofuran (0.5 M) was added dropwise. The reaction mixture was refluxed for 3 hours under inert atmosphere. The reaction mixture was concentrated and the crude product purified by silica gel chromatography eluting with 50% ethyl acetate-hexane to 100% ethyl acetate. The 9-BBN protected L-homoserine (**3b**) was obtained as a white solid (yield 80%); mp 112 – 115 °C.

^{1}H NMR (400 MHz, DMSO-d_6) δ 0.49 (d, J = 0.5 Hz, 2H), 1.33 – 1.83 (m, 13H), 1.94 – 2.01 (m, 1H), 3.58 – 3.67 (m, 3H), 4.80 (t, J = 4.8 Hz, 1H), 5.86 – 5.91 (m, 1H), 6.41 – 6.46 (m, 1H); **^{13}C NMR** (101 MHz, DMSO-d_6) δ 23.9, 24.4, 30.9, 31.3, 33.1, 52.3, 57.9, 174.0; **IR** (v_{max}) 3425, 3227, 2843, 1721, 1595, 1276; **HRMS** (ESI): calculated for $C_{12}H_{23}BNO_3^+$ [M + H]$^+$: 240.1771; found 240.1764.

Compound 5
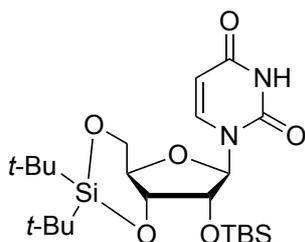
The reaction was conducted according to a published procedure with minor modifications.[2] Uridine (**4**) (1 g, 4.1 mmol) was dissolved in dry DMF (5 ml) and stirred at 0 °C. Di-*tert*-butylsilyl bis(trifluoromethanesulfonate) (1.6 ml, 2.2 g, 4.95 mmol, 1.2 eq.) was added dropwise. After 45 min, imidazole (1.4 g, 20.5 mmol, 5.0 eq.) was added and then the reaction was warmed to room temperature over the period of 30 min. Then, di-*tert*-butyldimethylsilyl chloride (0.75 g, 4.95 mmol, 1.2 eq.) was added and the reaction was left to stir overnight. Afterwards the reaction mixture was diluted with ethyl acetate and extracted with saturated

NaHCO$_3$ and water. The organic layer was dried over Na$_2$SO$_4$ and evaporated. The crude product was then purified by silica gel chromatography eluting with hexane/ethyl acetate (7/3, v/v) to afford the product **5** as a white solid (yield 86%).

¹H NMR (400 MHz, CDCl$_3$) δ 0.13 (s, 3H), 0.18 (s, 3H), 0.92 (s, 9H), 1.01 (s, 9H), 1.04 (s, 9H), 3.86 (dd, J = 9 Hz, J = 5 Hz, 1H), 3.97 (d, J = 9 Hz, 1H), 4.09 – 4.18 (m, 1H), 4.28 (d, J = 5 Hz, 1H), 4.50 (dd, J = 9 Hz, J = 5 Hz, 1H), 5.65 (s, 1H), 5.74 (d, J = 8 Hz, 1H), 7.25 (d, J = 8 Hz, 1H), 9.47 (s, 1H); **¹³C NMR** (101 MHz, CDCl$_3$) δ -4.9, -4.2, 18.4, 20.5, 22.9, 25.9, 27.1, 27.6, 67.7, 74.6, 75.5, 76.2, 94.0, 102.5, 139.5, 149.7, 163.1; **IR** (ν$_{max}$) 3227, 2920, 2844, 1707, 1597, 1452, 1357, 1259; **HRMS** (ESI): calculated for C$_{23}$H$_{43}$O$_6$N$_2$Si$_2^+$ [M + H]$^+$: 499.2660; found 499.2657.

Compound 6
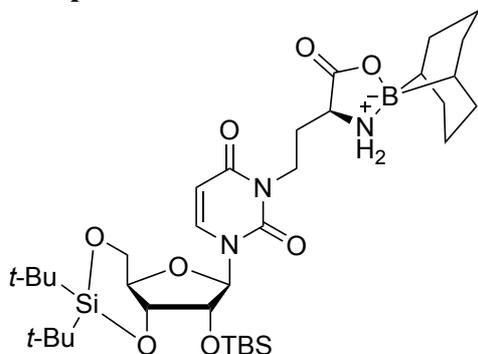
2'-*O*-(*tert*-butyldimethylsilyl)-3'-5'-*O*-(di-*tert*-butylsilylene)-uridine (**5**) (0.38 g, 0.7 mmol) was dissolved in dry tetrahydrofuran (7 mL) under nitrogen. Then 9-BBN protected homoserine (**3b**) (0.22 g, 0.91 mmol, 1.2 eq.) and triphenylphosphine (0.24 g, 0.91 mmol, 1.2 eq.) were added. Afterwards DIAD (0.19 ml, 0.2 g, 0.99 mmol, 1.3 eq.) was added dropwise at 0 °C. The reaction mixture was left to stir for 3 hours at room temperature and then volatiles were removed *in vacuo*. The residue was purified by silica gel chromatography eluting with hexane/ethyl acetate (1/1, v/v) to afford the compound **6** as a white solid (yield 88%).

¹H NMR (400 MHz, CDCl$_3$) δ 0.13 (s, 3H), 0.18 (s, 3H), 0.56 – 0.63 (m, 2H), 0.92 (s, 9H), 1.02 (s, 9H), 1.05 (s, 9H), 1.29 – 1.91 (m, 12H), 1.96 – 2.06 (m, 1H), 2.49 – 2.58 (m, 1H), 3.35 – 3.41 (m, 1H), 3.84 (dd, J = 9 Hz, J = 5 Hz, 1H), 3.95 – 4.03 (m, 2H), 4.16 – 4.25 (m, 2H), 4.28 (d, J = 5 Hz, 1H), 4.39 – 4.4.6 (m, 1H), 4.50 (dd, J = 9 Hz, J = 5 Hz, 1H), 5.62 (s, 1H), 5.76 – 5.79 (m, 1H), 5.81 (d, J = 8 Hz, 1H), 7.33 (d, J = 8 Hz, 1H); **¹³C NMR** (101 MHz, CDCl$_3$) δ -4.9, -4.1, 18.3, 20.5, 22.9, 25.9, 27.1, 27.6, 29.0, 30.9, 31.3, 31.6, 32.2, 53.4, 67.7, 74.8, 75.4, 76.0, 94.8, 101.6, 138.2, 150.7, 163.5, 173.0; **IR** (ν$_{max}$) 2929, 2858, 1695, 1458,

1266, 1053, 800; **HRMS** (ESI): calculated for $C_{35}H_{63}BN_3O_8Si_2^+$ [M + H]$^+$: 720.4247; found 720.4243.

Compound 7
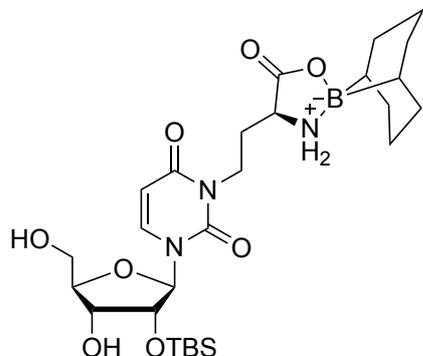
Compound **6** (0.5 g, 0.69 mmol) was dissolved in dry CH_2Cl_2 (7 ml), transferred in a falcon tube and cooled to 0 °C. Then pyridine was added (1 ml) followed by addition of HF-Pyridine (0.131 ml). After 1 hour the reaction was quenched with saturated $NaHCO_3$, washed with CH_2Cl_2, dried and evaporated. The residue was purified by silica gel chromatography eluting with CH_2Cl_2/CH_3OH (9/1, v/v) to afford alcohol **7** as a white solid (yield 99%).

^{1}H NMR (400 MHz, Acetone-d_6) δ 0.13 (s, 3H), 0.14 (s, 3H), 0.51 – 0.67 (m, 2H), 0.93 (s, 9H), 1.36 – 1.94 (m, 12H), 1.95 – 2.02 (m, 1H), 2.38 – 2.48 (m, 1H), 3.66 – 3.74 (m, 1H), 3.83 (dd, J = 12 Hz, J = 2.5 Hz, 1H), 3.94 (dd, J = 12 Hz, J = 2.5 Hz, 1H), 4.06 – 4.10 (m, 1H), 4.11 – 4.16 (m, 2H), 4.21 (t, J = 5 Hz, 1H), 4.39 (t, J = 4 Hz, 1H), 5.69 (t, J = 11 Hz, 1H), 5.77 (d, J = 8 Hz, 1H), 5.92 (d, J = 4 Hz, 1H), 6.05 – 6.18 (m, 1H), 8.27 (d, J = 8 Hz, 1H); **^{13}C NMR** (101 MHz, Acetone-d_6) δ -4.7, -4.6, 18.7, 24.9, 25.4, 26.1, 31.9, 32.1, 32.2, 32.6, 38.4, 53.7, 61.2, 70.4, 77.4, 85.4, 91.0, 101.7, 140.2, 152.5, 164.3, 172.9; **IR** (v_{max}) 3220, 3131, 2924, 2854, 1699, 1662, 1622, 1462, 1260, 1219, 1147, 1098, 966, 864, 837; **HRMS** (ESI): calculated for $C_{27}H_{46}O_8N_3BNaSi^+$ [M + Na]$^+$: 602.3045; found 602.3044.

Compound 8
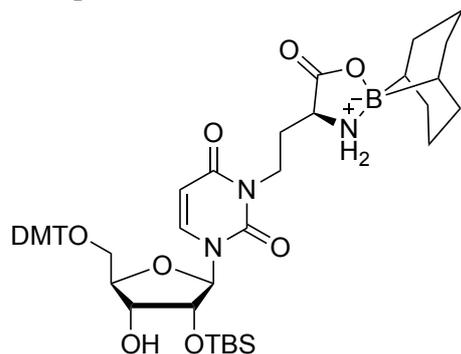
To the solution of compound **7** (0.25 g, 0.34 mmol) in dry pyridine (3 ml) 4,4′-dimethoxytrityl chloride (0.22 g, 0.65 mmol, 1.5 eq.) was added. The reaction mixture was stirred overnight under nitrogen atmosphere at room temperature. After evaporation of reaction mixture the residue was dissolved in CH_2Cl_2, washed with saturated $NaHCO_3$ and brine, dried over anhydrous Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography eluting with CH_2Cl_2/CH_3OH (9/1, v/v) containing 0.1 % of pyridine to afford the product **8** as white foam (yield 71%).

^{1}H NMR (400 MHz, Acetone-d_6) δ 0.16 (s, 3H), 0.20 (s, 3H), 0.51 – 0.63 (m, 2H), 0.95 (s, 9H), 1.42 – 2.01 (m, 13H), 2.38 – 2.47 (m, 1H), 3.55 (dd, J = 11 Hz, J = 3 Hz, 1H), 3.47 (dd, J = 11 Hz, J = 3 Hz, 1H), 3.68 – 3.77 (m, 1H), 3.80 (s, 6H), 4.10 – 4.20 (m, 3H), 4.41 – 4.4.6 (m, 2H), 5.39 (d, J = 8 Hz, 1H), 5.69 (t, J = 11 Hz, 1H), 5.86 (d, J = 2 Hz, 1H), 6.05 – 6.18 (m, 1H), 6.91 (d, J = 8.9 Hz, 4H), 7.26 – 7.49 (m, 9H), 8.07 (d, J = 8 Hz, 1H); **^{13}C NMR** (101 MHz, Acetone-d_6) δ -4.54, -4.5, 18.7, 22.2, 24.9, 25.5, 26.2, 31.9, 32.0, 32.2, 32.6, 38.4, 53.6, 55.5, 62.8, 69.1, 70.4, 77.4, 83.5, 87.5, 91.6, 101.6, 114.0, 126.1, 127.8, 128.7, 128.9, 129.7, 131.0, 136.1, 136.5, 139.8, 145.7, 152.4, 157.2, 159.7, 164.1, 172.9; **IR** (v_{max}) 2925, 2855, 1705, 1652, 1607, 1508, 1458, 1301, 1249, 1219, 1176, 1110, 833; **HRMS** (ESI): calculated for $C_{48}H_{64}O_{10}N_3BNaSi^+$ [M + Na]$^+$: 904.4352; found 904.4333.

Compound 1-PA
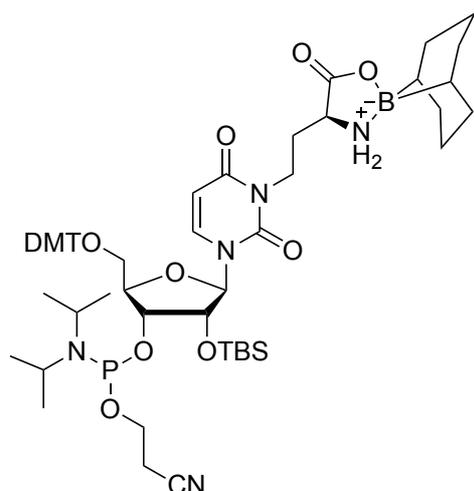
The DMT-protected nucleoside **8** (0.1 g, 0.11 mmol) was co-evaporated with pyridine and dried under high vacuum overnight. It was further dissolved in anhydrous CH₂Cl₂ (3 ml) and cooled to 0 °C. Then DIPEA (0.08 mL, 0.44 mmol, 4 eq.) and 2-cyanoethyl-*N*,*N*,-diisopropyl chlorophosphoramidite **(**0.063 mL, 0.28 mmol, 2.5 eq.) were added under nitrogen atmosphere. The reaction mixture was brought to room temperature and stirred for 3 hours. The reaction mixture was quenched by the addition of saturated NaHCO₃ solution and extracted with CH₂Cl₂. The combined organic extract was dried over anhydrous NaSO₄, filtered, concentrated *in vacuo*, and the residue was purified by silica gel column chromatography eluting with hexane/ethyl acetate (2/1, v/v, HPLC grade solvents) containing 0.1 % of pyridine to afford phosphoramidite **1-PA** as a white solid after lyophilization from benzene (yield of the mixture of diastereomers 89%).

Diastereomers can be separated during column chromatography; ¹H and ¹³C spectra are given of one of the isomers. **¹H NMR** (400 MHz, Acetone-*d₆*) δ 0.22 (s, 3H), 0.24 (s, 3H), 0.60 – 0.62 (br s, 2H), 0.95 (s, 9H), 1.07 (d, *J* = 6.8 Hz, 6H), 1.18 (d, *J* = 6.8 Hz, 6H), 1.43 – 1.97 (m, 12H), 1.98 – 2.01 (m, 1H), 2.38 – 2.47 (m, 1H), 2.71 – 2.79 (m, 2H), 3.51 (dd, *J* = 11, *J* = 3 Hz, 1H), 3.62 (dd, *J* = 11 Hz, *J* = 3 Hz, 1H), 3.60 – 3.73 (m, 2H), 3.80 (s, 6H), 3.82 – 3.86 (m, 1H), 3.94 – 4.16 (m, 3H), 4.29 – 4.32 (m, 1H), 4.45 – 4.51 (m, 1H), 4.57 (t, *J* = 4 Hz, 1H), 5.26 (d, *J* = 8 Hz, 1H), 5.60 – 5.66 (m, 1H), 5.85 (d, *J* = 3 Hz, 1H), 6.04 – 6.10 (m, 1H), 6.89 – 6.92 (m, 4H), 7.26 – 7.36 (m, 7H), 7.45 – 7.49 (m, 2H), 8.05 (d, *J* = 8 Hz, 1H); **¹³C NMR** (101 MHz, Acetone-*d₆*) δ -4.4, -4.2, 18.6, 21.1, 21.2, 24.8, 24.9, 25.0, 25.1, 25.5, 26.3, 32.0, 32.1,

32.2, 32.6, 38.3, 43.7, 43.8, 53.6, 55.5, 59.0, 59.2, 62.4, 72.5, 72.7, 76.5, 82.7, 82.8, 87.7, 91.1, 101.7, 114.0, 119.5, 127.9, 128.8, 129.1, 131.2, 135.9, 136.1, 139.6, 145.4, 152.1, 159.8, 164.4, 172.9; ^{31}P NMR (162 MHz, Acetone-d_6) δ 150.1, 148.8; IR (ν_{max}) 2927, 2853, 1703, 1654, 1608, 1508, 1459, 1363, 1302, 1252, 1178, 834, 809; HRMS (ESI): calculated for $C_{57}H_{82}BN_5O_{11}PSi^+$ [M+H]$^+$: 1082.5611; found 1082.5603.

3. Synthesis and Purification of Oligonucleotides

Sequence: 5' CAUGacp^{3}UUGCA 3' (ODN1)
Sequence: 5' GACUGACacp^{3}UCGUAGCacp^{3}UAACUCAU 3' (ODN4)

All of the oligonucleotides used in this study were synthesized on a 1 μmol scale using a DNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleosides were purchased from Glen Research and Sigma-Aldrich. Oligonucleotide containing acp^{3}U nucleoside was synthesized in DMT-OFF mode using phosphoramidites (Bz-A, Dmf-G, Ac-C, U) with BTT in CH$_3$CN as an activator, DCA in CH$_2$Cl$_2$ as a deblocking solution and Ac$_2$O in pyridine/THF as a capping reagent. The cleavage and deprotection of the CPG bound oligonucleotides were performed with aqueous NH$_4$OH/MeNH$_2$ (1/1, v/v, 1 mL) at room temperature for 1 h. The resin was removed by filtration and the solution was evaporated at under reduced pressure. The residue was subsequently heated with a solution of triethylamine trihydrofluoride (125 μL) in DMSO (50 μL) at 65 °C for 1.5 h. Upon cooling on ice bath, NaOAc (3.0 M, 25 μL) and n-BuOH (1 mL) were added. The resulting suspension was vortexed and cooled in a freezer (-80 °C) for 1 h. After the centrifugation, supernatant was removed and the remaining oligonucleotide pellet was dried under vacuum. The oligonucleotides were further purified by reverse-phase HPLC using a Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel. Oligonucleotides were purified using the following buffer system: buffer A: 100 mM NEt$_3$/HOAc, pH 7.0 in H$_2$O and buffer B: 100 mM NEt$_3$/HOAc in 80 % (v/v) acetonitrile. A flow rate of 5 mL/min with a gradient of 0-25 % of buffer B in 30 min was applied for the purifications. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow of 0.5 mL/min, a gradient of 0-30% of buffer B in 45 min was applied. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies:

https://eu.idtdna.com/calc/analyzer). For strands containing non-canonical base, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. The structural integrity of the synthesized oligonucleotides was analyzed by MALDI-TOF mass measurement.

a)

b)

c)

d)

Figure S1. (a) raw-HPL chromatogram of **ODN4**; (b) HPL chromatogram of purified **ODN4**; (c) MALDI-TOF mass spectrum of raw **ODN4**; (d) MALDI-TOF mass spectrum of purified **ODN4**.

4. UV Melting Curve Measurements

The UV melting curves were measured on JASCO V-650 spectrometer using 10 mm QS cuvettes, purchased from Hellma Analytics. A solution (80 μL) of equimolar amounts of oligonucleotides (4 μM each) in the buffer solution containing 10 mM sodium phosphate buffer (pH 7.0) and 150 mM NaCl was heated at 50 °C for 5 min and gradually cooled to 4 °C prior to the measurement. Melting profiles were recorded at temperatures between 5 and 75 °C with a ramping and scanning rate of 1 °C/min at 260 nm. All samples were measured at least three times. T_m values from each measurement were calculated using the "fitting curve" method and presented as an average of three independent measurements.

