Charles University

Faculty of Science

Study programme: Organic chemistry

Branch of study: Organic chemistry

Mgr. Ivana Ivancová

Modification of nucleic acids by reactive groups for bioconjugations and cross-linking with lysine containing peptides and proteins

Modifikace nukleových kyselin reaktivními skupinami pro biokonjugace a cross-linkování s peptidy a proteiny obsahujícími lysin

Dissertation

Supervisor: Prof. Ing. Michal Hocek, Ph.D., DSc.

Prague, 2022

This dissertation was worked out at the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague form September 2016 to December 2022.

Declaration:

I hereby declare that I have worked out my dissertation by myself and that all literature sources used are listed in the reference list. Neither this work, nor its significant part was used to obtain other academic title.

Prague,

Ivana Ivancová

Acknowledgement

Above all, I would like to thank my supervisor Prof. Michal Hocek for his guidance, patience, and trust. I thank Dr. Radek Pohl for his help with measurement and interpretation of complex NMR spectra, people from MS department at IOCB, namely, Ing. Kateřina Nováková for measurement of MALDI-TOF spectra and Mgr. Marta Vlková for proteomic analysis. Dr. Eva Konkoľová and Ing. Eliška Koutná for their advice on how to work with proteins, Mgr. Juraj Konč and Dr. Anton Škríba for advice on MS analysis of proteins.

I also want to thank to my colleagues for creative working atmosphere, especially to Dr. Agata Olszewska for introducing me to the world of biochemistry and for her friendship, good mood, inspiring life attitude, advice, and encouragement. Dr. Nemanja Milisavljevič, Dr. Lucia Veselovská, Dr. Ján Matyašovský, Ing. Vojtěch Havlíček, Dr. Olena Mayboroda, Dr. Michal Tichý and Dr. Veronika Sýkorová for chitchat and laughter when needed and Tereza Schröpferová who kept our labs properly stocked.

A special mention goes to Dr. Miroslav Kuba, my dearest friend and partner in crime, for his constructive criticism that made me mad, as he was always right.

Finally, yet importantly I am grateful to my mother and sisters for always believing in me.

My Ph.D. work was a part of multidisciplinary project performed by Prof. Hocek's research group at the Institute of Organic Chemistry and Biochemistry (IOCB) AS CR. All the experiments described in the thesis were performed by me in person. Measurement and interpretation of complex NMR spectra was done by Dr. Radek Pohl. Ing. Kateřina Nováková measured MALDI-TOF mass spectra of prepared oligonucleotides and the mass spectra of all the synthesised compounds were measured by the mass spectrometry department at IOCB Prague. The measurements of DNA-protein conjugates were performed by Dr. Martin Hubálek and the measurement of RNA-protein conjugate and proteomic analysis was performed by Mgr. Marta Vlková.

This work was supported by Czech Academy of Sciences (Praemium Academiae award to M.Hocek), by the Czech Science Foundation (18-03305S) and European Regional Development fund; OP RDE; Project: “ChemBioDrug” (No. CZ.02.1.01/0.0/0.0/16_019/0000729).

Abstract

In the first part of this thesis, I developed reactive DNA probe for selective cross-linking with lysine residues of DNA-binding proteins. I synthesized 2'-deoxycytidine 5'-*O*-mono- and triphosphate bearing squaramate moiety tethered to the position 5 via propargylamine linker. The monophosphate was used as a model compound to test the reactivity of this mixed squaramate in cross-linking reactions with lysine and short lysine containing peptides. Squaramate modified 2'-deoxycytidine 5'-*O*-triphosphate was found to be suitable substrate for KOD XL polymerase in both PEX and PCR synthesis of modified DNA. Squaramate modified DNA forms stable diamide linkage with primary amines. I tested the reactivity of this DNA probe in bioconjugation reactions with sulfo-Cy5-amine and lysine containing peptides. Afterwards, squaramate-linked DNA was successfully cross-linked with lysine rich histone proteins. This reactive squaramate modified nucleotide showed potential for following bioconjugation reactions of nucleic acids with amines or lysine containing peptides and proteins without the need of external reagent.

Based on positive results of experiments with squaramate modified DNA, in the second part of the thesis I developed and synthesized squaramate modified ribonucleotide to study cross-linking with RNA binding proteins. After optimisation of reaction conditions, squaramate modified cytidine 5'-*O*-triphosphate was incorporated into the RNA by T7 RNA polymerase in *in vitro* transcription reaction. Next, the modified RNA was used for post-synthetic labelling with sulfo-Cy-5-amine. Finally, I tested the squaramate modified RNA probe in reaction with various model proteins (i.e., JEV and YFV NS5, SARS-CoV-2 RdRp, SARS-CoV-2 nucleoprotein and HIV-rt). The modified ribonucleotide was also used in RNA extension reactions catalysed by RNA dependent