5. NMR Spectra of Synthesized Compounds

¹H NMR and ¹³C NMR of compound **3b**
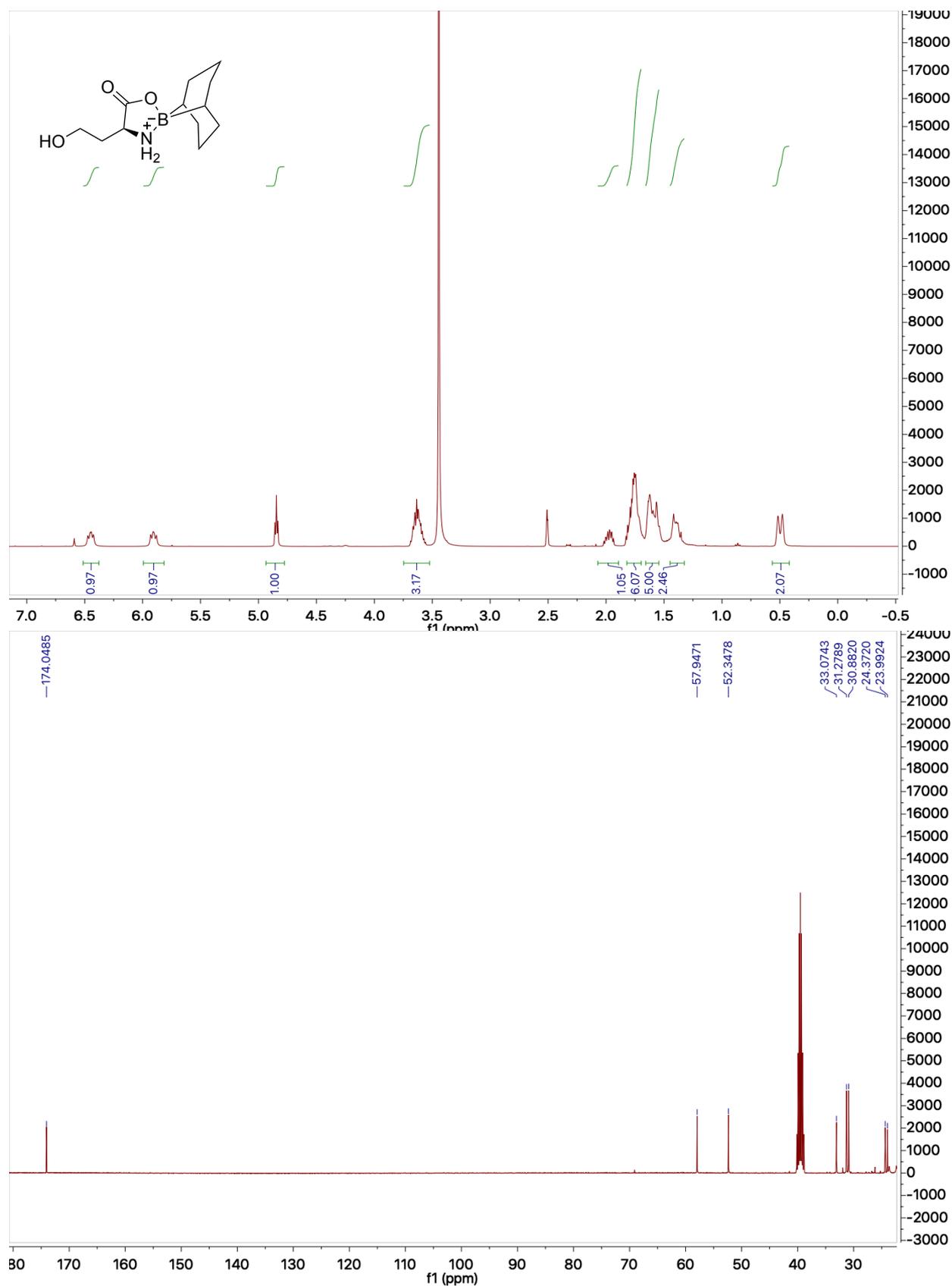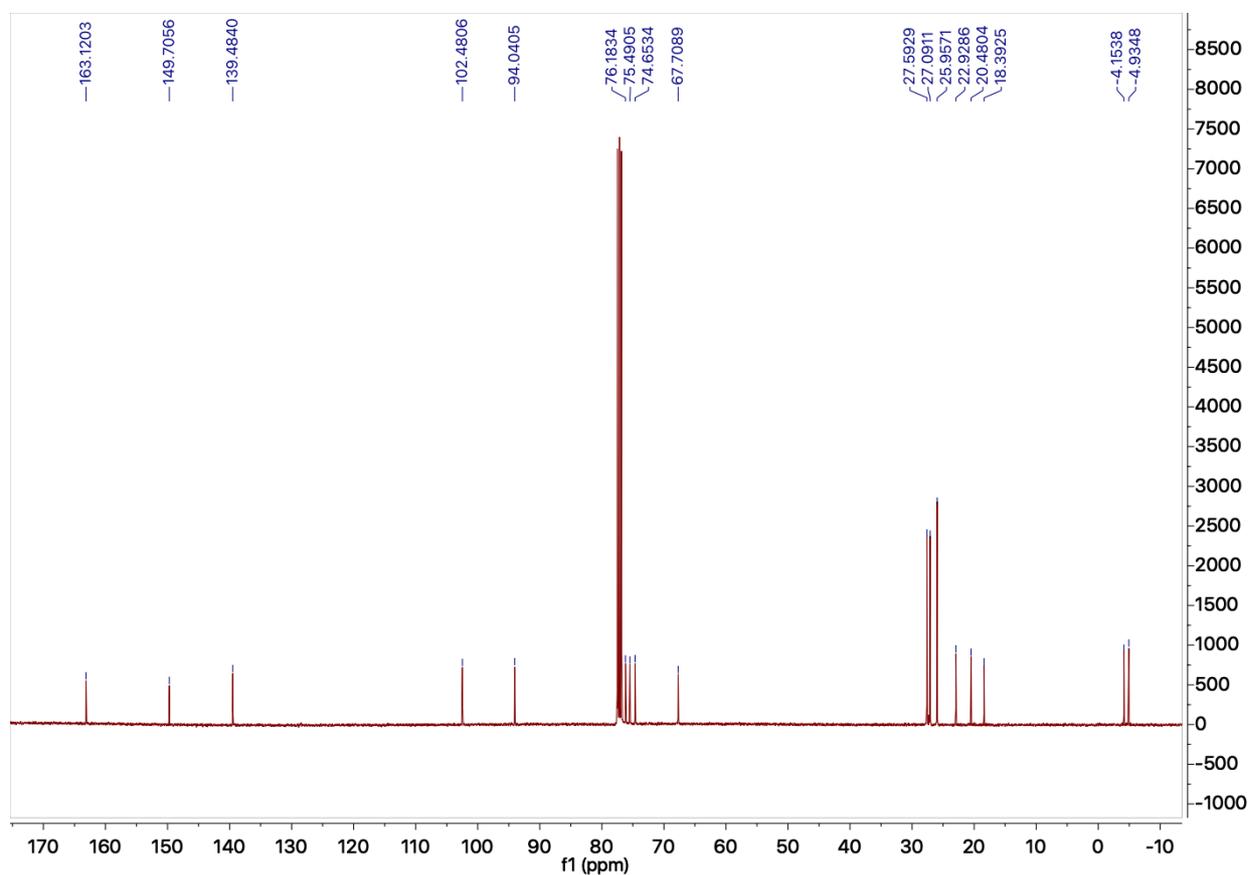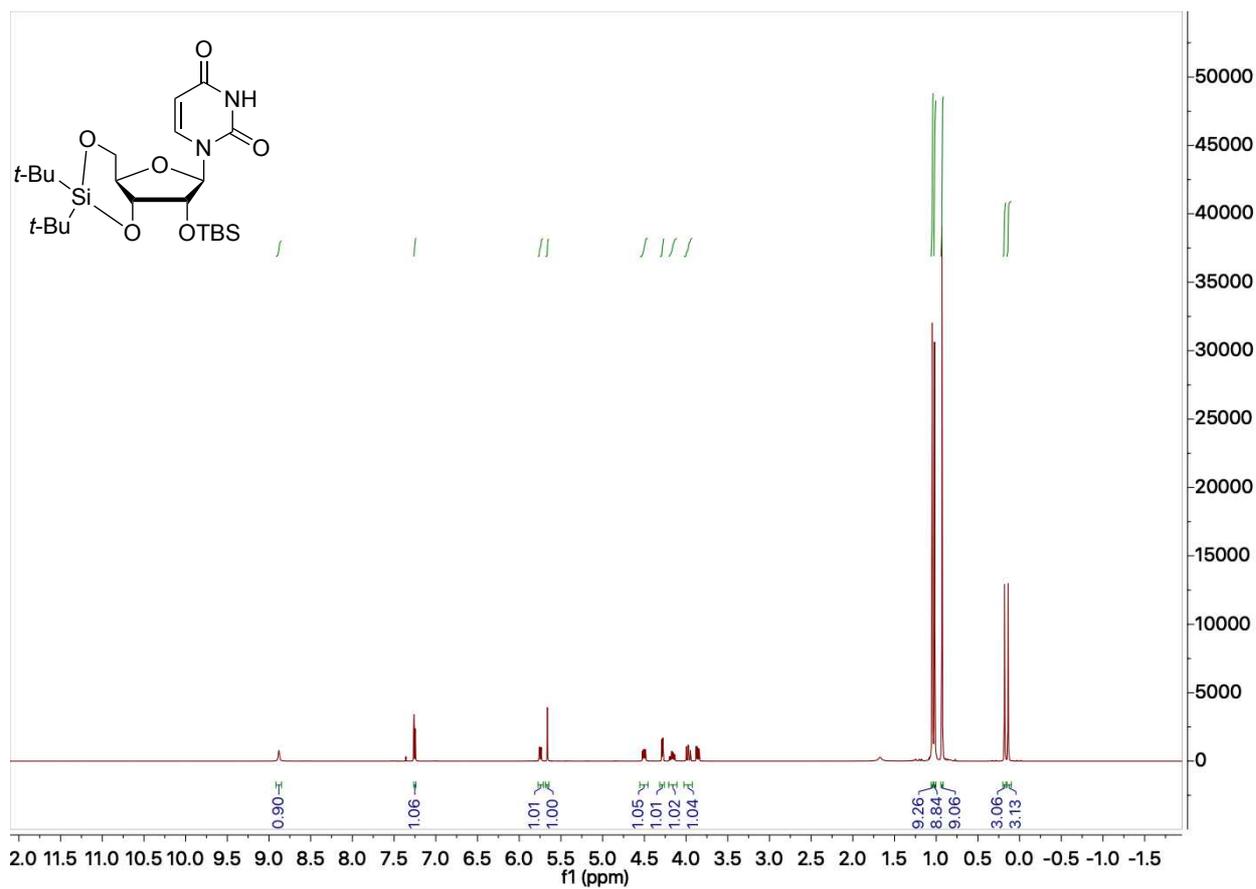
¹H NMR and ¹³C NMR of compound **6**
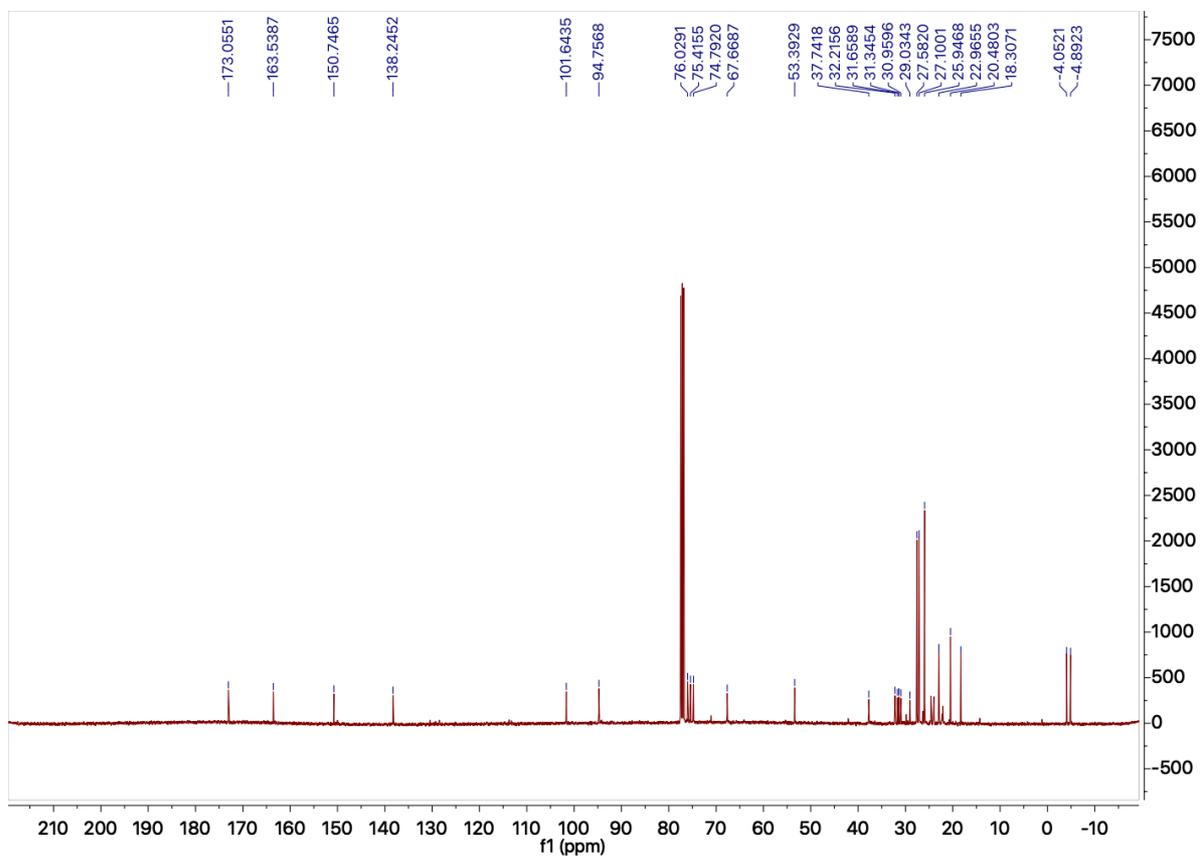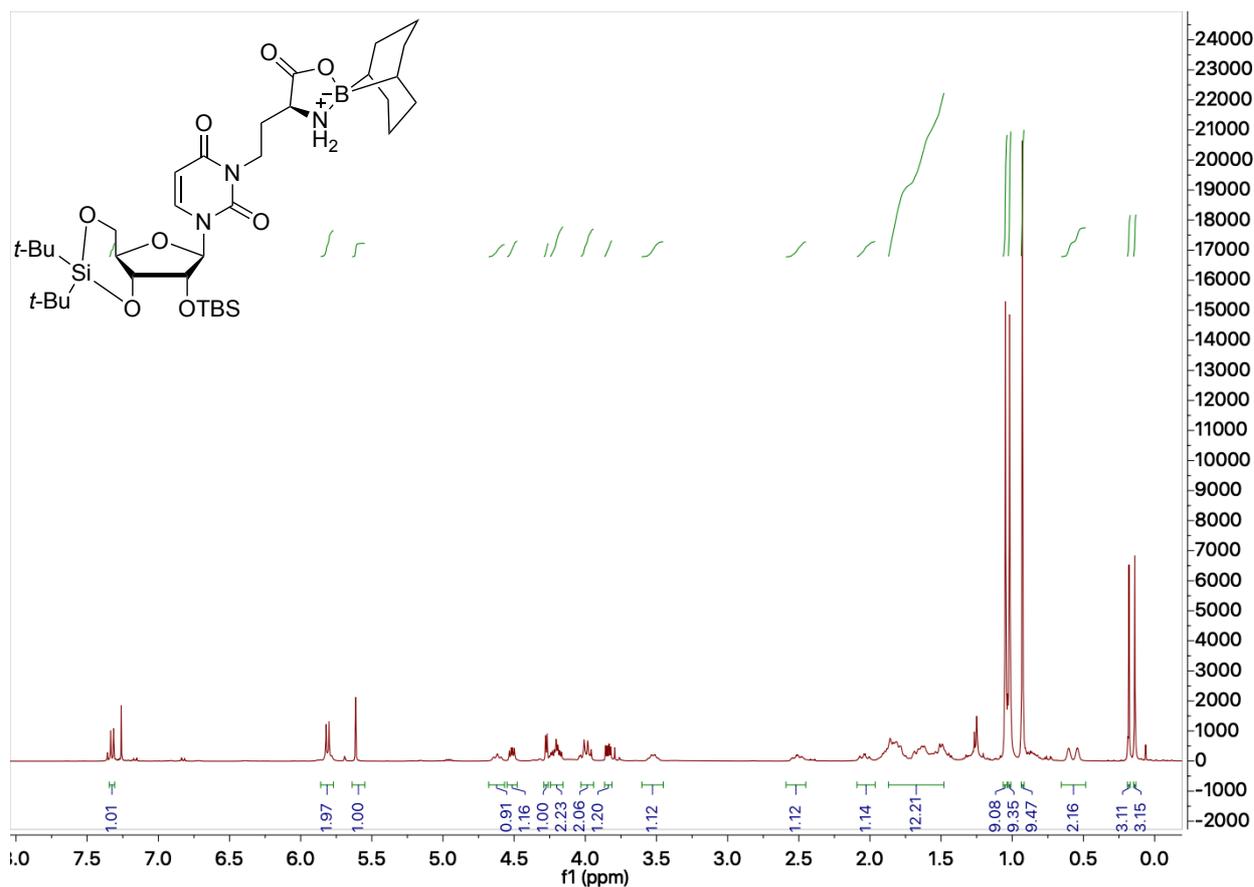
¹H NMR and ¹³C NMR of compound **7**
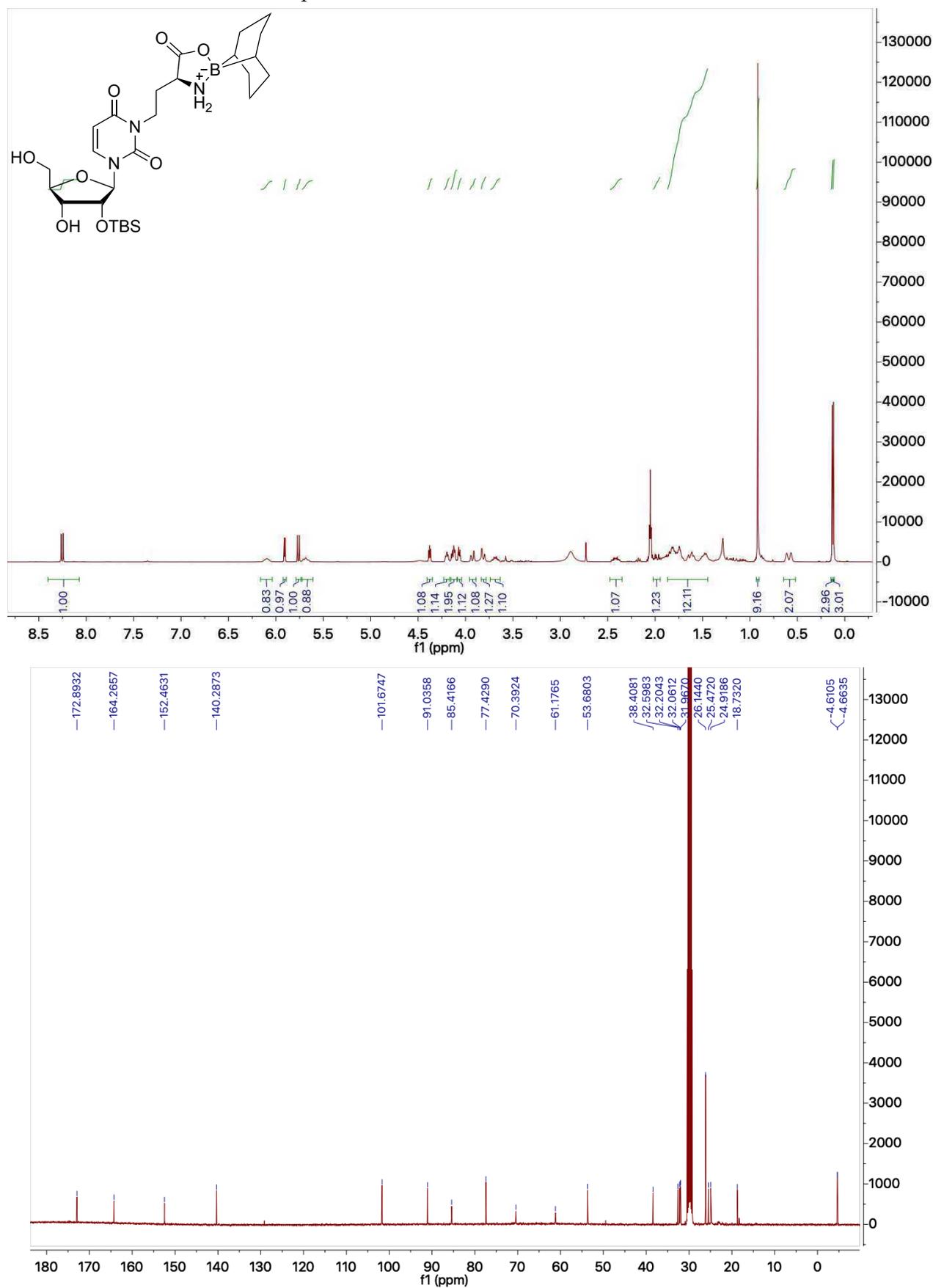
^{1}H NMR and ^{13}C NMR of compound **8**
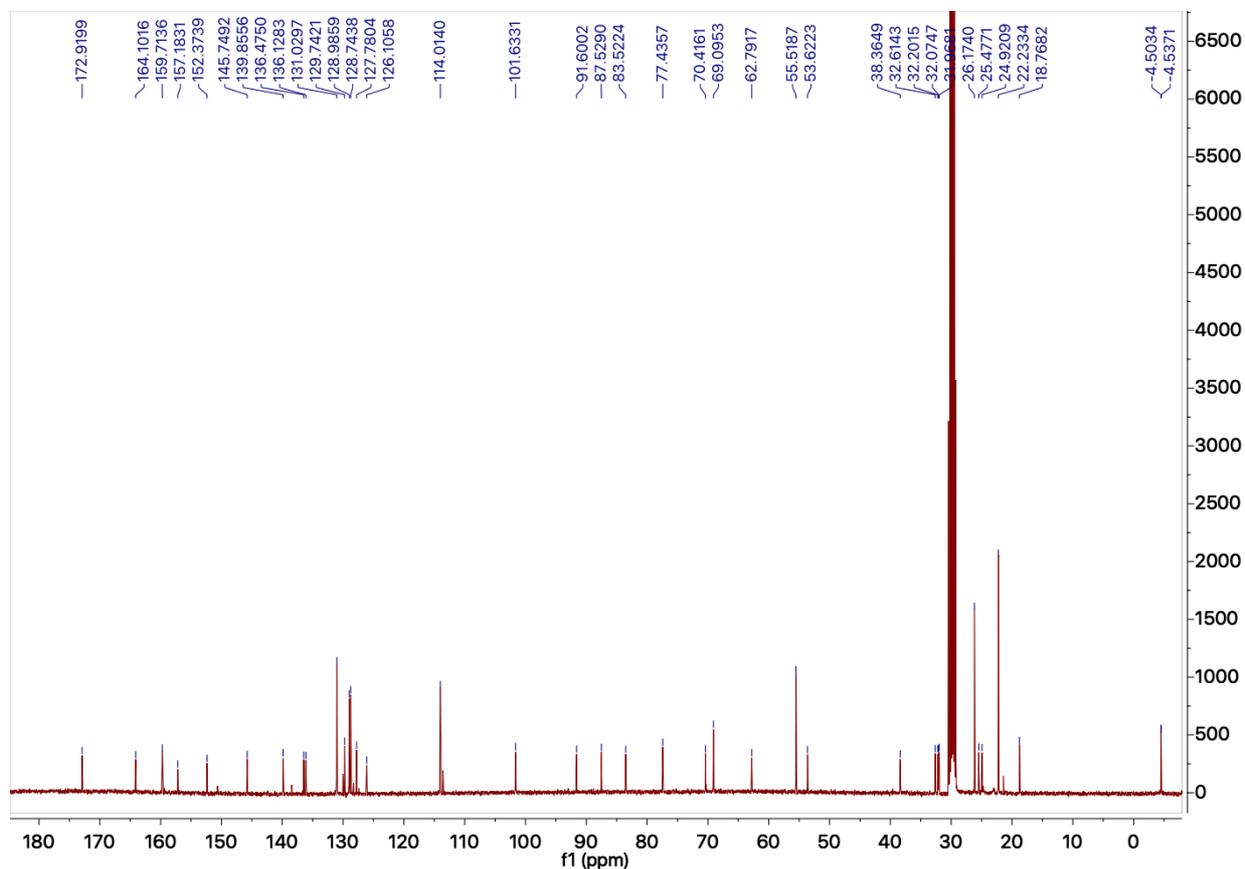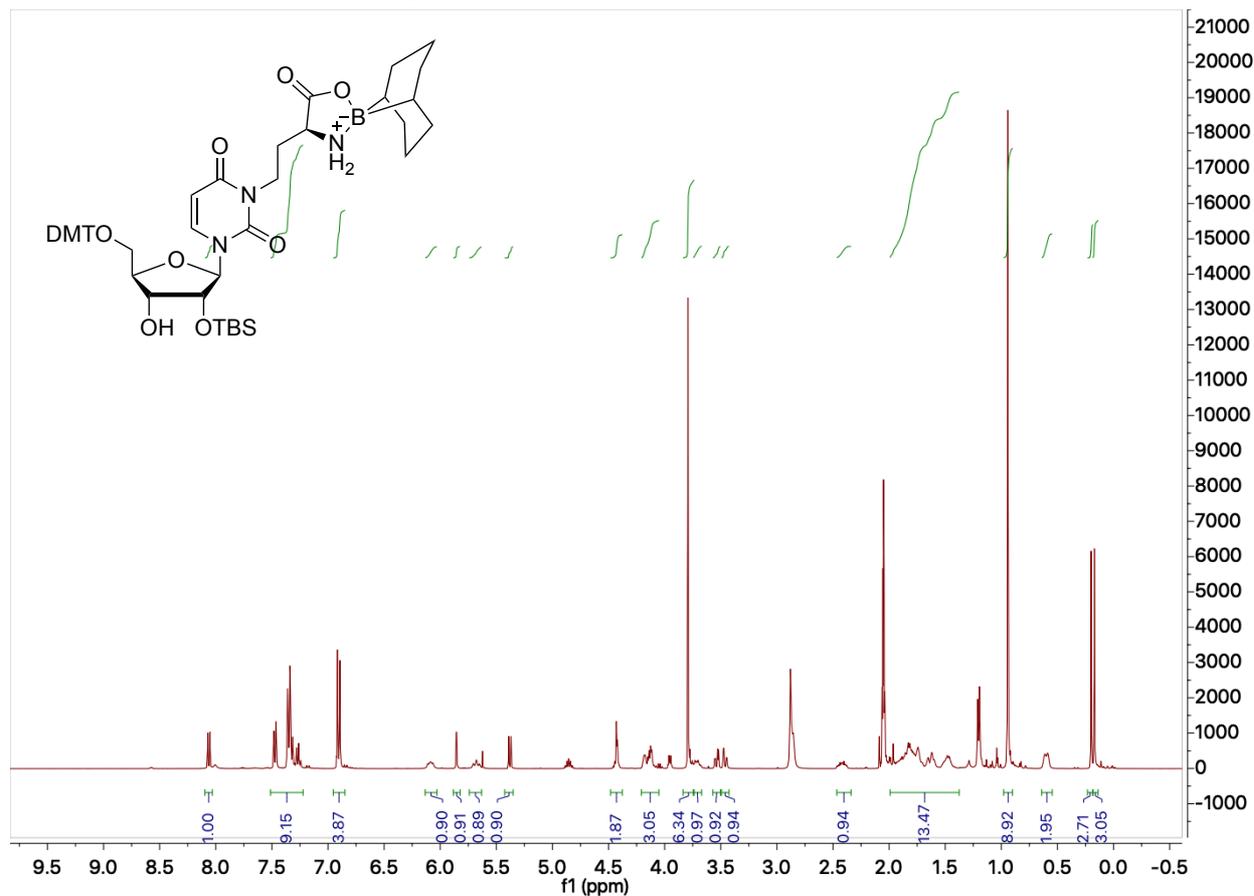
¹H NMR, ¹³C NMR and ³¹P NMR of compound **1-PA**
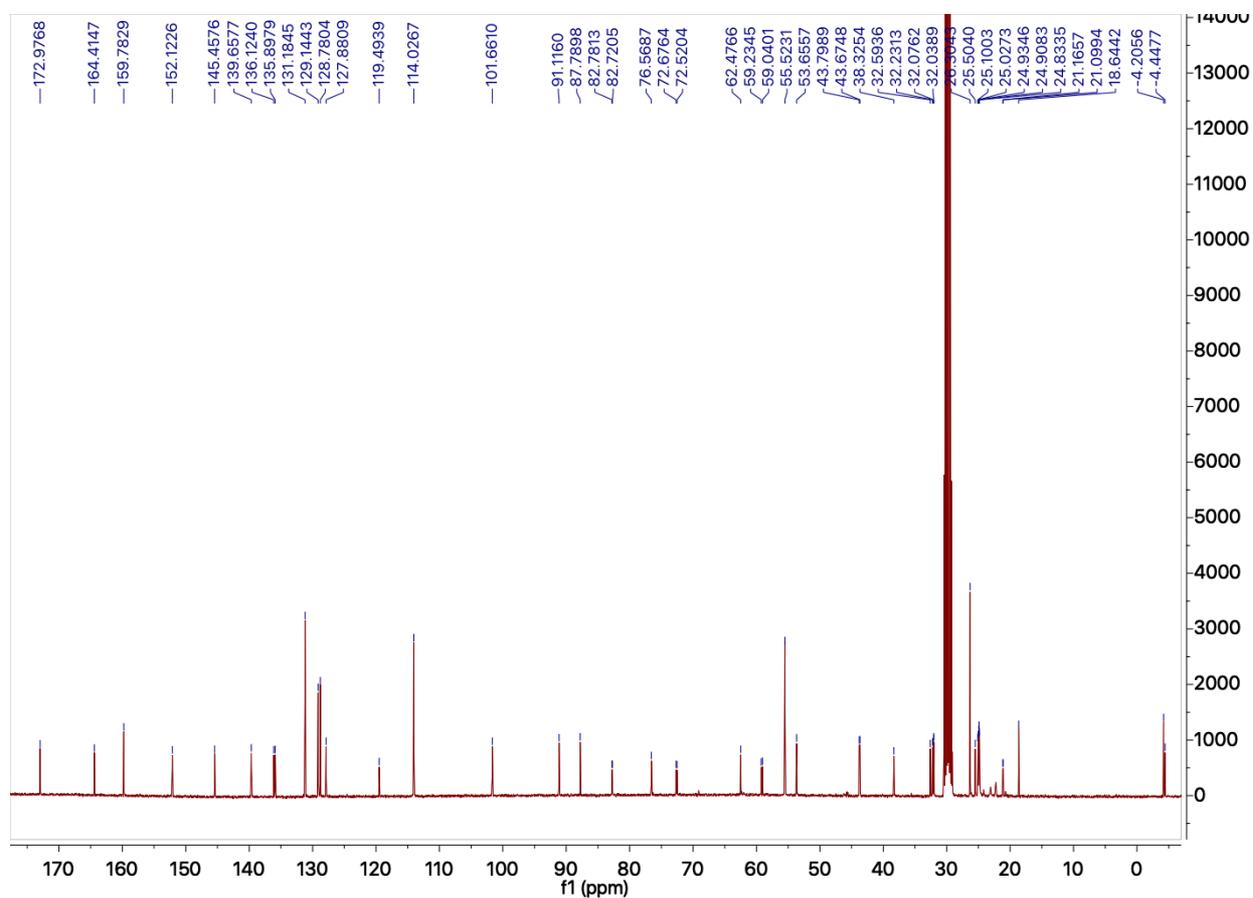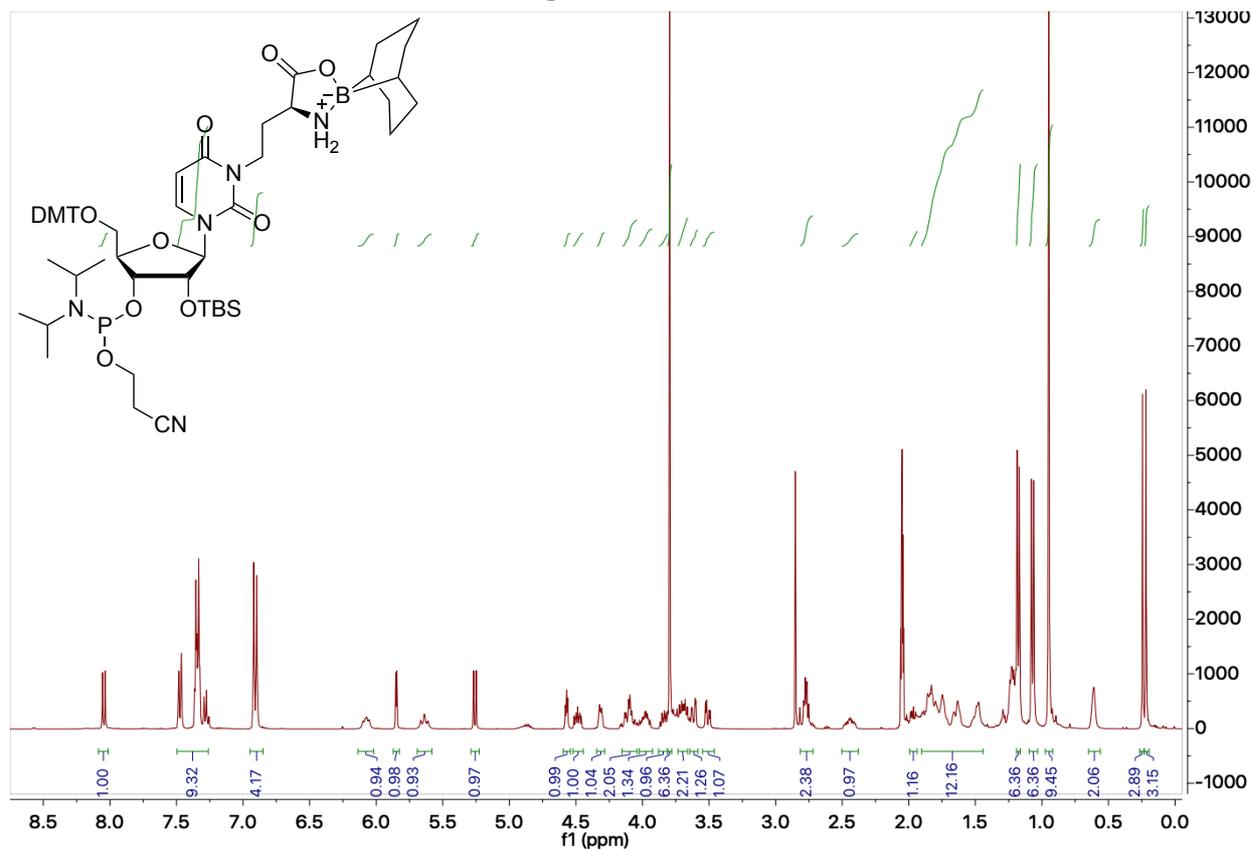
6. References

1. H. Taskent-Sezgin, J. Chung, P. S. Banerjee, S. Nagarajan, R. B. Dyer, I. Carrico, D. P. Raleigh, *Angew. Chem. Int. Ed.*, 2010, **49**, 7473–7475.

2. C. Riml, T. Amort, D. Rieder, C. Gasser, A. Lusser and R. Micura, *Angew. Chem., Int. Ed.*, 2017, **56**, 13479–13483.

8 Appendix II

Supporting information of the publication "Amino Acid Modified Bases as Building Blocks of an Early Earth RNA-Peptide World"

Supporting information

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

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

1. General Experimental Methods

Chemicals were purchased from Sigma-Aldrich, TCI, Fluka, ABCR, Carbosynth or Acros Organics and used without further purification. Some of the strands were purchased from Metabion or Ella Biotech. Reagent-grade dry solvents (Sigma-Aldrich, Acros Organics) were stored over molecular sieves and handled under inert gas atmosphere. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel F254 TLC plates from Merck KGaA. Flash column chromatography was performed on Geduran® Si60 (40-63 µm) silica gel from Merck KGaA. NMR spectra were recorded on Bruker AVIIIHD 400 spectrometers (400 MHz). ^{1}H NMR shifts were calibrated to the residual solvent resonances: DMSO-d_6 (2.50 ppm), CDCl$_3$ (7.26 ppm), Acetone-d_6 (2.05 ppm), CD$_2$Cl$_2$ (5.32 ppm). ^{13}C NMR shifts were calibrated to the residual solvent: DMSO-d_6 (39.52 ppm), CDCl$_3$ (77.16 ppm), Acetone-d_6 (29.84 ppm), CD$_2$Cl$_2$ (53.84 ppm). All NMR spectra were analysed using the program MestreNova 10.0.1 from Mestrelab Research S. L. High resolution mass spectra were measured by the analytical section of the Department of Chemistry of the Ludwigs-Maximilians-Universität München on the spectrometer MAT 90 (ESI) from Thermo Finnigan GmbH. IR spectra were recorded on a PerkinElmer Spectrum BX II FT-IR system. All substances were directly applied as solids or on the ATR unit. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow of 0.5 ml/min, a gradient of 0-30% of buffer B in 45 min was applied. Preparative RP-HPLC was performed on a HPLC Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel using a flow of 5 ml/min, a gradient of 0-25% of buffer B in 45 min was applied for the purifications. Oligonucleotides were purified using the following buffer system: buffer A: 100 mM NEt$_3$/HOAc (pH 7.0) in H$_2$O and buffer B: 100 mM NEt$_3$/HOAc in 80% (v/v) acetonitrile. The pH values of buffers were adjusted using a MP 220 pH-meter (Metter Toledo). Oligonucleotides were detected at wavelength: 260 nm. Melting profiles were measured on a JASCO V-650 spectrometer. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). For strands containing artificial bases, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a Bruker Autoflex II. For MALDI-TOF measurements, the samples were desalted on a 0.025 µm VSWP filter (Millipore)

against ddH$_2$O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

2. Synthesis of the Phosphoramidite Building-Blocks

2.1. Synthesis of RNA building blocks

Npe-protection of carboxy group of amino acid

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

L-amino acid (1 eq.), 2-(4-nitrophenyl)ethanol (npe-OH, 3 eq.) and TsOH (3 eq.) were refluxed in toluene overnight in a *Dean-Stark* apparatus. The solution was cooled to room temperature and Et$_2$O 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$_2$O.

Compound 4
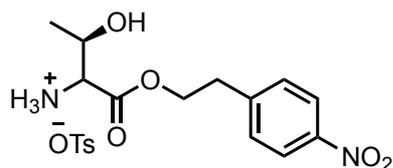
Yield: 70%; **IR**: $\tilde{v}$ = 3401 (w), 2930 (s), 2892 (s), 2858 (s), 1730 (s), 1510 (vs), 1465 (s), 1300 (vs), 1258 (s), 1167 (s), 1010 (s), 895 (w), 832 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.2 – 8.1 (m, 5H), 7.58 (d, J = 8.8 Hz, 2H), 7.48 (d, J = 8.2 Hz, 2H), 7.11 (d, J = 8.2 Hz, 2H), 5.6 (br. s., 1H), 4.45 (t, J = 6.2 Hz, 2H), 4.1 – 4.0 (m, 1H), 3.89 (d, J = 4.0 Hz, 1H), 3.10 (t, J = 6.2 Hz, 2H), 2.29 (s, 3H), 1.14 (d, J = 6.6 Hz, 3H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.2, 146.5, 146.3, 137.7, 130.4, 128.1, 125.6, 123.5, 65.4, 64.9, 57.9, 33.8, 20.7, 20.0; **HRMS (ESI)**: calculated for C$_{12}$H$_{17}$N$_2$O$_5$$^+$: m/z = 269.1137 [M+H]$^+$; found: m/z = 269.1140 [M+H]$^+$.

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

Compound 5

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

Compound 6

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

General procedure of hydroxy group protection with TBSCl

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

Npe-protected ester (1 eq.) was dissolved in pyridine and treated with one half of TBSCl (3 eq.) and 1H-imidazole (3 eq.). After 10 min, the second half was added, and the reaction mixture was left to stir at room temperature overnight. The mixture was diluted with CH_2Cl_2 and washed successively with sat. NaHCO$_3$ solution and H$_2$O. The organic layer was dried, evaporated and purified by flash chromatography eluting with CH_2Cl_2/MeOH (10/1, v/v) to afford the target compound as an oil.

Compound 7

Yield: 94%; **IR**: $\tilde{v}$ = 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$_2$O$_5$Si$^+$: m/z = 383.2002 [M+H]$^+$; found: m/z = 383.1997 [M+H]$^+$.

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

Compound 8

Yield: 96%; **IR**: $\tilde{v}$ = 3854 (w), 3745 (w), 2955 (w), 2930 (w), 2856 (w), 1735 (vs), 1601 (s), 1518 (vs), 1472 (w), 1374 (w), 1344 (vs), 1251 (s), 1155 (s), 1075 (s), 967 (s), 855 (w), 835 (s), 775 (s), 747 (s) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 8.13 (d, J = 8.6 Hz, 2H), 7.36 (d, J = 8.6 Hz, 2H), 4.34 (q, J = 6.7 Hz, 2H), 3.82 (dd, J = 9.8, 3.7 Hz, 1H), 3.73 (dd, J = 9.8, 3.0 Hz, 1H), 3.47 (s, 1H), 3.04 (t, J = 6.7 Hz, 2H), 0.81 (s, 9H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 173.9, 146.9, 145.6, 129.8, 123.8, 65.4, 64.4, 56.5, 34.9, 25.7, 18.2, -5.5, -5.6; **HRMS (ESI)**: calculated for C$_{17}$H$_{29}$N$_2$O$_5$Si$^+$: m/z = 369.1846 [M+H]$^+$; found: m/z = 369.1842 [M+H]$^+$.

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

Boc-amino acid (1 eq.) was dissolved in CH$_2$Cl$_2$ under inert atmosphere and cooled to 0 ^{o}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$_2$SO$_4$ and evaporated. The crude product was purified by flash chromatography to afford the target product.

Compound 12

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

Yield: 96%; **IR**: $\tilde{v}$ = 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$_2$O$_6^+$: m/z = 367.1869 [M+H]$^+$; found: m/z = 367.1874 [M+H]$^+$.

Compound 13

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

Yield: 85%; **IR**: $\tilde{v}$ = 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$_2$O$_6^+$: m/z = 325.1400 [M+H]$^+$; found: m/z = 325.1398 [M+H]$^+$.

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

Compound 14a

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

Yield: 81%; **IR**: $\tilde{v}$ = 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$_2$Cl$_2$) 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$_2$O$_6$Na$^+$: m/z = 437.1683 [M+Na]$^+$; found: m/z = 437.1684 [M+Na]$^+$.

Compound 14b

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

Yield: 85%; **IR**: $\tilde{v}$ = 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$_3$O$_6$$^+$: m/z = 432.2129 [M+NH$_4$]$^+$; found: m/z = 432.2131 [M+NH$_4$]$^+$.
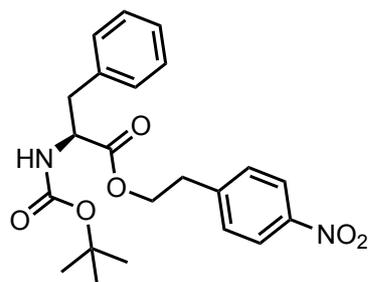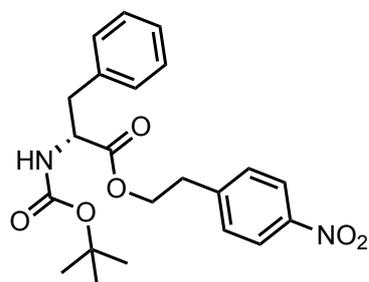
Compound 19
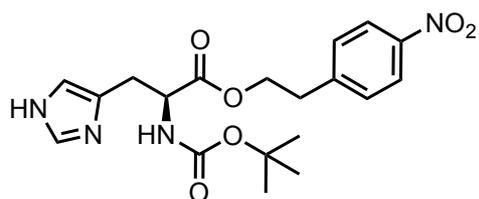
Boc-histidine (0.5 g, 1.96 mmol, 1 eq.), 2-(4-nitrophenyl)ethanol (0.655 g, 3.92 mmol, 2 eq.), DMAP (0.048 g, 0.39 mmol, 0.20 eq.) and HBTU (0.967 g, 2.55 mmol, 1.3 eq.) were dissolved in DMF (4 ml) under inert atmosphere. Diisopropylamine (686 µl, 4.90 mmol, 2.5 eq.) was added dropwise and the reaction mixture was stirred overnight at room temperature. The resulting solution was diluted with EtOAc (30 ml) and quenched with saturated 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**: $\tilde{v}$ =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; **¹H NMR (400 MHz, CDCl₃) 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); **¹³C NMR (101 MHz, CDCl₃) δ**: 172.1, 155.7, 147.1, 145.7, 135.2, 134.2, 130.0, 123.9, 115.9, 80.2, 64.9, 53.6, 34.9, 29.7, 28.5; **HRMS (ESI)**: calculated for $C_{19}H_{25}N_4O_6^+$: m/z = 405.1769 [M+H]⁺; found: m/z = 405.1765 [M+H]⁺.

General procedure for deprotection of Boc-protecting group

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

Compound 15
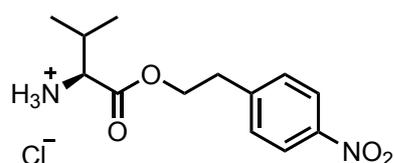
Yield: 99%; **IR**: $\tilde{v}$ = 3335 (w), 2964 (w), 2850 (w), 1741 (s), 1604 (w), 1516 (vs), 1464 (w), 1379 (vs), 1232 (vs), 1215 (s), 1170 (w), 1043 (w), 969 (w), 857 (s), 751 (s), 700 (s) cm⁻¹; **¹H**

NMR (400 MHz, DMSO-d_6) δ: 8.66 (br. s, 3H), 8.17 (d, J = 8.6 Hz, 2H), 7.62 (d, J = 8.6 Hz, 2H), 4.46 (dtd, J = 23.3, 11.1, 6.3 Hz, 2H), 3.76 (d, J = 4.5 Hz, 1H), 3.11 (t, J = 6.3 Hz, 2H), 2.10 (tt, J = 11.6, 5.8 Hz, 1H), 0.84 (d, J = 3.0 Hz, 3H), 0.82 (d, J = 3.0 Hz, 3H). **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.8, 146.3, 130.4, 123.4, 65.3, 57.2, 33.8, 29.2, 18.3, 17.4 **HRMS (ESI)**: calculated for $C_{13}H_{19}N_2O_4^+$: m/z = 267.1339 [M+H]$^+$; found: m/z = 267.1139 [M+H]$^+$.

Compound 16
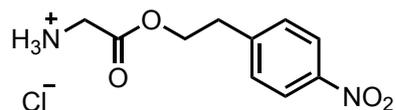
Yield: 99%; **IR**: $\tilde{v}$ = 2949 (w), 1746 (s), 1515 (vs), 1310 (vs), 1238 (vs), 1053 (w), 955 (s), 905 (s) 856 (s), 698 (s) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.30 (br. s, 3H), 8.18 (d, J = 8.7 Hz, 2H), 7.59 (d, J = 8.7 Hz, 2H), 4.44 (t, J = 6.4 Hz, 2H), 3.77 (s, 2H), 3.09 (t, J = 6.4 Hz, 2H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 167.6, 156.2, 146.3, 130.4, 123.5, 67.9, 33.9, 28.2, 22.0; **HRMS (ESI)**: calculated for $C_{10}H_{13}N_2O_4^+$: m/z = 225.0870 [M+H]$^+$; found: m/z = 225.0868 [M+H]$^+$.

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

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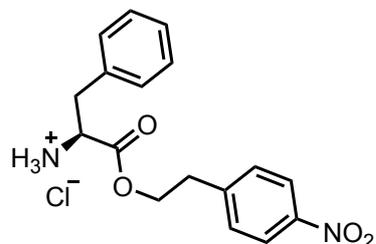
Yield: 99%; **IR**: $\tilde{v}$ = 3142 (w), 2988 (w), 2802 (w), 1740 (vs), 1601 (s), 1518 (vs), 1490 (vs), 1351 (vs), 1232 (vs), 1191 (s), 1102 (s), 981 (s), 856 (vs), 755 (vs), 736 (vs), 706 (vs) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.74 (br. s, 3H), 8.16 (d, J = 8.7 Hz, 2H), 7.49 (d, J = 8.7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

Compound 17b
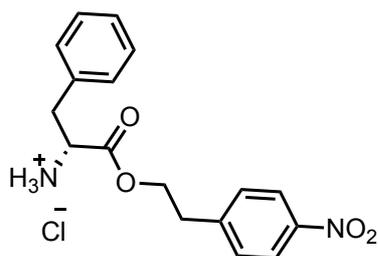
Yield: 99%; **IR**: $\tilde{v}$ = 3146 (w), 2988 (w), 2802 (w), 1740 (vs), 1602 (s), 1518 (vs), 1490 (vs), 1351 (vs), 1232 (vs), 1192 (s), 1102 (s), 856 (vs), 755 (vs), 736 (vs), 705 (vs) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.91 (br. s., 3H), 8.12 (d, J = 8.7 Hz, 2H), 7.46 (d, J = 8.7 Hz, 2H), 7.27 – 7.20 (m, 3H), 7.12 (d, J = 6.4 Hz, 2H), 4.29 (t, J = 6.4 Hz, 2H), 4.19 – 4.08 (m, 1H), 3.21 (dd, J = 14.1, 5.0, 1H), 3.10 – 2.88 (m, 3H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 168.9, 146.3, 134.8, 130.4, 129.4, 128.5, 127.2, 123.4, 65.4, 53.3, 35.8, 33.7; **HRMS (ESI)**: calculated for $C_{17}H_{19}N_2O_4^+$: m/z = 315.1339 [M+H]$^+$; found: m/z = 315.1332 [M+H]$^+$.

Compound 21
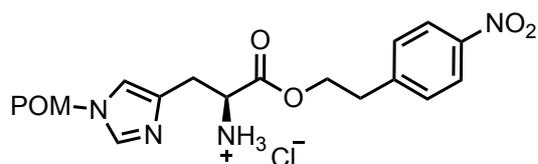
Yield: 98%; **IR**: $\tilde{v}$ = 2960 (w), 1737 (vs), 1624 (w), 1598 (w), 1516 (vs), 1454 (w), 1415 (w), 1351 (vs), 1282 (w), 1188 (w), 1124 (vs), 1045 (w), 1008 (w), 854 (s), 774 (w), 749 (w), 701 (w) cm^{-1}; **^{1}H NMR (400 MHz, DMSO-d_6) δ**: 8.69 (s, 3H), 8.21 – 8.15 (m, 2H), 7.59 – 7.55 (m, 2H), 7.51 (s, 1H), 6.04 (s, 2H), 4.39 (m, 3H), 3.56 (s, 1H), 3.17 (d, J = 6.9 Hz, 2H), 3.04 (td, J = 6.4, 3.9 Hz, 2H), 1.13 (s, 9H); **^{13}C NMR (101 MHz, DMSO-d_6) δ**: 176.7, 168.2, 146.4, 146.1, 137.4, 130.4, 123.5, 120.0, 69.2, 66.4, 65.6, 51.1, 38.2, 33.7, 26.5; **HRMS (ESI)**: calculated for $C_{20}H_{27}N_4O_6^+$: m/z = 419.1925 [M+H]$^+$; found: 419.1919 [M+H]$^+$.

Compound 20
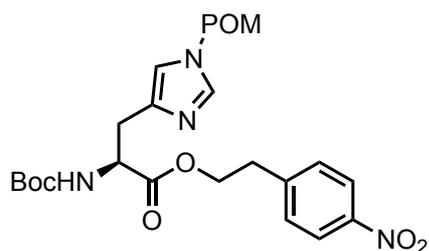
The histidine derivative **19** (0.745 g, 1.84 mmol, 1 eq.) was dissolved in DMF (12 mL) under N$_2$ atmosphere at 0 °C. Subsequently, K$_2$CO$_3$ (0.509 g, 3.68 mmol, 2 eq.) was added and the mixture was stirred for 40 min. Chloromethylpivalate (319 µl, 2.21 mmol, 1.2 eq.) was added dropwise at 0°C and the reaction mixture was left to warm to room temperature while stirring for 5 h. Catalytic amounts of KI were added and the mixture was stirred for another 1 h. The resulting suspension was diluted with EtOAc (75 ml) and quenched with saturated NH$_4$Cl solution (35 ml). The organic phase was washed with water, dried over Na$_2$SO$_4$ and concentrated *in vacuo*. The crude mixture was purified by flash chromatography on silica gel (40% isohexane in EtOAc to pure EtOAc). The pivalate protected histidine derivative was obtained as a yellow oil.

Yield: 55%; **IR**: 2850 (w), 2600 (w), 1738 (vs), 1624 (w), 1598 (w), 1573 (w), 1516 (vs), 1454 (w), 1414 (w), 1350 (vs), 1282 (w), 1191 (w), 1124 (vs), 1044 (w), 1008 (w), 854 (vs), 773 (w), 749 (w), 701 (w) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 8.19 – 8.14 (m, 2H), 7.54 (d, J = 1.4 Hz, 1H), 7.42 – 7.36 (m, 2H), 6.71 (s, 1H), 5.81 (d, J = 8.2 Hz, 1H), 5.73 (s, 2H), 4.51 (dt, J = 8.2, 5.3 Hz, 1H), 4.34 (t, J = 6.7 Hz, 2H), 3.04 (t, J = 6.7 Hz, 2H), 2.97 (t, J = 5.2 Hz, 2H), 1.42 (s, 9H), 1.14 (s, 9H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 177.9, 172.0, 155.7, 147.0, 145.8, 138.3, 130.0, 123.9, 117.3, 79.9, 77.4, 67.7, 64.7, 53.5, 38.9, 35.0, 30.1, 28.5, 27.0; **HRMS (ESI)**: calculated for C$_{25}$H$_{35}$N$_4$O$_8^+$: m/z = 519.2449 [M+H]$^+$; found: m/z = 519.2441 [M+H]$^+$.

Compound 22
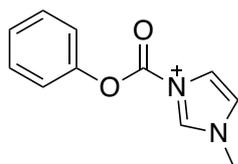
The reaction was conducted according to a published procedure.[3]

Phenyl chloroformate (4 ml, 31.9 mmol, 1 eq.) was dissolved in dry CH$_2$Cl$_2$ under nitrogen and cooled to 0 °C. Then *N*-methylimidazole (2.54 ml, 31.9 mmol, 1 eq.) was added dropwise. The mixture was allowed to stir at room temperature for 2 hours. Afterwards the reaction mixture was filtered, the precipitate was washed with CH$_2$Cl$_2$ and dried.

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

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

Compound 24
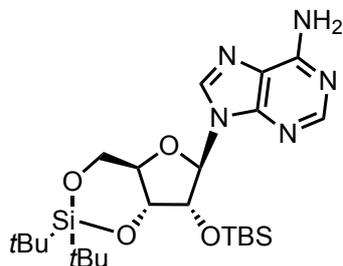
Compound was synthesized following the procedure published earlier.[4]

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

Yield: 76%; **IR**: $\tilde{v}$ = 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$_5$O$_4$Si$_2$$^+$: m/z = 522.2932 [M+H]$^+$; found: m/z = 522.2926 [M+H]$^+$.

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

General procedure for amino acid attachment to the adenosine derivative

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

Compound 25

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

Yield: 91%; **IR**: $\tilde{v} = 3237$ (w), 2931 (s), 2857 (s), 1737 (s), 1701 (vs), 1610 (s), 1520 (vs), 1465 (s), 1345 (s), 1250 (s), 1136 (w), 1057 (s), 998 (w), 894 (w), 840 (s), 777 (s) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 10.05 (d, *J* = 9.0 Hz, 1H), 8.56 (s, 1H), 8.41 (s, 1H), 8.16 (s, 1H), 7.94 (d, *J* = 8.6 Hz, 2H), 7.32 (d, *J* = 8.6 Hz, 2H), 5.99 (s, 1H), 4.64 (d, *J* = 4.6 Hz, 1H), 4.60-4.46 (m, 3H), 4.33-4.24 (m, 2H), 4.20-4.29 (m, 1H), 4.05 (dd, *J* = 10.5, 9.1 Hz, 1H), 3.03 (t, *J* = 6.5 Hz, 2H), 1.25 (d, *J* = 6.5 Hz, 3H), 1.08 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.90 (s, 9H), 0.19 (s, 3H), 0.16 (s, 3H), 0.07 (s, 3H), -0.04 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 171.1, 154.6,

151.3, 150.3, 149.8, 146.8, 145.7, 141.6, 129.9, 123.7, 121.1, 92.6, 76.0, 75.7, 75.0, 68.8, 68.0, 64.8, 59.8, 35.0, 27.7, 27.2, 26.1, 25.7, 22.9, 21.3, 20.6, 18.5, 18.0, -4.1, -4.8, -5.2; **HRMS (ESI)**: calculated for $C_{43}H_{72}N_7O_{10}Si_3^+$: m/z = 930.4643 [M+H]$^+$; found: m/z = 930.4640 [M+H]$^+$.

Compound 26

Eluent: 10% CH_2Cl_2 in EtOAc to pure EtOAc.

Yield: 86%; **IR**: $\tilde{v}$ = 3854 (w), 3745 (w), 3650 (w), 2932 (w), 2858 (w), 2361 (w), 2341 (w), 1735 (s), 1670 (s), 1654 (w), 1610 (w), 1587 (w), 1521 (vs), 1472 (s), 1395 (w), 1345 (vs), 1252 (w), 1166 (w), 1118 (vs), 1055 (vs), 999 (w), 894 (w), 826 (vs), 781 (s), 750 (w) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 10.05 (dd, J = 7.6, 3.7 Hz, 1H), 8.45 (s, 1H), 8.08 (s, 1H), 8.06 – 8.02 (m, 3H), 7.59 (d, J = 1.3 Hz, 1H), 7.41 – 7.35 (m, 2H), 6.86 (dd, J = 5.3, 1.3 Hz, 1H), 5.96 (d, J = 7.0 Hz, 1H), 5.74 (d, J = 2.6 Hz, 2H), 4.94 – 4.85 (m, 1H), 4.60 (dd, J = 6.0, 4.6 Hz, 1H), 4.53 – 4.46 (m, 2H), 4.42 (tq, J = 6.5, 1.7 Hz, 2H), 4.24 (tdd, J = 9.8, 5.0, 2.8 Hz, 1H), 4.08 – 4.00 (m, 1H), 3.23 – 3.12 (m, 2H), 3.07 (t, J = 6.5 Hz, 2H), 1.09 (s, 9H), 1.08 (s, 9H), 1.05 (s, 9H), 0.94 (s, 9H), 0.17 (s, 3H), 0.15 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 177.8, 171.7, 153.6, 151.3, 150.2, 149.8, 146.9, 145.8, 141.2, 138.4, 138.1, 123.0, 123.8, 121.1, 117.3, 92.6, 76.0, 75.7, 74.9, 67.9, 67.7, 64.7, 55.5, 38.8, 35.0, 30.7, 27.7, 27.2, 26.9, 26.1, 22.9, 20.5, 18.5, 1.3, -4.2; **HRMS (ESI)**: calculated for $C_{45}H_{68}N_9O_{11}Si_2^+$: m/z = 966.4571 [M+H]$^+$; found: m/z = 966.4576 [M+H]$^+$.

Compound 27

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

Yield: 78%; **IR**: $\tilde{v}$ = 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$_7$O$_9$Si$_2$$^+$: m/z = 814.3991 [M+H]$^+$; found: m/z = 814.3976.
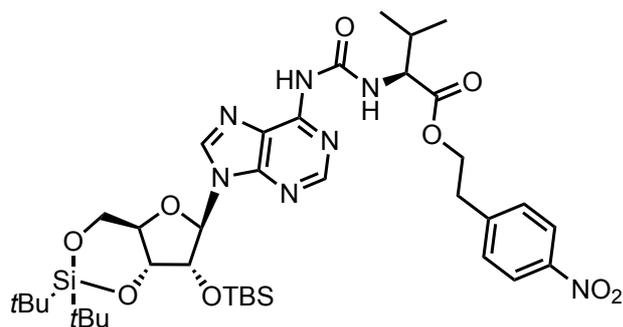
Compound 28
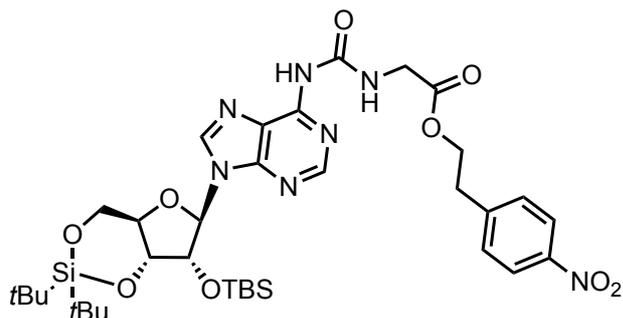
Eluent: Hex/EtOAc (4/3, v/v).

Yield: 72%; **IR**: $\tilde{v}$ = 3239 (w), 2932 (s), 2858 (s), 1749 (s), 1703 (vs), 1611 (s), 1520 (vs), 1468 (s), 1345 (s), 1252 (s), 1141 (w), 1055 (s), 990 (w), 894 (w), 826 (s), 750 (s) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 9.95 (br. s, 1H), 8.50 (s, 1H), 8.21 – 8.03 (m, 3H), 7.38 (d, J = 8.6 Hz, 2H), 5.98 (s, 1H), 4.60 (d, J = 4.7 Hz, 1H), 4.57 – 4.40 (m, 4H), 4.30 – 4.17 (m, 3H), 4.10 –

4.03 (m, 1H), 3.09 (t, J = 6.6 Hz, 2H), 1.08 (s, 9H), 1.05 (s, 9H) 0.94 (s, 9H), 0.17 (s, 3H), 0.16 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 170.0, 154.2, 151.2, 150.2, 149.9, 147.0, 145.5, 141.4, 129.9, 123.8, 121.1, 92.5, 76.0, 75.7, 74.9, 67.9, 42.2, 35.0, 27.6, 27.2, 26.0, 22.9, 20.5, 18.5, -4.1, -4.9; **HRMS (ESI)**: calculated for $C_{35}H_{54}N_7O_9Si_2^+$: m/z = 772.3522 [M+H]$^+$; found: m/z = 772.3504 [M+H]$^+$.

Compound 29
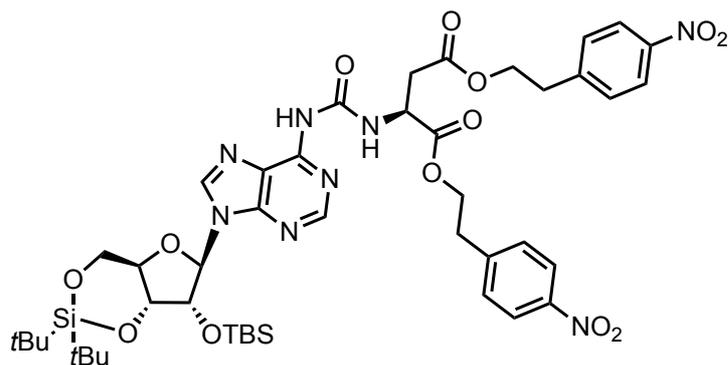
Eluent: Hex/EtOAc (2/1, v/v).

Yield: 87%; **IR**: $\tilde{v}$ = 3229 (w), 2953 (w), 2929 (w), 2857 (w), 1735 (s), 1693 (s), 1607 (s), 1589 (s), 1517 (vs), 1469 (s), 1391 (w), 1310 (vs), 1292 (w), 1251 (w), 1210 (w), 835 (s) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 10.28 (s, 1H), 8.42 (s, 1H), 8.22 – 8.12 (m, 1H), 8.07 (d, J = 8.6 Hz, 2H), 7.96 (d, J = 8.6 Hz, 2H), 7.37 – 7.30 (m, 4H), 6.00 (s, 1H), 4.93 – 4.89 (m, 1H), 4.64 (d, J = 4.6 Hz, 1H), 4.51 (ddd, J = 9.2, 4.9, 2.7 Hz, 2H), 4.46 – 4.40 (m, 2H), 4.39 – 4.22 (m, 3H), 4.09 – 4.05 (m, 1H), 3.08 – 2.94 (m, 6H), 1.07 (s, 9H), 1.05 (s, 9H), 0.95 (s, 9H), 0.19 (s, 3H), 0.17 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 179.8, 170.7, 153.7, 149.8, 146.9, 146.8, 145.5, 145.4, 129.8, 123.8, 123.6, 120.9, 92.6, 75.9, 75.7, 74.9, 67.9, 65.0, 64.5, 49.7, 36.5, 34.9, 34.8, 27.6, 27.1, 26.0, 22.8, 20.5, 18.4, -4.2, -4.9; **HRMS (ESI)**: calculated for $C_{45}H_{63}N_8O_{13}Si_2^+$: m/z = 979.4053 [M+H]$^+$; found: m/z = 979.4056.

Compound 30a
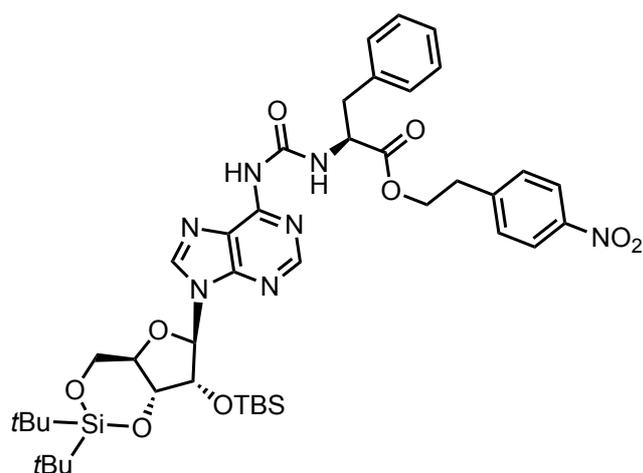
Yield: 68%; **IR**: $\tilde{v}$ = 3190 (w), 2933 (w), 2858 (w), 1742 (w), 1702 (vs), 1612 (s), 1587 (w), 1521 (vs), 1469 (vs), 1345 (vs), 1253 (s), 1057 (s), 1000 (w), 828 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ**: 10.00 (d, J = 7.5 Hz, 1H), 8.94 (s, 1H), 8.37 (d, J = 1.2 Hz, 2H), 8.00 (d, 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$_2$Cl$_2$) 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$_7$O$_9$Si$_2^+$: m/z = 862.3986 [M+H]$^+$; found: m/z = 862.3995 [M+H]$^+$.

Compound 30b
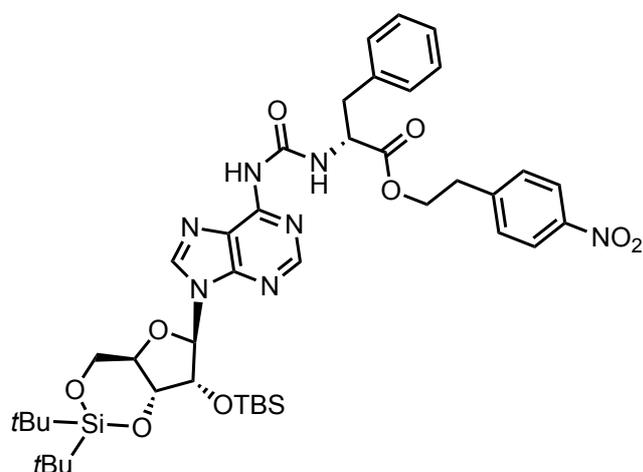
Eluent: Hex/EtOAc (4/1, v/v).

Yield: 71%; **IR**: $\tilde{v}$ = 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$_2$Cl$_2$) 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$_2$Cl$_2$) 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$_7$O$_9$Si$_2^+$: m/z = 862.3986 [M+H]$^+$; found: m/z = 862.3996 [M+H]$^+$.

Compound 31
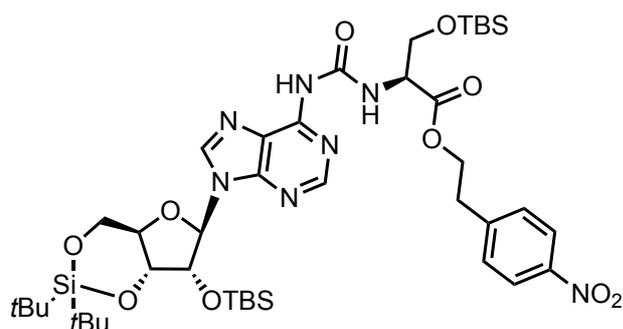
Eluent: Hex/EtOAc (1/1, v/v).

Yield: 85%; **IR**: $\tilde{v}$ = 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$_7$O$_{10}$Si$_3^+$: m/z = 916.4492 [M+H]$^+$; found: m/z = 916.4501 [M+H]$^+$.

General procedure for deprotection 3'-5'-silyl protecting group
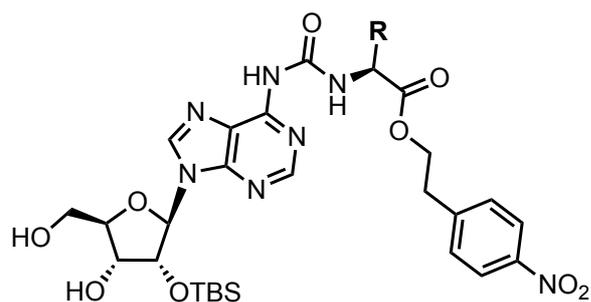
The modified adenosine (0.86 mmol) was dissolved in CH_2Cl_2 under N_2 atmosphere and transferred into a plastic flask. Pyridine (1 ml) was added and the solution was cooled in an ice-bath. Then $Py^*(HF)_n$ (140 µl) was added and the mixture was stirred at 0 °C for 2 h. The reaction was quenched with sat. $NaHCO_3$ and extracted with CH_2Cl_2. Organic phase was washed with water and dried over Na_2SO_4. The solvents were removed in vacuo. The crude product was purified by flash chromatography eluting with CH_2Cl_2/MeOH (9/1, v/v) to afford the product as a colourless foam.

Compound 32
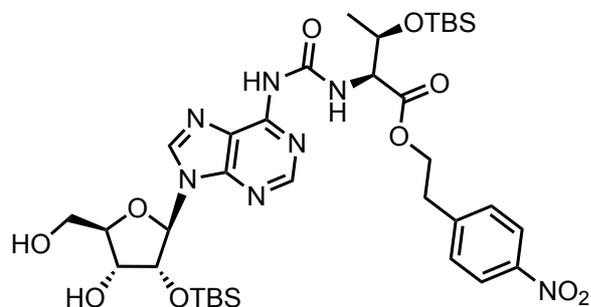
Yield: 95%; **IR**: $\tilde{v}$ = 3244 (w), 2952 (w), 2929 (w), 2856 (w), 1736 (w), 1695 (s), 1610 (s), 1588 (s), 1520 (vs), 1469 (s), 1345 (vs), 1313 (w), 1250 (vs), 1129 (w), 1093 (s), 835 (vs), 760 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 9.90 (d, J = 9.1 Hz, 1H), 8.50 (s, 1H), 8.46 (s, 1H), 8.10 – 8.08 (m, 3H), 7.37 (d, J = 8.6 Hz, 2H), 5.85 (d, J = 7.0 Hz, 1H), 5.10 (dd, J = 7.0, 4.7 Hz, 1H), 4.59 (dd, J = 9.1, 1.5 Hz, 1H), 4.53 – 4.41 (m, 2H), 4.41 – 4.34 (m, 2H), 4.34 – 4.26 (m, 1H), 3.98 (dd, J = 13.0, 1.5 Hz, 1H), 3.78 (dd, J = 13.0, 1.5 Hz, 1H), 3.07 (t, J = 6.7 Hz, 2H), 1.24 (d, J = 6.7 Hz, 3H), 0.89 (s, 9H), 0.81 (s, 9H), 0.05 (s, 3H), -0.06 (s, 3H), -0.14 (s, 3H), -0.34 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 170.9, 154.0, 151.3, 150.9, 149.3, 147.1, 145.6, 143.0, 130.0, 123.9, 91.5, 87.7, 74.9, 72.7, 68.8, 65.13, 63.3, 59.8, 35.0, 25.7, 21.3, 18.1,

18.0, -4.0, -5.0, -5.1, -5.2; **HRMS (ESI)**: calculated for $C_{35}H_{56}N_7O_{10}Si_2^+$: m/z = 790.3622 [M+H]$^+$; found: m/z = 790.3612 [M+H]$^+$.

Compound 33
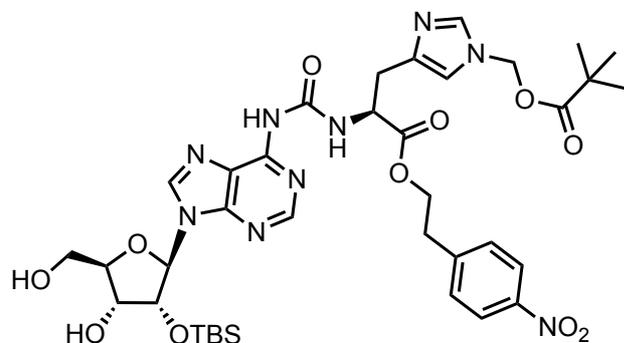
Yield: 75%; **IR**: $\tilde{v}$ = 3854 (w), 3745 (w), 3650 (w), 3230 (w), 2930 (w), 2857 (w), 2361 (w), 2341 (w), 1740 (s), 1699 (vs), 1611 (w), 1587 (w), 1520 (vs), 1472 (s), 1395 (w), 1345 (vs), 1253 (w), 1119 (vs), 1091 (w), 1058 (w), 1030 (w), 983 (w), 914 (w), 856 (vs), 779 (vs), 747(w), 670 (w) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 10.05 – 9.98 (m, 1H), 8.50 (d, J = 3.1 Hz, 1H), 8.18 – 8.10 (m, 3H), 8.01 (d, J = 9.1 Hz, 1H), 7.60 (dd, J = 8.2, 1.3 Hz, 1H), 7.44 – 7.37 (m, 2H), 6.84 (d, J = 1.4 Hz, 1H), 5.99 – 5.93 (m, 1H), 5.81 (dd, J = 7.3, 5.4 Hz, 1H), 5.77 – 5.71 (m, 2H), 5.08 (dt, J = 7.3, 5.1 Hz, 1H), 4.90 (dt, J = 7.2, 5.4 Hz, 1H), 4.42 (q, J = 6.3 Hz, 2H), 4.39 – 4.34 (m, 2H), 3.98 – 3.92 (m, 1H), 3.82 – 3.72 (m, 1H), 3.23 – 3.13 (m, 2H), 3.08 (dt, J = 13.8, 6.6 Hz, 2H), 2.83 (s, 1H), 1.09 (s, 9H), 0.80 (s, 9H), -0.18 (s, 3H), -0.40 (s, 3H); **^{13}C NMR (101 MHz, CDCl$_3$) δ**: 177.8, 171.8, 153.2, 151.1, 150.9, 149.3, 147.0, 145.8, 143.0, 138.2, 138.1, 130.0, 123.9, 123.1, 117.4, 91.4, 87.8, 74.7, 73.0, 67.7, 65.0, 63.5, 53.5, 38.8, 35.0, 30.6, 26.9, 25.7, 18.0, -5.1, -5.2; **HRMS (ESI)**: calculated for $C_{37}H_{52}N_9O_{11}Si^+$: m/z = 826.3550 [M+H]$^+$; found: m/z = 826.3559 [M+H]$^+$.

Compound 34
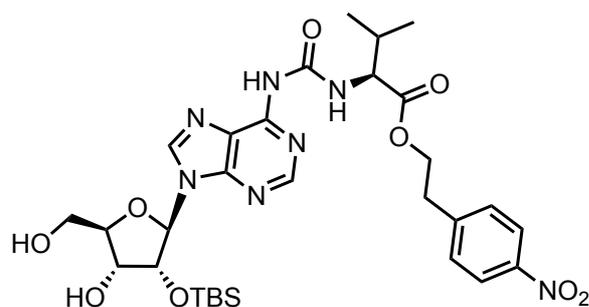
Yield: 95%; **IR**: ṽ = 3245 (w), 2930 (w), 2857 (w), 1743 (s), 1695 (vs), 1590 (s), 1518 (vs), 1471 (s), 1344 (vs), 1252 (s), 1212 (w), 1188 (s), 1130 (w), 1085 (s), 836 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.01 (d, J = 8.4 Hz, 1H), 8.50 (s, 1H), 8.17 (s, 1H), 8.07 (d, J = 8.3 Hz, 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$_7$O$_9$Si$^+$: m/z = 674.2970 [M+H]$^+$; found: m/z = 674.2974 [M+H]$^+$.

Compound 35
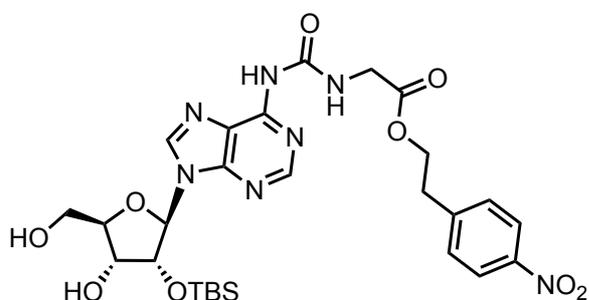
Yield: 97%; **IR**: ṽ = 3250 (w), 2932 (w), 2859 (w), 1740 (s), 1680 (vs), 1595 (s), 1520 (vs), 1470 (s), 1340 (vs), 1254 (s), 1213 (w), 1168 (s), 1130 (w), 1085 (s), 836 (vs) cm^{-1}; **^{1}H NMR (400 MHz, CDCl$_3$) δ**: 9.92 (t, J = 5.7 Hz, 1H), 8.92 (br. s, 1H), 8.54 (s, 1H), 8.25 (s, 1H), 8.13 (d, 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$_7$O$_9$Si$^+$: m/z = 632.2495 [M + H]$^+$; found m/z = 632.2492.

Compound 36
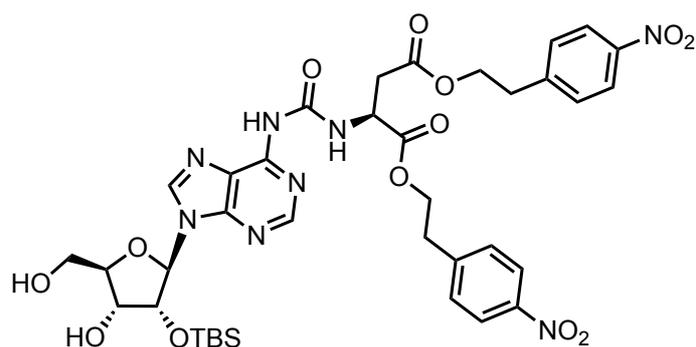
Yield: 98%; **IR**: $\tilde{v}$ = 3229 (w), 2953 (w), 2929 (w), 2857 (w), 1735 (s), 1693 (s), 1607 (s), 1589 (s), 1517 (vs), 1469 (s), 1391 (w), 1310 (vs), 1292 (w), 1251 (w), 1210 (w), 835 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.30 (d, J = 7.9 Hz, 1H), 9.41 (s, 1H), 8.77 (s, 1H), 8.50 (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$_8$O$_{13}$Si$^+$: m/z = 839.3032 [M+H]$^+$; found: m/z = 839.3041 [M+H]$^+$.

Compound 37a
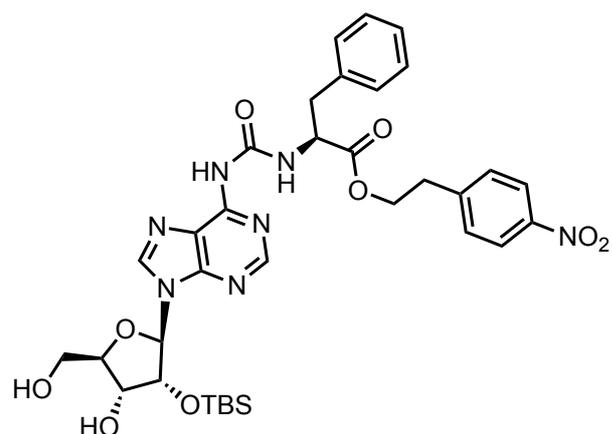
Yield: 80%; **IR**: $\tilde{v}$ = 3240 (w), 2929 (w), 2857 (w), 1742 (w), 1699 (vs), 1613 (s), 1520 (vs), 1471 (s), 1345 (vs), 1255 (w), 839 (s) cm^{-1}; **^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ**: 9.83 (d, J = 7.4 Hz, 1H), 8.63 (s, 1H), 8.40 (s, 1H), 8.18 (s, 1H), 8.10 (d, J = 8.7 Hz, 2H), 7.38 (d, J = 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$_2$Cl$_2$)** **δ**: 171.4, 153.1, 150.7, 150.5, 149.8, 149.3, 146.8, 145.8, 143.3, 136.3, 129.8, 129.4, 128.3, 127.2, 123.5, 91.0, 87.6, 74.8, 72.7, 64.7, 63.1, 54.9, 37.8, 34.7, 25.3, 17.7, -5.6, -5.7; **HRMS (ESI)**: calculated for $C_{34}H_{44}N_7O_9Si^+$: m/z = 722.2965 [M+H]$^+$; found: m/z = 722.2971 [M+H]$^+$.

Compound 37b
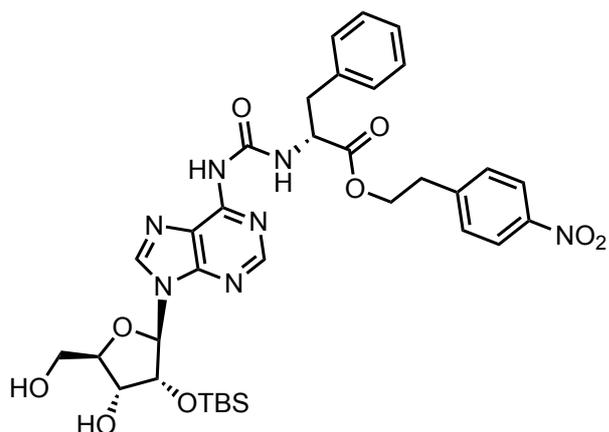
Yield: 78%; **IR**: $\tilde{v}$ = 3229 (w), 2930 (w), 2858 (w), 1740 w), 1697 (vs), 1612 (s), 1519 (vs), 1470 (s), 1345 (vs), 1254 (s), 1216 (w), 838 (vs), 781 (s), 736 (vs), 700 (s) cm^{-1}; **^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ**: 9.79 (br. s, 1H), 8.58 (br. s, 1H), 8.39 (s, 1H), 8.17 (t, J = 4.0 Hz, 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$_2$Cl$_2$) δ**: 171.4, 153.1, 150.7, 150.5, 149.8, 149.3, 146.8, 145.8, 143.3, 136.3, 129.8, 129.4, 128.5, 128.3, 127.2, 123.5, 91.0, 87.6, 74.8, 72.7, 64.7, 63.1, 54.9, 37.8, 34.7, 25.3, 17.7, -5.6, -5.7; **HRMS (ESI)**: calculated for $C_{34}H_{43}N_7O_9SiNa^+$: m/z = 744.2784 [M+Na]$^+$; found: m/z = 744.2776 [M+Na]$^+$.

Compound 38
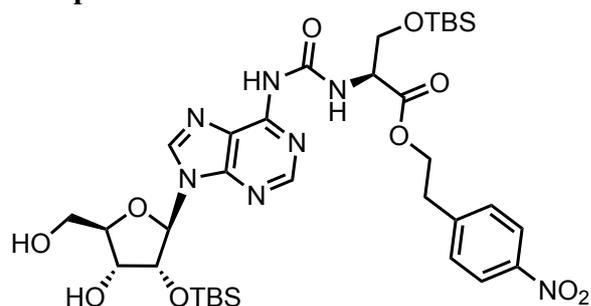
Yield: 96%; **IR**: $\tilde{v}$ = 3240 (w), 2950 (w), 2927 (w), 2856 (w), 1737 (w), 1695 (s), 1611 (s), 1589 (s), 1520 (vs), 1469 (s), 1346 (vs), 1313 (w), 1250 (vs), 1130 (w), 1093 (s), 835 (vs), 780 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.11 (d, J = 8.2 Hz, 1H), 9.13 (s, 1H), 8.72 (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$_7$O$_{10}$Si$_2$Na$^+$: m/z = 798.3290 [M+Na]$^+$; found: m/z = 798.3288 [M+Na]$^+$.

General procedure for DMT protection
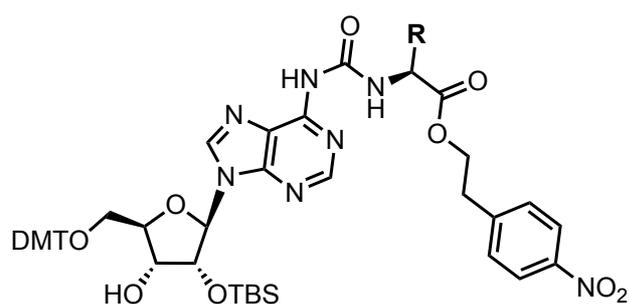
The 3'-5'-deprotected adenosine derivative (1 eq.) was dissolved in pyridine under N$_2$ atmosphere. DMT chloride (1.5 eq.) was added in two portions and the mixture was stirred at room temperature overnight. Then the volatiles were evaporated, and crude product was purified by flash chromatography eluting with CH$_2$Cl$_2$/MeOH (10/1/, v/v) with an addition of 0.1% of pyridine, unless otherwise specified, to afford the DMT protected derivative as white foam.

Compound 39
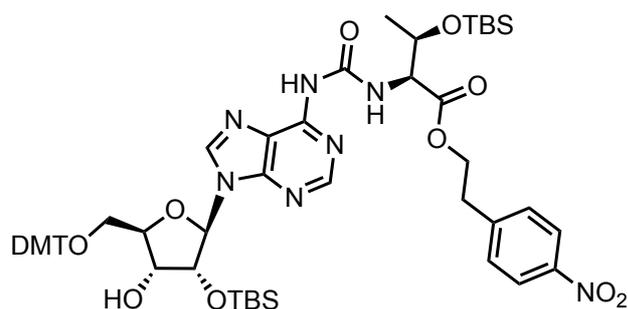
Yield: 90%; **IR**: $\tilde{v}$ = 3350 (w), 2930 (w), 2856 (w), 1729 (w), 1684 (s), 1608 (s), 1521 (vs), 1464 (s), 1345 (vs), 1248 (vs), 1174 (w), 1094 (w), 10033 (s), 827 (vs), 777 (vs) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.00 (d, J = 9.0 Hz, 1H), 8.95 (s, 1H), 8.55 (s, 1H), 8.40 (s, 1H), 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$_7$O$_{12}$Si$_2^+$: m/z = 1092.4929 [M+H]$^+$; found: m/z = 1092.4937 [M+H]$^+$.

Compound 40
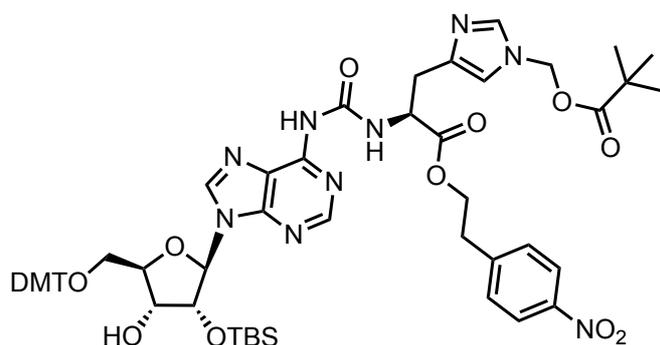
Eluent: 10% CH$_2$Cl$_2$ in EtOAc to pure EtOAc containing 0.1% of pyridine.

Yield: 85%; **IR**: $\tilde{v}$ = 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$_9$O$_{13}$Si$^+$: m/z = 1128.4857 [M+H]$^+$; found: m/z = 1128.4885 [M+H]$^+$.

Compound 41
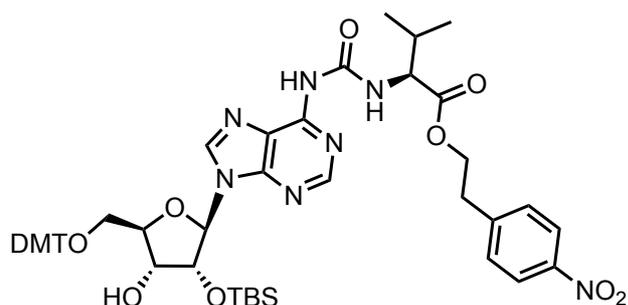
Yield: 70%; **IR**: $\tilde{v}$ = 2929 (w), 1735 (w), 1684 (w), 1569 (s), 1508 (vs), 1464 (s), 1344 (s), 1249 (vs), 1176 (w), 1015 (s), 800 (vs) 749 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.02 (d, J = 9.0 Hz, 1H), 8.55 (s, 1H), 8.51 (s, 1H), 8.11 (d, J = 8.7 Hz, 2H), 7.59 (d, J = 8.7 Hz, 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$_7$O$_{11}$Si$^+$: m/z = 976.4277 [M+H]$^+$; found: m/z = 976.4287 [M+H]$^+$.

Compound 42
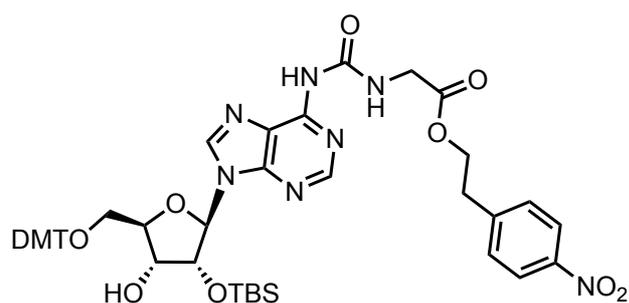
Yield: 85%; **IR**: $\tilde{v}$ = 2931 (w), 1748 (w), 1703 (s), 1609 (s), 1588 (w), 1509 (s), 1469 (s), 1345 (vs), 1250 (vs), 1177 (vs), 1035 (w), 835 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 9.93 (t, J = 5.5 Hz, 1H), 9.05 (br. s, 1H), 8.58 (s, 1H), 8.46 (s, 1H), 8.11 (d, J = 8.7 Hz, 2H), 7.58 (d, J = 8.7 Hz, 2H), 7.54 – 7.47 (m, 2H), 7.41 – 7.33 (m, 4H), 7.31 – 7.23 (m, 2H), 7.23 – 7.16 (m, 1H), 6.88 – 6.78 (m, 4H), 6.16 (d, J = 4.4 Hz, 1H), 5.10 (dd, J = 4.7 Hz, 1H), 4.53 (t, J = 5.1 Hz, 1H), 4.44 (t, J = 6.4 Hz, 2H), 4.30 (dt, J = 4.4 Hz, 1H), 4.13 (d, J = 5.5 Hz, 2H), 4.00 (d, J = 5.7 Hz, 1H), 3.75 (s, 6H), 3.53 – 3.42 (m, 2H), 3.13 (t, J = 6.4 Hz, 2H), 0.85 (s, 9H), 0.05 (s, 3H), -0.05 (s, 3H); **^{13}C NMR (101 MHz, Acetone-d_6) δ**: 170.6, 159.6, 154.6, 151.7, 151.3, 151.2, 147.6, 147.4, 146.1, 143.6, 136.7, 131.1, 130.9, 129.1, 129.0, 128.6, 127.5, 124.2, 121.7, 113.8, 90.0, 87.1, 84.8, 76.4, 72.0, 65.2, 64.4, 55.5, 42.6, 35.3, 26.1, 18.7, -4.7, -4.8; **HRMS (ESI)**: calculated for C$_{48}$H$_{55}$N$_7$O$_{11}$Si$^+$: m/z = 934.3802 [M + H]$^+$; found: m/z = 934.3812 [M + H]$^+$.

Compound 43
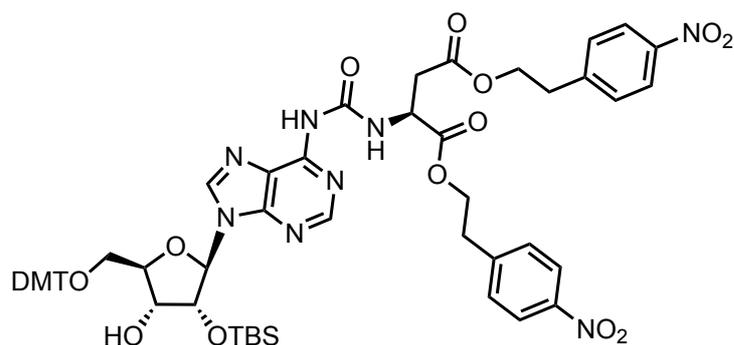
Yield: 95%; **IR**: $\tilde{v}$ = 2930 (w), 1734 (s), 1698 (s), 1607 (s), 1509 (vs), 1466 (s), 1370 (vs), 1248 (s), 1174 (s), 1032 (s), 829 (vs), 699 (s) cm^{-1}; **^{1}H NMR (400 MHz, Acetone-d_6) δ**: 10.29 (d, J = 7.9 Hz, 1H), 9.18 (s, 1H), 8.60 (s, 1H), 8.42 (s, 1H), 8.10 – 8.04 (m, 4H), 7.58 – 7.51 (m, 4H), 7.39 – 7.36 (m, 4H), 7.30 – 7.25 (m, 2H), 7.21 – 7.18 (m, 2H), 6.86 – 6.82 (m, 5H), 6.17

(d, J = 4.2 Hz, 1H), 5.10 (t, J = 4.5 Hz, 1H), 4.85 (dt, J = 7.9, 5.2 Hz, 1H), 4.56 (q, J = 5.2 Hz, 1H), 4.46 – 4.28 (m, 5H), 4.02 (d, J = 5.9 Hz, 1H), 3.75 (d, J = 2.1 Hz, 6H), 3.53 – 3.44 (m, 2H), 3.13 – 3.06 (m, 4H), 3.00 – 2.96 (m, 2H), 0.86 (s, 9H), 0.07 (s, 3H), 0.03 (s, 3H); ^{13}C NMR (101 MHz, Acetone-d_6) δ: 171.3, 171.2, 159.5, 154.2, 151.5, 150.6, 147.4, 147.2, 146.1, 143.6, 136.7, 131.0, 130.0, 129.0, 128.6, 127.5, 124.2, 113.8, 90.1, 87.0, 84.6, 76.4, 71.8, 65.7, 65.1, 64.3, 55.4, 50.6, 37.1, 35.2, 26.1, 18.6, -4.6, -4.8; **HRMS (ESI)**: calculated for $C_{58}H_{64}N_8O_{15}SiNa^+$: m/z = 1163.4158 [M+Na]$^+$; found: m/z = 1163.4159 [M+Na]$^+$.

Compound 44a
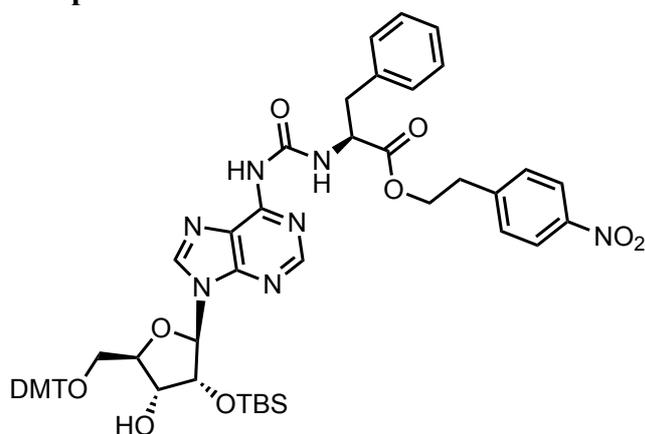
Eluent: CH$_2$Cl$_2$/MeOH (20/1, v/v).

Yield: 82%; **IR**: $\tilde{v}$ = 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$_2$Cl$_2$) 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$_2$Cl$_2$) δ**: 171.5, 153.3, 150.8, 150.6, 149.8, 149.4, 146.8, 145.8, 143.6, 136.2, 135.9, 129.9, 129.4, 128.5, 128.3, 127.2, 123.7, 123.6, 121.8, 91.0, 87.5, 74.8, 72.7, 64.8, 63.1, 54.8, 37.8, 34.7, 25.3, 17.7, -5.7; **HRMS (ESI)**: calculated for $C_{55}H_{62}N_7O_{11}Si^+$: m/z = 1024.4271 [M+H]$^+$; found: m/z = 1024.429 [M+H]$^+$.

Compound 44b
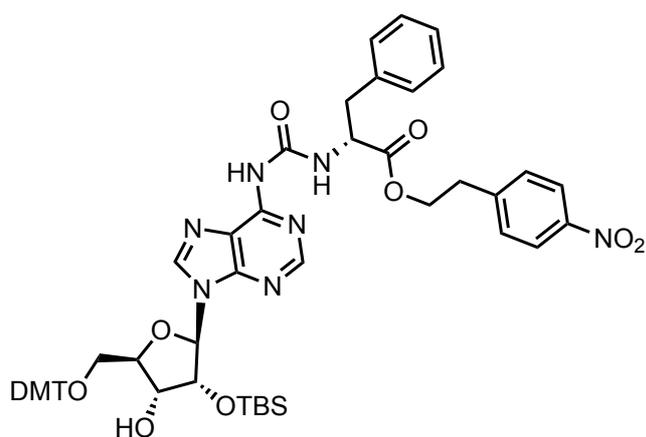
Eluent: CH$_2$Cl$_2$/MeOH (20/1, v/v).

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$_2$Cl$_2$) 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$_2$Cl$_2$) 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$_7$O$_{11}$Si$^+$: m/z = 1024.4271 [M+H]$^+$; found: m/z = 1024.4291 [M+H]$^+$.

Compound 45
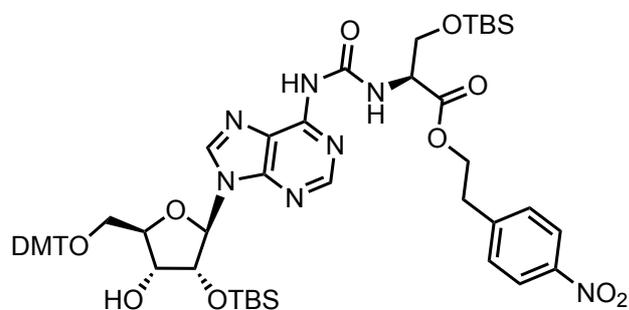
Yield: 90%; **IR**: ṽ = 2930 (w), 1740 (s), 1694 (s), 1607 (w), 1582 (w), 1519 (vs), 1465 (w), 1345 (s), 1249 (s), 1176 (w), 1110 (w), 1034 (w), 820 (vs), 778 (s), 699 (w) cm^{-1}; **^{1}H NMR**

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

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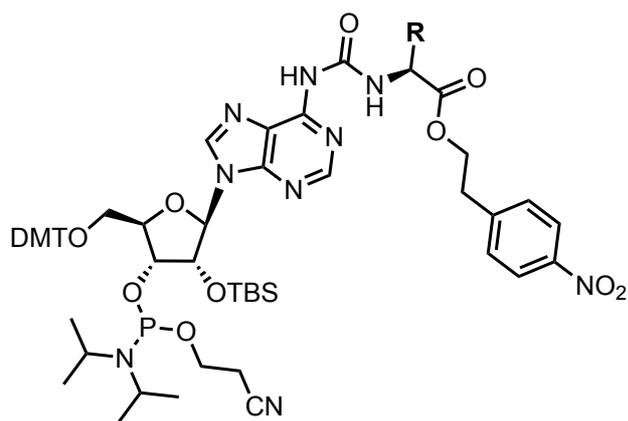
In an oven-dried flask under argon atmosphere, 5'-DMT protected compound (1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 °C. Hünig's base (4 eq.) was added dropwise followed by the addition of 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (2.5 eq.). The solution was stirred at room temperature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

Compound 46
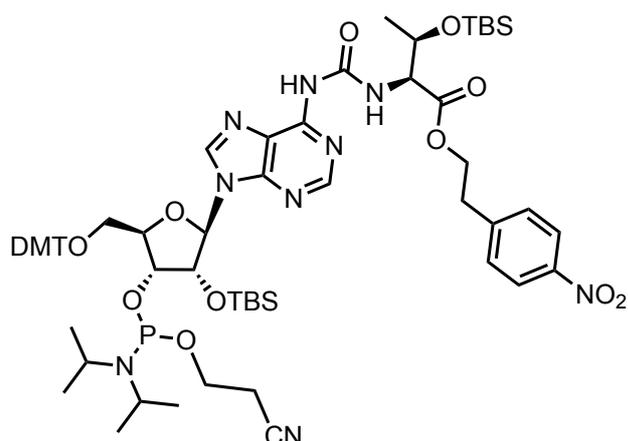
Yield: 85%; **^{31}P NMR (162 MHz, Acetone-d_6) δ**: 150.16, 148.45; **HRMS (ESI)**: calculated for $C_{65}H_{91}N_9O_{13}PSi_2^+$: m/z = 1292.6007 [M+H]$^+$; found: m/z = 1292.6033 [M+H]$^+$.

Compound 47
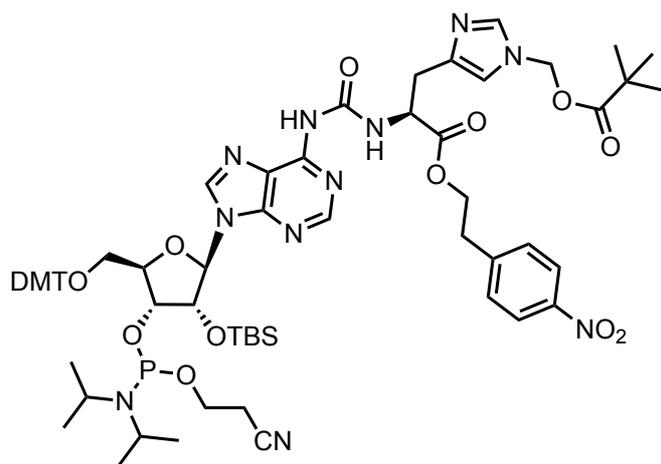
Eluent: 30% CH_2Cl_2 in EtOAc containing 0.1% of pyridine.

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

Compound 48
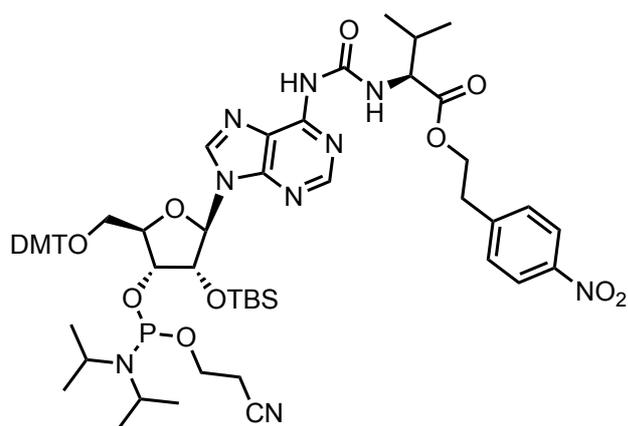
Yield: 87%; **^{31}P NMR (162 MHz, Acetone-d_6)** δ: 150.22, 148.65; **HRMS (ESI)**: calculated for $C_{60}H_{79}N_9O_{12}PSi^+$: m/z = 1176.5355 [M+H]$^+$; found: m/z = 1176.5359 [M+H]$^+$.

Compound 49
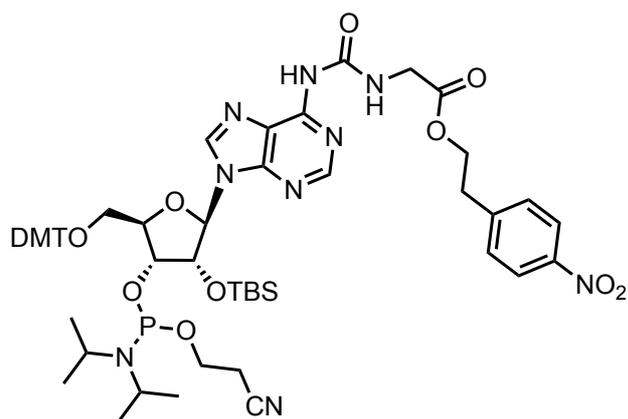
Yield: 86%; **^{31}P NMR (162 MHz, Acetone-d_6)** δ: 150.32, 148.60; **HRMS (ESI)**: calculated for $C_{57}H_{72}N_9O_{12}SiPSi^+$: m/z = 1134.4880 [M+H]$^+$; found: m/z = 1134.4894 [M + H]$^+$.

Compound 50
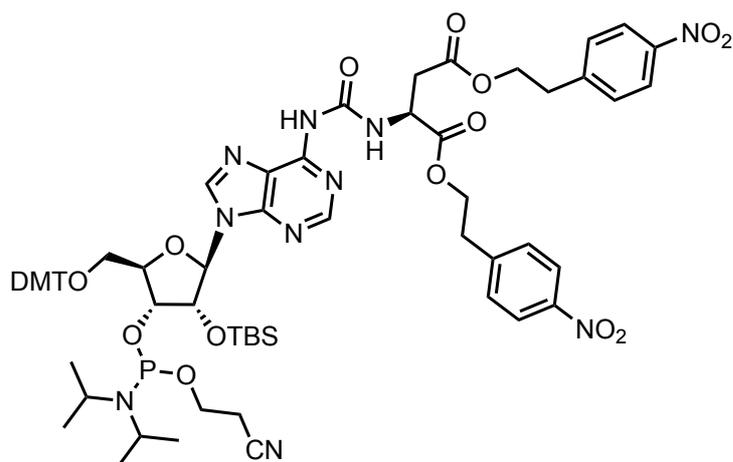
Yield: 65%; ^{31}**P NMR (162 MHz, Acetone-d_6) δ**: 150.15, 148.67; **HRMS (ESI)**: calculated for $C_{67}H_{82}N_{10}O_{16}PSi^+$: m/z = 1341.5417 [M+H]$^+$; found: m/z = 1341.5437 [M+H]$^+$.

Compound 51a
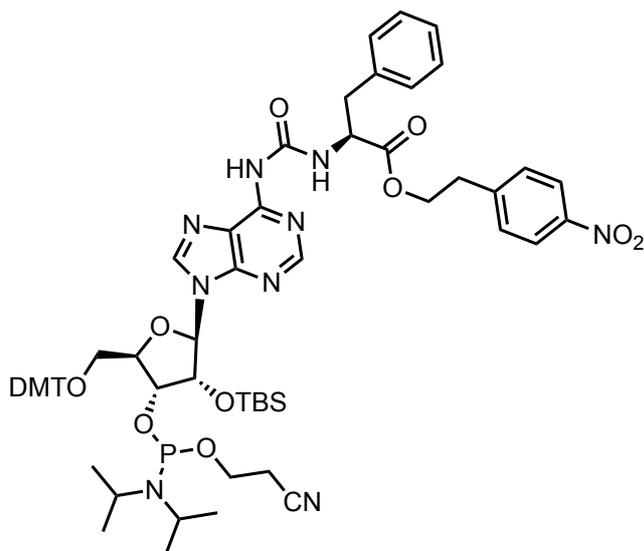
Yield: 67%; ^{31}**P NMR (162 MHz, CD$_2$Cl$_2$) δ**: 150.7, 149.1; **HRMS (ESI)**: calculated for $C_{64}H_{79}N_9O_{12}PSi^+$: m/z = 1224.5350 [M+H]$^+$; found: m/z = 1224.5374 [M+H]$^+$.

Compound 51b
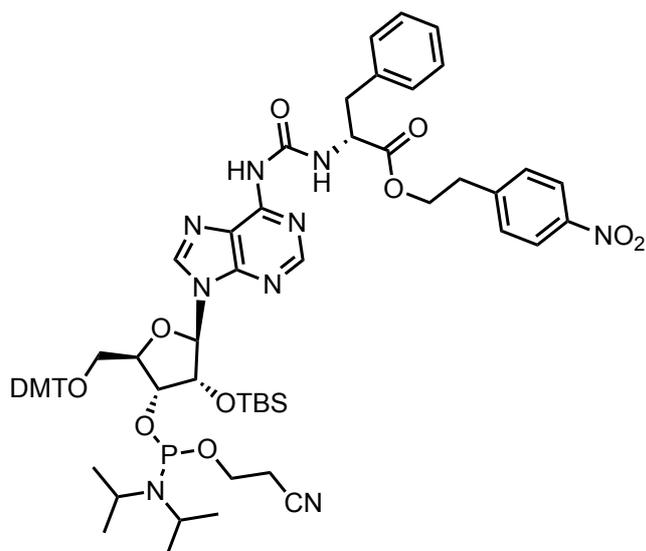
Yield: 67%; ^{31}**P NMR (162 MHz, CD$_2$Cl$_2$) δ**: 150.6, 148.9; **HRMS (ESI)**: calculated for C$_{64}$H$_{79}$N$_9$O$_{12}$PSi$^+$: m/z = 1224.5350 [M+H]$^+$; found: m/z = 1224.5383 [M+H]$^+$.

Compound 52
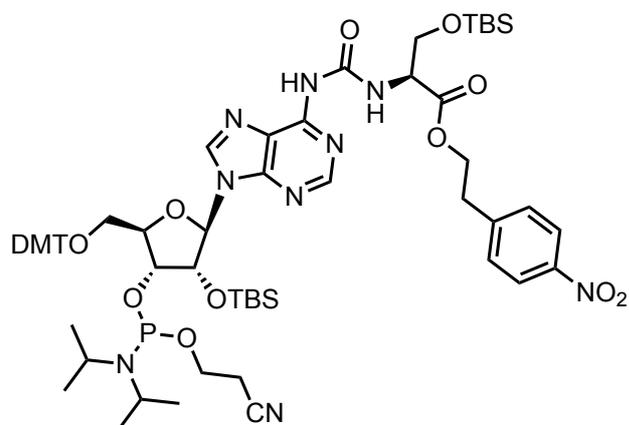
Yield: 70%; ^{31}**P NMR (162 MHz, Acetone-d_6) δ**: 150.34, 148.45; **HRMS (ESI)**: calculated for C$_{64}$H$_{89}$N$_9$O$_{13}$PSi$_2$$^+$: m/z = 1278.5856 [M+H]$^+$; found: m/z = 1278.5877 [M+H]$^+$.

2.2. Synthesis of DNA building block

Compound 54
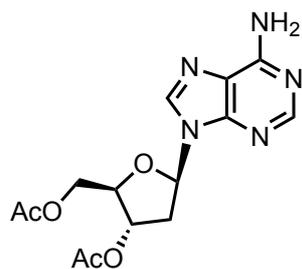
The compound was synthesized according to the published procedure.[5]

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

Yield: 99%; **¹H NMR (400 MHz, CDCl₃) δ**: 8.23 (s, 1H), 7.92 (s, 1H), 7.26 (s, 2H), 6.34 (dd, J = 7.9, 6.0 Hz, 1H), 5.34 – 5.32 (m, 1H), 4.32 – 4.24 (m, 3H), 2.87 – 2.80 (m, 1H), 2.57 – 2.52 (m, 1H), 2.04 (s, 3H), 2.00 (s, 3H). **HRMS (ESI)**: calculated for $C_{14}H_{18}N_5O_5^+$: m/z = 336.1302 [M+H]⁺; found: m/z = 336.1305 [M+H]⁺.

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

Compound 55
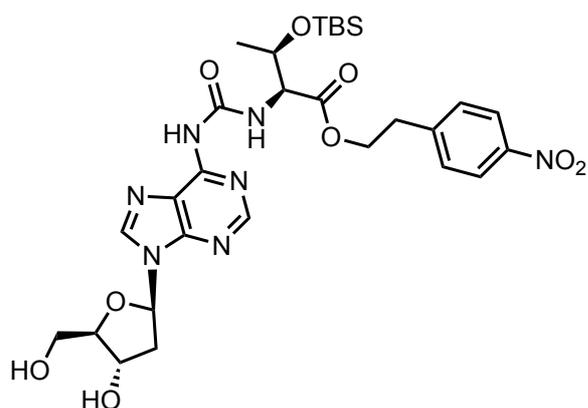
The acetyl-protected deoxyadenosine derivative **54** (0.5 g, 1.5 mmol, 1 eq.) was dissolved in dry CH₂Cl₂ under nitrogen atmosphere. 1-*N*-methyl-3-phenoxycarbonyl-imidazolium chloride (**22**; 0.71 g, 3 mmol, 2 eq.) was added to the reaction mixture and the resulting suspension was stirred at room temperature for 2 hours (the solution in time becomes clear). Afterwards the protected threonine derivative **7** (1.1 g, 3 mmol, 2 eq.) was added together with TEA (415 μl, 3 mmol, 2 eq.) as a solution in CH₂Cl₂ and the resulting solution was stirred overnight at room

temperature. The reaction was quenched by addition of saturated aqueous NaHCO$_3$ solution. The solution was extracted three times with CH$_2$Cl$_2$, and the organic phase was dried, filtered and concentrated *in vacuo*. The acetyl groups were immediately deprotected with 7N NH$_3$/MeOH (2 ml). After stirring 2 hours at room temperature, the reaction mixture was evaporated. The residue was purified via flash chromatography eluting with CH$_2$Cl$_2$/MeOH (10/1, v/v).

Yield: 65%; **IR**: $\tilde{v}$ = 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$_7$O$_9$Si$^+$: m/z = 660.2813 [M+H]$^+$; found: m/z = 660.2812 [M+H]$^+$.

Compound 56
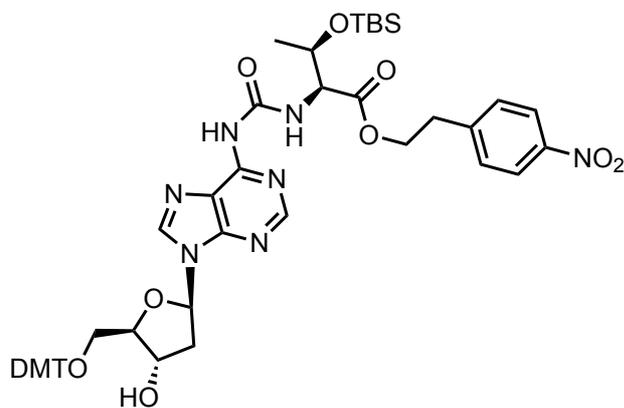
The 3'-5'-deprotected adenosine derivative **55** (0.34 g, 0.52 mmol, 1 eq.) was dissolved in pyridine under N$_2$ atmosphere. DMT chloride (0.26 g, 0.77 mmol, 1.5 eq.) was added in two portions and the mixture was stirred at room temperature overnight. Then the volatiles were evaporated, and crude product was purified by flash chromatography eluting with CH$_2$Cl$_2$/MeOH (10/1/, v/v) with an addition of 0.1% of pyridine to afford the DMT protected derivative.

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

Compound 57
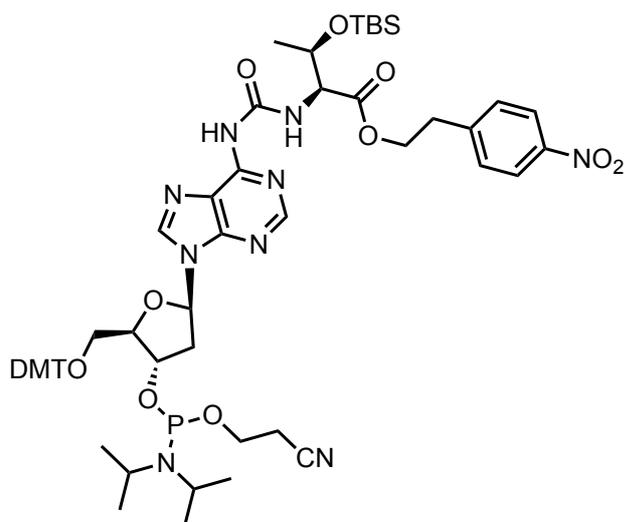
In an oven-dried flask under argon atmosphere, 5'-DMT protected compound **56** (0.1 g, 0.1 mmol, 1 eq.) was dissolved in CH_2Cl_2 and cooled to 0 ℃. Hünig's base (72 μl, 0.4 mmol, 4 eq.) was added dropwise followed by the addition of 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (60 μl, 0.25 mmol, 2.5 eq.). The solution was stirred at room temperature for 2 h. The reaction was quenched by addition of sat. NaHCO$_3$ solution, and then extracted three times with CH_2Cl_2. The organic phase was dried over Na$_2$SO$_4$, filtered and concentrated in vacuo. The residue was purified by flash chromatography, eluting with Hex/EtOAc (1/2, v/v) containing 0.1% of pyridine. The phosphoramidite was obtained as a mixture of diastereomers as white foam.

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

3. Synthesis and Purification of Oligonucleotides

All of the oligonucleotides used in this study were synthesized on a 1 μmol scale using a DNA automated synthesizer (Applied Biosystems 394 DNA/RNA Synthesizer) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleosides were purchased from Glen Research and Sigma-Aldrich. Oligonucleotides containing non-canonical nucleosides were synthesized in DMT-OFF mode using phosphoramidites (Bz-A, Dmf-G, Ac-C, U) with BTT in CH_3CN as an activator, DCA in CH_2Cl_2 as a deblocking solution and Ac_2O in pyridine/THF as a capping reagent. For deprotection of npe-group the solid support was treated with DBU solution (10% in THF, 1 ml) for 2 h at room temperature. Subsequently, the supernatant was removed, and the beads were washed with THF and dried under vacuum. The solid support was suspended in a mixture of aqueous ammonia and methylamine (1:1, 1 ml) at room temperature for 1 h. The supernatant was removed, and the beads were washed with water. The supernatant and washings were combined, and the solvents were evaporated under reduced pressure. The residue was subsequently heated with a solution of triethylamine trihydrofluoride (125 μl) in DMSO (50 μl) at 65 °C for 1.5 h. The RNA was precipitated by addition of aqueous NaOAc solution (3M, 25 μl) and *n*-butanol (1 ml). To ensure complete precipitation, the sample was incubated at -80 °C for 1 h. After centrifugation (4 °C, 4000 rpm, 15 min), the supernatant was removed, and the precipitated RNA was lyophilized. The oligonucleotides were further purified by reverse-phase HPLC using a Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel. Oligonucleotides were purified using the following buffer system: buffer A: 100 mM NEt_3/HOAc, pH 7.0 in H_2O and buffer B: 100 mM NEt_3/HOAc in 80 % (v/v) acetonitrile. A flow rate of 5 ml/min with a gradient of 0-25 % of buffer B in 30 min was applied for the purifications. Analytical RP-HPLC was performed on an analytical HPLC Waters Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-2 C18 from Macherey Nagel using a flow of 0.5 mL/min, a gradient of 0-30% of buffer B in 45 min was applied. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0 (Integrated DNA Technologies: https://eu.idtdna.com/calc/analyzer). For strands containing non-canonical base, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. The structural integrity of the synthesized oligonucleotides was analyzed by MALDI-TOF mass measurements.

Sequences of synthesized strands:

ON1: 5' GUCt^{6}ACCUGA 3'

ON2: 5' GUCg^{6}ACCUGA 3'

ON3: 5' GUCval^{6}ACCUGA 3'

ON4: 5' GUChis^{6}ACCUGA 3'

ON5: 5' GUCasp^{6}ACCUGA 3'

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

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

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

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

ODN1: 5' d(GTCt^{6}dACCTGA) 3'
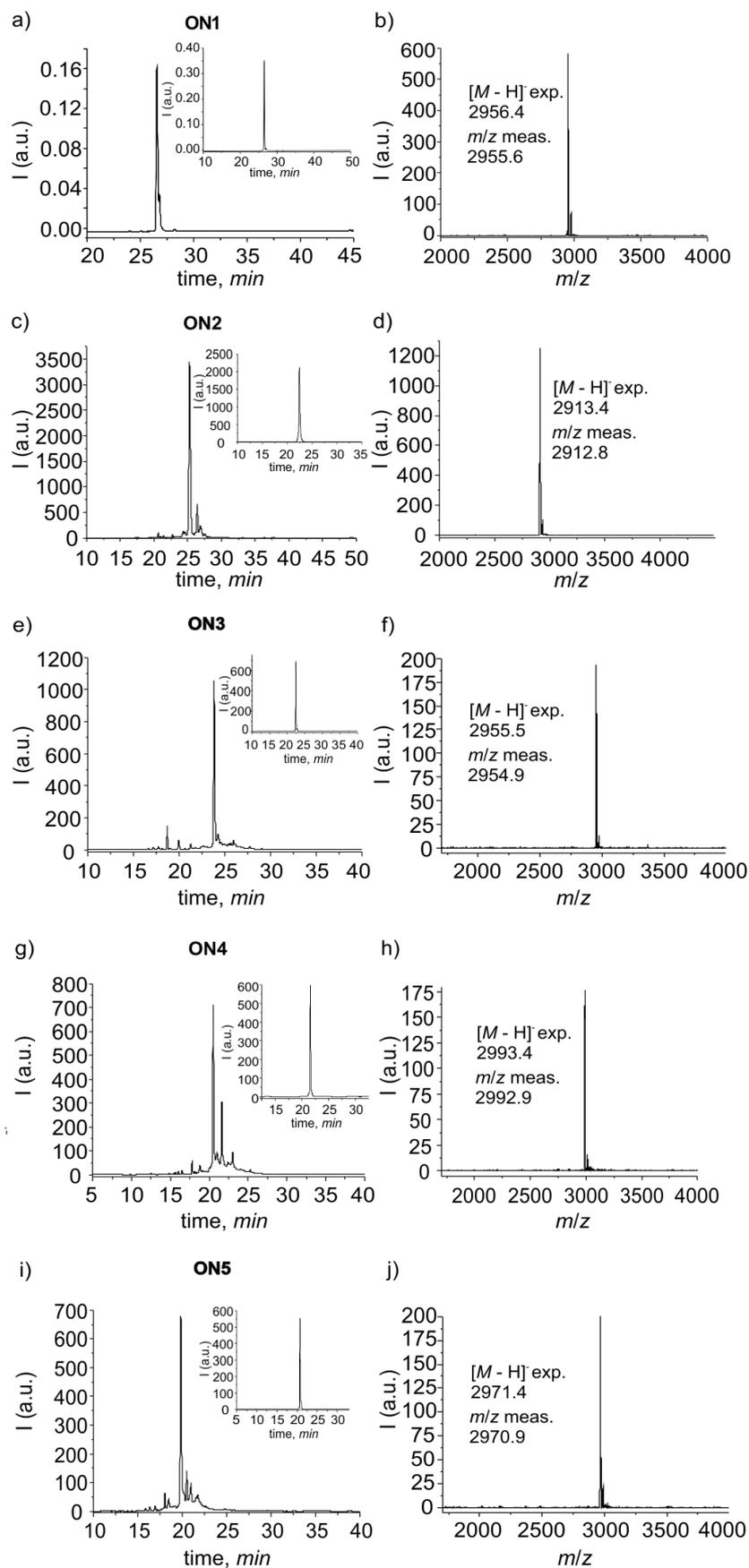
Figure S1. HPLC and MALDI data of oligonucleotides **ON1-ON5**.
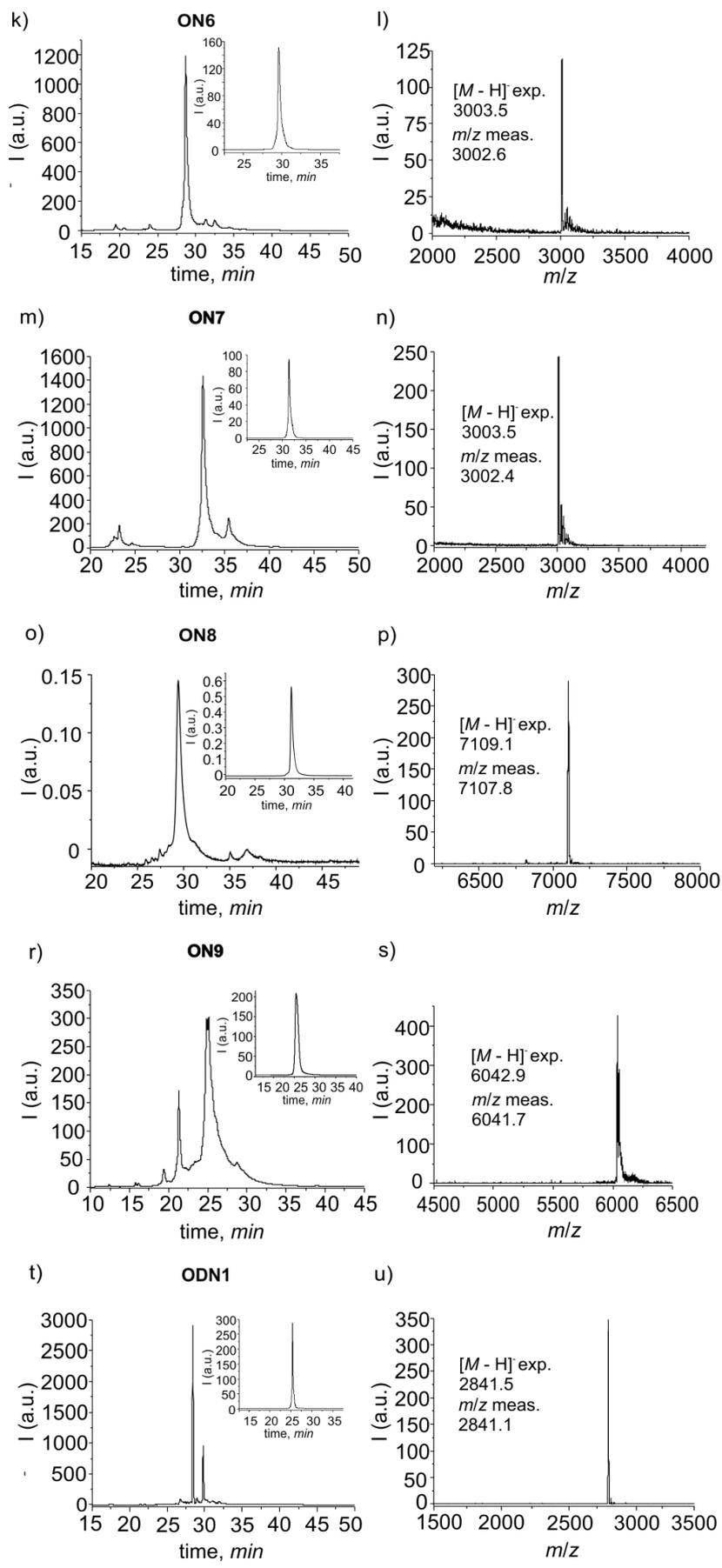
Figure S2. HPLC and MALDI data of oligonucleotides **ON6-ODN1**.

4. UV Melting Curve Measurements

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

5 NMR spectra

Compound 5

Compound 6

Compound 8

Compound 12

Compound 13

Compound 14a

Compound 14b

Compound 15

8.657

8.1774
8.1558

7.6261
7.6045

4.5231
4.5068
4.4954
4.4792
4.4629
4.4543
4.4389
4.4239
4.4112
4.3960

3.7664
3.7551

3.1285
3.1129
3.0971

2.1418
2.1245
2.1126
2.1072
2.0955
2.0901
2.0782
2.0611

0.8420
0.8346
0.8246
0.8174

2.91
2.00
2.02
2.01
0.97
1.96
0.98
2.84
2.86

168.7676

146.3384

130.3962
123.4502

65.2876
57.2031

33.8191
29.2389

18.3021
17.4450

Compound 16

Compound 17a

Compound 17b

Compound 19

Compound 20

Compound 21

Compound 25

Compound 26

Compound 27

Compound 28

Compound 29

Compound 30a

Compound 30b

Compound 31

Compound 32

Compound 33

Compound 34

Compound 35

Compound 36

Compound 37a

Compound 37b

Compound 38

Compound 39

Compound 40

Compound 41

229

Compound 42

Compound 43

Compound 44a

Compound 44b

Compound 45

Compound 46

150.1589
148.4478

250 230 210 190 170 150 130 110 90 70 50 30 10 −10 −30 −50 −
f1 (ppm)

150.27
148.61

Compound 47

260 250 240 230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 −10 −20 −30 −40 −50 −
f1 (ppm)

Compound 48

Compound 49

Compound 50

150.1514
148.6748

Compound 51a

150.7141
149.1259

Compound 51b

150.6164
148.9161

Compound 52

150.3381
148.4481

238

Compound 55

Compound 56

Compound 57

148.0022
146.5879

190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
−10

250 230 210 190 170 150 130 110 90 70 50 30 10 −10 −30 −50 −70 −90
f1 (ppm)

6. References

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

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

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

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

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

9 List of abbreviations

AA	N-arachidonoylethanolamine
A_m	2'-O-methyladenosine
aaRS	aminoacyl tRNA synthetase
AcOH	acetic acid
acp^{3}U	3-(3-amino-3-carboxy-n-propyl)uridine
Ago	argonaute
AIBN	2,2'-azobisisobutyronitrile
AMN	amino acid modified nucleobases
ASGPR	asialoglycoprotein
ASL	anticodon stem loop
Boc	$tert$-butoxycarbonyl
BOP	(benzotriazole-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate
BTT	5-(benzylthio)-1H-tetrazole
Chol	cholesterol
C_m	2'-O-methylcytidine
cRGD	cyclo(Arg-Gly-Asp)
ct^{6}A	cyclic N^6-threonylcarbamoyladenosine
CuAAC	copper catalyzed azide-alkyne cycloaddition
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCL1	Dicer-like 1
DCM	dichloromethane
ddH$_2$O	deionized water
DEA	diethylamine
DIPEA	diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	N,N-dimethylformamide
DMSO	dimethyl sulfoxide
DMTCl	dimethoxytritylchloride
DMTMM	4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride
DNA	deoxyribonucleic acid
DOTMA	N-[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride

dsRNA	double-stranded RNA
DTBS	di-*tert*-butylsilyl
E	envelope protein
E. coli	*Escherichia coli*
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
ESI	electrospray ionization
EtOAc	ethyl acetate
EWS-FLI1	Ewing's sarcoma fusion gene
FBS	fetal bovine serum
FDA	Food and Drug Administration
fMet	*N*-formyl methionine
g^6A	N^6-glycinylcarbamoyladenosine
GalNAc	*N*-acetylgalactosamine
Glc	glucose
Gly	glycine
GTP	guanosine-5'-triphosphate
hATTR	hereditary transthyretin amyloidosis
HBSS	Hanks' balanced salt solution
HBTU	2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBV	Hepatitis-B
HDL	high-density lipoproteins
HPLC	high-performance liquid chromatography
HRMS	high-resolution mass spectrometry
IFN-a	interferon-alpha
IGF-1	insulin-like growth factor 1
k^2C or L	lysidine
LDL	low-density lipoproteins
Leu	leucine
LNA	locked nucleic acid
lncRNA	long non-coding RNA
M	membrane protein
m^7G	7-methylguanosine
m^6t^6A	N^6-methyl-N^6-threonylcarbamoyladenosine

MALDI	matrix-assisted laser ionization
Met	methionine
miRNA	microRNA
mnm^{5}U	5'-methylaminomethyluridine
MOE	methoxyethyl
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
N	nucleocapsid protein
NBS	N-bromosuccinimide
nm^{5}U	5'-aminomethyluridine
NMR	nuclear magnetic resonance
npe	4-nitrophenylethyl
Nsp	non-structural proteins
nts	nucleotides
PA	phosphoramidite
PCR	polymerase chain reaction
PEI	polyethylenimine
Pivom^{5}U	5-pivaloylmethoxyuridine
PTC	peptidyl transferase center
RABV	rabies virus
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RPMI	Roswell Park Memorial Institute medium
rRNA	ribosomal RNA
RTC	replicase-transcriptase complex
S	spike protein
SARS-CoV-2	severe acute respiratory syndrome betacoronavirus 2
siRNA	small interfering RNA
SNALP	stable nucleic acid-lipid particle
snRNA	small nuclear RNA
ssRNA	single-stranded RNA
t^{6}A	N^6-threonylcarbamoyladenosine
TBS	*tert*-butylsilyl

TBTA	tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine
TEA	triethylamine
Thr	threonine
TLC	thin-layer chromatography
TLR	toll-like receptor
TRBP	transactivation response element RNA-binding protein
tRNA	transfer RNA
U_m	2'-*O*-methyluridine
UTR	untranslated region
Val	valine

10 References

[1]   P. F. Agris, *RNA* **2015**, *21*, 552-554.

[2]   P. F. Agris, *Nucleic Acids Res*. **2004**, *32*, 223-238.

[3]   B. S. Zhao, I. A. Roundtree, C. He, *Nat. Rev. Mol. Cell Biol*. **2017**, *18*, 31-42.

[4]   M. Duechler, G. Leszcynska, E. Sochacka, B. Nawrot, *Cell. Mol. Life Sci*. **2016**, *73*, 3075-3095.

[5]   W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. Vendeix, *Nucleic Acids Res*. **2011**, *39*, D195-D201.

[6]   M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res*. **2013**, *41*, D262-D267.

[7]   G. R. Björk, *Chem. Scr*. **1986**, *26B*, 91-95.

[8]   W. E. V. Cohn, V. Elliot, *Nature* **1951**, *167*, 483-484.

[9]   S. Takemura, H. Kasai, M. Goto, *J. Biochem*. **1974**, *75*, 1169-1172.

[10]   S. Diafa, M. Hollenstein, *Molecules* **2015**, *20*, 16643-16671.

[11]   P. E. Burmeister, S. D. Lewis, R. F. Silva, J. R. Preiss, L. R. Horwitz, P. S. Pendergrast, T. G. McCauley, J. C. Kurz, D. M. Epstein, C. Wilson, A. D. Keefe, *Chem. Biol*. **2005**, *12*, 25-33.

[12]   A. Hottin, A. Marx, *Acc. Chem. Res*. **2016**, *49*, 418-427.

[13]   F. Tolle, G. M. Brandle, D. Matzner, G. A. Mayer, *Angew. Chem. Int. Ed*. **2015**, *54*, 10971-10974.

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

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

[16]   J. Sulston, R. Lohrmann, L. E. Orgel, H. T. Miles, *Proc. Natl. Acad. Sci. USA* **1968**, *59*, 726-733.

[17]   J. Sulston, R. Lohrmann, L. E. Orgel, H. T. Miles, *Proc. Natl. Acad. Sci. USA* **1968**, 60, 409-415.

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

[19]   A. M. Poole, D. C. Jeffares, D. J. Penny, *Mol. Evol*. **1998**, 46, 1–17.

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

[21]   Y. I. Wolf, E. V. Koonin, *Biol. Direct*. **2007**, *2*, 14.

[22]   H. S. Bernhardt, W. P. Tate, *Biol. Direct*. **2012**, *7*, 4.

[23]   C. R. Woese, The genetic code, Harper and Row, New York, Evanston and London, **1967**.

[24]   C. Woese, *Nature* **1970**, *226*, 817-820.

[25]   M. Di Giulio, *Biol. Dir.* **2008**, *3*, 37.

[26]   M. Di Giulio, *J. Theor. Biol.* **1998**, *191*, 191-196.

[27]   H. Grosjean, V. de Crécy-Lagard, G. R. Björk, *Trends Biochem.* **2004**, *29*, 519-522.

[28]   E. Szathmary, *Trends Genet.* **1999**, *15*, 223-229.

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

[30]   M. Ishigami, K. Nagano, N. Tonotsuka, *Biosystems* **1977**, *9*, 229-243.

[31]   M. Paecht-Horowitz, M. J. Berger, A. Katchalsky, *Nature* **1970**, *228*, 636-639.

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

[33]   K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *36*, 842-845.

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

[35]   J. Maynard Smith, E. Szathmary, The origins of life, Oxford University Press: New York, NY, USA, **1999**.

[36]   S. Chatterjee, S. Yadav, *Life* **2019**, *9*, 25.

[37]   D. Kunnev, A. Gospodinov, *Life* **2018**, *8*, 44.

[38]   B. Zhang, T. R. Cech, *Nature* **1997**, *390*, 96-100.

[39]   R. M. Turk, N. V. Chumachenko, M. Yarus, *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 4585-4589.

[40]   A. Harish, G. Caetano-Anolles, *PLoS ONE* **2001**, *7*, e32776.

[41]   J. C. Bowman, N. V. Hud, J. D. Williams, *J. Mol. Evol.* **2015**, *80*, 143-161.

[42]   S. A. Chatterjee, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20033-20046.

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

[44]   H. F. Noller, On the origin of ribosome: Coevolution of subdomains of tRNA and rRNA. In *The RNA World*; Cold Spring Harbor Laboratory Press: Plainview, NY, USA, **1993**, 137-156.

[45]   R. Zhou, K. Basu, H. Hartman, C. J. Matocha, S. K. Sears, H. Vali, M. I. Guzman, *Sci. Rep.* **2017**, *7*, 533.

[46]   P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 920-929.

[47]   C. Orelle, E. D. Carlson, T. Szal, T. Florin, M. C. Jewett, A. S. Mankin, *Nature* **2015**, *524*, 119-124.

[48]   Y. Savir, T. Tlusty, *Cell* **2013**, *153*, 471-479.

[49]   T. M. Schmeing, V. Ramakrishnan, *Nature* **2009**, *461*, 1234-1242.

[50]   T. M. Schmeing, K. S. Huang, S. A. Strobel, T. A. Steitz, *Nature* **2005**, *438*, 520-524.

[51]   T. W. Dreher, *Wiley Interdiscip. Rev. RNA 1*, **2010**, 402-414.

[52]   M. Helm, J. D. Alfonzo, *Chem. Biol.* **2014**, *21*, 174-185.

[53]   P. F. Agris, A. Narendran, K. Sarachan, V. Y. P. Vare, E. Eruysal, *Enzymes* **2017**, *41*, 1-50.

[54]   F. Jühling, M. Mörl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, J. Pütz, *Nucleic Acids Res.* **2009**, *37*, D159-D162.

[55]   N. Ranjan, M. V. Rodnina, *Translation* **2016**, *4*, e1143076.

[56]   T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, S. Yokoyama, *Nature* **1988**, *336*, 179-181.

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

[58]   P. C. Thiaville, B. El Yacoubi, C. Köhrer, J. Thiaville, C. Deutsch, D. Iwata-Reuyl, J. M. Bacusmo, J. Armengaud, Y. Bessho, C. Wetzel, X. Cao, P. A. Limbach, U. L. RajBhandary, V. de Crécy-Lagard, *Mol. Microbiol.* **2015**, *98*, 1199-1221.

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

[60]   M. Sundaram, P. C. Durant, D. R. Davis, *Biochemistry (Mosc.)* **2000**, *39*, 12575-12584.

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

[62]   T. Niimi, O. Tureko, T. Yokogawa, N. Hayashi, K. Nishikawa, K. Watanabe, S. Yokoyama, *Nucleosides Nucleotides* **1994**, *13*, 1231-1237.

[63]   B. El Yacoubi, I. Hatin, C. Deutsch, T. Kahveci, J. P. Iwata-Reuyl, A. G. Murzin, V. de Crécy-Lagard, *EMBO J.* **2011**, *30*, 882-893.

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

[65]   F. Kimura-Harada, D. L. von Minden, J. A. McCloskey, S. Nishimura, *Biochemistry* **1972**, *11*, 3910-3915.

[66]   L. Clarke, J. Carbon, *J. Biol. Chem.* **1974**, *249*, 6874-6885.

[67]   Y. Komine, H. Inokuchi, *Nucleic Acids Res.* **1992**, *20*, 4089.

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

[69]   M. P. Schweizer, K. McGrath, L. Baczynskyj, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1046-1052.

[70]   M. Levitt, *Nature* **1969**, *224*, 759-763.

[71]   E. Y. Chen, B. A. Roe, *Biochem. Biophys. Res. Commun.* **1978**, *82*, 235-246.

[72]   J. W. Stuart, M. M. Basti, W. S. Smith, B. Forrest, R. Guenther, H. Sierzputowska-Grcz, B. Nawrot, A. Malkiewicz, P. F. Agris, *Nucleosides Nucleotides* **1996**, *15*, 1009-1028.

[73]   D. Bullinger, H. Neubauer, T. Fehm, S. Laufer, C. H. Gleiter, B. Kammerer, *BMC Biochem.* **2007**, *8*, 25.

[74]   S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet.* **2016**, *17*, 333–351.

[75]   J. C. Kaczmarek, P. S. Kowalski, D. G. Anderson, *Genome Medicine* **2017**, *9*, 60.

[76]   D. Castanotto, J. J. Rossi, *Nature* **2009**, *457*, 426-433.

[77]   C. F. Bennett, E. E. Swayze, *Annu. Rev. Pharmacol. Toxicol.* **2020**, *50*, 259-293.

[78]   A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, *Nature* **1998**, *391*, 806-811.

[79]   R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* **1993**, *75*, 843-854.

[80]   B. Wightman, I. Ha, G. Ruvkun, *Cell* **1993**, *75*, 855-862.

[81]   V. Ambros, B. Bartel, D. P. Bartel, C. B. Burge, J. C. Carrington, X. Chen, G. Dreyfuss, S. R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, T. Tuschl, *RNA-Publ. RNA Soc.* **2003**, *9*, 277-279.

[82]   V. N. Kim, J. Han, M. C. Siomi, *Nat. Rev. Mol. Biol.* **2009**, *10*, 126-139.

[83]   W. M. Jones-Rhoades, D. P. Bartel, B. Bartel, *Annu. Rev. Plant Biol.* **2006**, *57*, 19-53.

[84]   A. J. Hamilton, D. C. Baulcombe, *Science* **1999**, *286*, 950-952.

[85]   S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, T. Tuschl, *Nature* **2001**, *411*, 494-498.

[86]   D. de Paula, M. V. Bentley, R. I. Mahato, *RNA* **2007**, *13*, 431-456.

[87]   A. Khvorova, A. Reynolds, S. D. Jayasena, *Cell* **2003**, *115*, 209-216.

[88]   K. Nakanishi, *Wiley Interdiscip. Rev. RNA* **2016**, *7*, 637-660.

[89]   Y. Tomari, C. Matranga, B. Haley, N. Martinez, P. D. Zamore, *Science* **2004**, *306*, 1377-1380.

[90]   J. B. Ma, K. Ye, D. J. Patel, *Nature* **2004**, *429*, 318-322.

[91]   J. Liu, M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J. J. Song, S. M. Hammond, L. Joshua-Tor, G. J. Hannon, *Science* **2004**, *305*, 1437-1441.

[92]   P. Muhonen, T. Tennilä, E. Azhayeva, R. N. Parthasarathy, A. J. Janckila, H. K. Väänänen, A. Azhayev, T. Laitala-Leinonen, *Chem. Biodivers*. **2007**, *4*, 858-873.

[93]   J. H. Park, C. Shin, *Nucleic Acids Res*. **2015**, *43*, 9418-9433.

[94]   K. A. Whitehead, R. Langer, D. G. Anderson, *Nat. Rev. Drug Discov*. **2009**, *8*, 129-138.

[95]   S. M. Elbashir, J. Martinez, A. Patkaniowska, W. Lendeckel, T. Tuschl, *EMBO J*. **2001**, *20*, 6877-6888.

[96]   M. Amarzguioui, T. Holen, E. Babaie, H. Prydz, *Nucleic Acids Res*. **2003**, *31*, 589-595.

[97]   N. J. Caplen, S. Parrish, F. Imani, A. Fire, R. A. Morgan, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9742-9747.

[98]   T. Holen, M. Amarzguioui, M. T. Wiiger, E. Babaie, H. Prydz, *Nucleic Acids Res*. **2002**, *30*, 1757-1766.

[99]   G. Hutvagner, P. D. Zamore, *Science* **2002**, *297*, 2056-2060.

[100]   T. Tuschl, P. D. Zamore, R. Lehmann, D. P. Bartel, P. A. Sharp, *Genes Dev.* **1999**, *13*, 3191-3197.

[101]   A. L. Jackson, S. T. Bartz, J. Schelter, S. B. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, P. S. Linsley, *Nat. Biotechnol*. **2003**, *21*, 635-637.

[102]   J. Brennecke, A. Stark, R. B. Russel, S. M. Cohen, *PLoS Biol*. **2005**, *3*, e85.

[103]   K. Kai-How Farh, A. Grimson, C. Jan, B. P. Lewis, W. K. Johnston, L. P. Lim, C. B. Burge, D. P. Bartel, *Science* **2005**, *310*, 1817-1821.

[104]   D. Grun, Y. L. Wang, D. Langenberger, K. C. Gunsalus, N. Rajewsky, *PLoS Comput. Biol*. **2005**, *1*, e13.

[105]   A. Krek, D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsulus, M. Stoffel, N. Rajewsky, *Nat. Genet*. **2005**, *37*, 495-500.

[106]   D. Semizarov, L. Frost, A. Sarthy, P. Kroeger, D. N. Halbert, S. W. Fesik, *Proc. Natl. Acad. Sci*. **2003**, *10*, 6347-6352.

[107]   S. P. Persengiev, X. Zhu, M. R. Green, *RNA* **2004**, *10*, 12-18.

[108]   A. Judge, I. MacLachlan, *Hum. Gene Ther*. **2008**, *19*, 111-124.

[109]   M. Schlee, V. Hornung, G. Hartmann, *Mol. Ther*. **2006**, *14*, 463-470.

[110]   S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, T. Tuschl, *Nature*, **2001**, *411*, 6545-6549.

[111]   K. Kariko, P. Bhuyan, J. Capodici, D. Weismann, *J. Immunol*. **2004**, *172*, 6545-6549.

[112] M. Sioud, D. R. Sorensen, *Biochem. Biophys. Res. Commun.* **2003**, *312*, 1220-1225.

[113] V. Hornung, M. Guenthner-Biller, C. Bourquin, A. Ablasser, M. Schlee, S. Uematsu, A. Noronha, M. Manoharan, S. Akira, A. de Fougerolles, S. Endres, G. Hartmann, *Nat. Med.* **2005**, *11*, 263-270.

[114] A. D. Judge, V. Sood, J. R. Shaw, D. Fang, K. McClintock, I. MacLachlan, *Nat. Biotech.* **2005**, *23*, 457-462.

[115] S. S. Diebold, T. Kaisho, H. Hemmi, S. Akira, C. Reis e Sousa, *Science* **2004**, *303*, 1526-1529.

[116] F. Heil, H. Hemmi, H. Hochrein, F. Amperberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, S. Bauer, *Science* **2004**, *303*, 1526-1529.

[117] M. Sioud, *J. Mol. Biol.* **2005**, *348*, 1079-1090.

[118] W. Greulich, M. Wagner, M. M. Gaidt, C. Stafford, Y. Cheng, A. Linder, T. Carell, V. Hornung, *Cell* **2019**, *179*, 1264-1275.

[119] M. Sioud, *Eur. J. Immunol.* **2006**, *36*, 1222-1230.

[120] M. Manoharan, *Curr. Opin. Chem. Biol.* **2004**, *8*, 570-579.

[121] M. A. Behlke, *Oligonucleotides* **2008**, *18*, 305-319.

[122] C. Chen, Z. Yang, X. Tang, *Med. Res. Rev.* **2018**, *38*, 829-869.

[123] F. Eberle, K. Giessler, C. Deck, K. Heeg, M. Peter, C. Richert, A. H. Dalpke, *J. Immunol.* **2008**, *180*, 3229-3237.

[124] Ya-Lin Chiu, T. M. Rana, *Mol. Cell* **2002**, *10*, 549-561.

[125] R. Breslow, W. H. Jr. Chapman, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10018-10021.

[126] T. P. Prakash, C. R. Allerson, P. Dande, T. A. Vickers, N. Sioufi, R. Jarres, B. F. Baker, E. E. Swayze, R. H. Griffey, B. Bhat, *J. Med. Chem.* **2005**, *48*, 4247-4253.

[127] K. Garber, *Nat. Biotechnol.* **2016**, *34*, 1213-1214.

[128] A. Khvorova, *N. Engl. J. Med.* **2017**, *376*, 4-7.

[129] K. K. Ray, U. Landmesser, L. A. Leiter, D. Kallend, R. Dufour, M. Karakas, T. Hall, R. P. T. Troquay, T. Turner, F. L. J. Visseren, P. Wijngaard, R. S. Wright, J. J. P. Kastelein, *N. Engl. J. Med.* **2017**, *376*, 1430-1440.

[130] F. Eckstein, *J. Am. Chem. Soc.* **1970**, *92*, 4718-4723.

[131] F. Eckstein, *Nucleic Acids Ther.* **2014**, *24*, 374-387.

[132] J. Elmén, H. Thonberg, K. Ljungberg, M. Frieden, M. Westergaard, Y. Xu, B. Wahren, Z. Liang, H. Ørum, T. Koch, C. Wahlestedt, *Nucleic Acids Res.* **2005**, *33*, 439-447.

[133] A. H. Hall, J. Wan, A. Spesock, Z. Sergueeva, B. R. Shaw, K. A. Alexander, *Nucleic Acids Res*. **2006**, *34*, 2773-2781.

[134] J. Soutschek, A. Akinc, B. Bramlage, K. Charisse, R. Constien, M. Donoghue, S. Elbashir, A. Geick, P. Hadwiger, J. Harborth, M. John, V. Kesavan, G. Lavine, R. K. Pandey, T. Racie, K. G. Rajeev, I. Rohl, I. Toudjarska, G. Wang, S. Wuschko, D. Bumcrot, V. Koteliansky, S. Limmer, M. Manoharan, H.-P. Vornlocher, *Nature* **2004**, *432*, 173-178.

[135] D. S. Schwarz, Y. Tomari, P. D. Zamore, *Curr. Biol*. **2004**, *14*, 787-791.

[136] S. Wang, N. Allen, T. A. Vickers, A. S. Revenko, H. Sun, X. H. Liang, S. T. Crooke, *Nucleic Acids Res*. **2018**, *46*, 3579-3594.

[137] M. Lee, A. D. Simon, C. A. Stein, L. E. Rabbani, *Antisense Nucleic Acid Drug. Dev*. **1999**, *9*, 487-492.

[138] S. Weitzer, J. Martinez, *Nature* **2007**, *447*, 222-226.

[139] F. Frank, N. Sonenberg, B. Nagar, *Nature* **2010**, *465*, 818-822.

[140] C. R. Allerson, N. Sioufi, R. Jarres, T. P. Prakash, N. Naik, A. Berdeja, L. Wanders, R. H. Griffey, E. E. Swayze, B. Bhat, J. *Med. Chem*. **2005**, *48*, 901-904.

[141] P. Y. Chen, L. Weinmann, D. Gaidatzis, Y. Pei, M. Zavolan, T. Tuschl, G. Meister, *RNA* **2008**, *14*, 263-274.

[142] D. M. Kenski, G. Butora, A. T. Willingham, A. J. Cooper, W. Fu, N. Qi, F. Soriano, I. W. Davis, M. Flanagan, *Mol. Ther. Nucleic Acids* **2012**, 1:e5.

[143] W. F. Lima, T. P. Prakash, T. P. Murray, G. A. Kinberger, W. Li, A. E. Chapell, C. S. Li, S. F. Murray, H. Gaus, P. P. Seth, E. E. Swayze, S. T. Crooke, *Cell* **2012**, *150*, 883-894.

[144] E. Elkayam, R. Parmer, C. R. Brown, J. L. Willoughby, C. S. Theile, M. Manoharan, L. Joshua-Tor, *Nucleic Acids Res*. **2017**, *45*, 3528-3536.

[145] T. P. Prakash, W. F. Lima, H. M. Murray, W. Li, G. A. Kinberger, A. E. Chappell, H. Gaus, P. P. Seth, B. Bhat, S. T. Crooke, E. E. Swayze, *Nucleic Acids Res*. **2015**, *43*, 2993-3011.

[146] V. P. Torchilin, *Nature Rev. Drug. Discov*. **2005**, *4*, 145-160.

[147] I. MacLachlan in *Antisense Drug Technology: Principles, Strategies and Applications*. 2nd edn Ch. 9 (ed. S. T. Crooke) 237-270 (CRC, Boca Raton, 2007).

[148] P. L. Felgner, T. R. Gadek, M. Holm, R. Roman, H. W. Chan, M. Wenz, J. P. Northrop, G. M. Ringold, M. Danielsen, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7413-7417.

[149] A. Mohammadi, L. Kudsiova, M. F. M. Mustapa, F. Campbell, D. Vlaho, K. Welser, H. Story, A. D. Tagalakis, S. L. Hart, D. J. Barlow, A. B. Tabor, M. J. Lawrence, H. C. Hailes, *Org. Biomol. Chem*. **2019**, *17*, 945-957.

[150] D. V. Morrissey, J. A. Lockridge, L. Shaw, K. Blanchard, K. Jense, W. Breen, K. Hartsough, L. Machemer, S. Radka, V. Jadhav, N. Vaish, S. Zinnen, C. Vargeese, K. Bowman, C. S. Shaffer, L. B. Jeffs, A. Judge, I. MacLachlan, B. Polisky, *Nat. Biotechnol*. **2005**, *23*, 1002-1007.

[151] H. Lv, S. Zhang, B. Wang, S. Cui, J. Yan, *J. Control. Release* **2006**, *114*, 100-109.

[152] Z. Ma, J. Li, F. He, A. Wilson, B. Pitt, S. Li, *Biochem. Biophys, Res. Commun*. **2005**, *330*, 755-759.

[153] S. Akhtar, I. Benter, *Adv. Drug Deliv. Rev*. **2007**, *59*, 164-182.

[154] S. Akhtar, I. F. Benter, *J. Clin. Invest*. **2007**, *117*, 3623-3632.

[155] A. J. Hollins, Y. Omidi, I. F. Benter, S. Akhtar, *J. Drug Target*. **2007**, *15*, 83-88.

[156] A. Akinc, M. A. Maier, M. Monoharan, K. Fitzgerald, M. Jayaraman, S. Barros, S. Ansell, X. Du, M. J. Hope, T. D. Madden, B. L. Mui, S. C. Semple, Y. K. Tam, M. Ciufolini, D. Witzigmann, J. A. Kulkarni, R. van der Meel, P. R. Cullis, *Nat. Nanotechnol*. **2019**, *14*, 1084-1087.

[157] S. M. Hoy, *Drugs* **2018**, *78*, 1625-1631.

[158] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7297-7301.

[159] D. Putnam, *Nature Mater*. **2006**, *5*, 439-451.

[160] A. Zintchenko, A. Philipp, A. Dehshahri, E. Wagner, *Bioconjug. Chem*. **2008**, *19*, 1448-1455.

[161] P. H. Tan, L. C. Yang, H. C. Shih, K. C. Lan, J. T. Cheng, *Gene Ther*. **2004**, *12*, 59-66.

[162] B. Urban-Klein, S. Werth, S. Abuharbeid, F. Czubayko, A. Aigner, *Gene Ther*. **2005**, *12*, 461-466.

[163] A. Kichler, *J. Gene Med*. **2004**, *6*, S3-S10.

[164] R. Kircheis, L. Wightman, E. Wagner, *Adv. Drug Deliv. Rev*. **2001**, *53*, 341-358.

[165] S. Hu-Lieskovan, J. D. Heidel, D. W. Bartlett, M. E. Davis, T. J. Triche, *Cancer Res*. **2005**, *65*, 8984-8992.

[166] M. Thomas, S. A. Kularatne, L. Qi, P. Kleindl, C. P. Leamon, M. J. Hansen, P. S. Low, *Ann. N. Y. Acad. Sci*. **2009**, *1175*, 32-39.

[167] E. Song, P. Zhu, S. K. Lee, D. Chowdhury, S. Kussman, D. M. Dykxhoorn, Y. Feng, D. Palliser, D. B. Weiner, P. Shankar, W. A. Marasco, J. Lieberman, *Nat. Biotechnol.* **2005**, *23*, 709-717.

[168] J. P. Dassie, X. Y. Liu, G. S. Thomas, R. M. Whitaker, K. W. Thiel, K. R. Stockdale, D. K. Meyerholz, A. P. McCaffrey, J. O. McNamara, P. H. Giangrande, *Nat. Biotechnol.* **2009**, *27*, 839-846.

[169] C. F. Xia, R. J. Boado, W. M. Pardridge, *Mol. Pharm.* **2009**, *6*, 747-751.

[170] G. Cesarone, O. P. Edupuganti, C. P. Chen, E. Wickstrom, *Bioconjug. Chem*. **2007**, *18*, 1831-1840.

[171] S. Lau, B. Graham, N. Cao, B. J. Boyd, C. W. Pouton, P. J. White, *Mol. Pharm*. **2012**, *9*, 71-80.

[172] N. Aronin, *Gene Ther*. **2006**, *13*, 509-516.

[173] J. O. McNamara II, E. R. Andrechek, Y. Wang, K. D. Viles, R. E. Rempel, E. Gilboa, B. A. Sullenger, P. H. Giangrande, *Nat. Biotechnol*. **2006**, *24*, 1005-1015.

[174] J. K. Nair, J. L. Willoughby, A. Chan, K. Charisse, M. R. Alam, Q. Wang, M. Hoekstra, P. Kandasamy, A. V. Kel'in, S. Milstein, N. Taneja, J. O'Shea, S. Kuchimanchi, K. Fitzgerald, T. Zimmermann, T. J. C. van Berkel, M. A. Maier, K. G. Rajeev, M. Manoharan, *J. Am. Chem. Soc*. **2014**, *136*, 16958-16961.

[175] C. Lorenz, P. Hadwiger, M. John, H. P. Vornlocher, C. Unverzagt, *Bioorg. Med. Chem. Lett*. **2004**, *14*, 4975-4977.

[176] A. Biscans, A. Coles, R. Haraszti, D. Echeverria, M. Hassler, M. Osborn, A. Khvorova, *Nucleic Acids Res*. **2019**, *47*, 1082-1096.

[177] K. Nishina, T. Unno, Y. Uno, T. Kubodera, T. Kanouchi, H. Misuzawa, T. Yokota, *Mol. Ther*. **2008**, *16*, 734-740.

[178] A. Kwiatkowska, M. Sobczak, B. Mikolajczyk, S. Janczak, A. B. Olejniczak, M. Sochacki, Z. J. Lesnikowski, B. Nawrot, *J. Am. Chem. Soc*. **2013**, *24*, 1017-1026.

[179] B. R. Meade, K. Gogoi, A. S. Hamil, C. Palm-Apergi, A. van den Berg, J. C. Hagopian, A. D. Springer, A. Eguchi, A. D. Kacsinta, C. F. Dowdy, A. Presente, P. Lönn, M. Kaulich, N. Yoshioka, E. Gros, X.-S. Cui, S. F. Dowdy, *Nat. Biotechnol*. **2014**, *32*, 1256-1261.

[180] J. J. Turner, G. D. Ivanova, B. Verbeure, D. Williams, A. A. Arzumanov, S. Abes, B. Lebleu, M. J. Gait, *Nucleic Acids Res*. **2005**, *33*, 6837-6849.

[181] I. Dovydenko, I. Tarassov, A. Venyaminova, N. Entelis, *Biomaterials* **2016**, *76*, 408-417.

[182]   D. B. Rozema, D. L. Lewis, D. H. Wakefield, S. C. Wong, J. J. Klein, P. L. Roesch, S. L. Bertin, T. W. reppen, Q. Chu, A. V. Blokhin, J. E. Hagstrom, J. A. Wolff, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12982-12987.

[183]   J. Yang, C. Chen, X. Tang, *Bioconjug. Chem.* **2018**, *29*, 1010-1015.

[184]   R. L. Letsinger, G. R. Zhang, D. K. Sun, T. Ikeuchi, P. S. Sarin, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 6553-6556.

[185]   J. D. Brunzell, M. Davidson, C. D. Furberg, R. B. Goldberg, B. V. Howard, J. H. Stein, J. L. Witztum, *Diabetes Care* **2008**, *31*, 811-822.

[186]   C. Wolfrum, S. Shi, K. N. Jayaprakash, M. Jayaraman, G. Wang, R. K. Pandey, K. G. Rajeev, T. Nakayama, K. Charrise, E. M. Ndungo, T. Zimmermann, V. Koteliansky, M. Manoharan, M. Stoffel, *Nat. Biotechnol.* **2007**, *25*, 1149-1157.

[187]   K. Brunner, J. Harder, T. Halboch, J. Willibald, F. Spada, F. Gnerlich, K. Sparrer, A. Beil, L. Möckl, C. Bräuchle, K. Conzelmann, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 1946-1949.

[188]   M. Glass, M. Dragunow, R. L. Faull, *Neuroscience* **1997**, *77*, 299-318.

[189]   A. A. Turanov, A. Lo, M. R. Hassler, A. Makris, A. Ashar-Patel, J. F. Alterman, A. H. Coles, R. A. Haraszti, L. Roux, B. M. D. Godinho, D. Echeverria, S. Pears, J. Iliopoulos, R. Shanmugalingam, R. Ogle, Z. K. Zsengeller, A. Hennessy, S. A. Karumanchi, M. J. Moore, A. Khvorova, *Nat. Biotechnol.* **2018**, *36*, 1164-1173.

[190]   E. Ansseau, C. Vanderplanck, A. Wauters, S. Q. Harper, F. Coppée, A. Belayew, *Genes* **2017**, *8*, 1-21.

[191]   G. Cesarone, O. P. Edupuganti, C. P. Chen, E. Wickstrom, *Bioconjug. Chem.* **2007**, *18*, 1831-1840.

[192]   M. R. Alam, X. Ming, M. Fisher, J. G. Lackey, K. G. Rajeev, M. Manoharan, R. D, Juliano, *J. Am. Chem. Soc.* **1011**, *22*, 1673-1681.

[193]   X. Liu, W. Wang, D. Samarsky, L. Liu, Q. Xu, W. Zhang, G. Zhu, P. Wu, X. Zuo, H. Deng, J. Zhang, Z. Wu, X. Chen, L. Zhao, Z. Qiu, Z. Zhang, Q. Zeng, W. Yang, B. Zhang, A. Ji, *Nucleic Acids Res.* **2014**, *42*, 11805-11817.

[194]   A. G. Morell, G. Gregoriadis, I. H. Scheinberg, J. Hickman, G. Ashwell, *J. Biol. Chem.* **1971**, *246*, 1461-1467.

[195]   M. Yasuda, L. Gan, B. Chen, S. Kadirvel, C. Yu, J. D. Phillips, M. I. New, A. Liebow, K. Fitzgerald, W. Querbes, R. J. Desnick, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7777-7782.

[196]   A. Aviñó, S. M. Ocampo, R. Lucas, J. J. Reina, J. C. Morales, J. C. Perales, R. Eritja, *Mol. Divers*. **2011**, *15*, 751-757.

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

[198]   J. Kimura, K. Yagi, H. Suzuki, O. Mitsunobu, *Bull. Chem. Soc. Jpn*. **1980**, *53*, 3670-3677.

[199]   B. J. Kopina, C. H. Lauhon, *Org. Lett*. **2012**, *14*, 4118-4121.

[200]   B. Uprety, C. Arderne, I. Bernal, *Eur. J. Inorg. Chem*. **2018**, *47*, 5058-5067.

[201]   J. J. Butzow, G. L. Eichhorn, *Biopolymers* **1965**, *3*, 95-107.

[202]   S. Matsushita, F. Ibuki, *Mem. Res. Inst. Food Sci. Kyoto Univ*. **1960**, *22*, 32-38.

[203]   J. W. Huff, S. Sastry, M. P. Gordon, W. E. C. Wacker, *Biochemistry* **1964**, *3*, 501-506.

[204]   M. Hutchby, C. E. Houlden, J. Gair Ford, S. N. G. Tyler, M. R. Gagné, G. C. Lloyd-Jones, K. I. Booker-Milburn, *Angew. Chem. Int. Ed*. **2009**, *121*, 8877-8880.

[205]   V. Boudou, J. Langridge, A. van Aerschot, C. Hendrix, A. Millar, P. Weiss, P. Herdewijn, *Helv. Chim. Acta* **2000**, *83*, 152-161.

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

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

[208]   V. Serebryany, L. Beigelman, *Tetrahedron Lett*. **2002**, *43*, 1983-1985.

[209]   B. M. Trost, C. G. Caldwell, *Tetrahedron Lett*. **1981**, *22*, 4999-5002.

[210]   M. Sekine, S. Limura, K. Furusawa, *J. Org. Chem*. **1993**, *58*, 3204-3208.

[211]   H. Schaller, G. Weinmann, B. Lerch, H. G. Khorana, *J. Am. Chem. Soc*. **1963**, *85*, 3821-3827.

[212]   H. Seliger, D. Zeh, G. Azuru, J. B. Chattopadhyaya, *Chem. Scr*. **1983**, *22*, 95-102.

[213]   M. H. Caruthers, D. Dellinger, K. Prosser, A. D. Brone, J. W. Dubendorff, R. Kierzek, M. Rosendahl, *Chem. Scr*. **1986**, *26*, 25-30.

[214]   R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, S. M. Freier, *Biochemistry* **1986**, *25*, 7840-7846.

[215]   I. Hirao, M. Ishikawa, K. Miura, *Nucleic Acids Symp. Ser*. **1985**, *16*, 173-176.

[216]   H. Tanimura, T. Fukuzawa, M. Sekine, T. Hata, J. W. Efcavitch, G. Zon, *Tetrahedron Lett*. **1988**, *29*, 577-578.

[217]   H. Tanimura, M. Sekine, T. Hata, *Nucleosides Nucleotides* **1986**, *5*, 363-383.

[218] H. Tanimura, T. Imada, *Chem. Lett.* **1990**, 1715-1718.

[219] A. Tanpure, S. Balasubramanian, *ChemBioChem* **2017**, *18*, 2236-2241.

[220] K. Bartosik, E. Sochacka, G. Leszczynska, *Org. Biomol. Chem.* **2017**, *15*, 2097-2103.

[221] H. S. Bernhardt, W. P. Tate, *Biology Direct* **2012**, *7*, 4.

[222] B. Barbier, A. Brack, *J. Am. Chem. Soc.* **1988**, *110*, 6880-6882.

[223] A. R. Fehr, S. Perlman, *Methods Mol. Biol.* **2015**, *1282*, 1-23.

[224] P. Zhou, Z.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, H.-R. Sui, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, *Nature* **2020**, *579*, 270-273.

[225] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, S. Pöhlmann, *Cell* **2020**, *181*, 271-280.

[226] J. Willibald, J. Harder, K. Sparrer, K.-K. Conzelmann, T. Carell, *J. Am. Chem. Soc.* **2012**, *134*, 12330-12333.

[227] R. Huisgen, *Proc. Chem. Soc.* **1961**, 357-396.

[228] H. Yan, K. Tram, *Glycoconjugate J.* **2007**, *24*, 107-123.

[229] C. Uttamapinant, A. Tangpeerachaikul, S. Grecian, S. Clarke, U. Singh, P. Slade, K. R. Gee, A. Y. Ting, *Angew. Chem. Int. Ed.* **2012**, *51*, 1-6.

[230] A. L. Allan, P. L. Gladstone, M. L. P. Price, S. A. Hopkins, J. C. Juarez, F. Donate, R. J. Ternansky, D. E. Shaw, B. Ganem, Y. Li, W. Wang, S. Ealick, *J. Med. Chem.* **2006**, *49*, 7807-7815.

[231] K. Bartosik, E. Sochacka, G. Leszczynska, *Org. Biomol. Chem.* **2017**, *15*, 2097-2103.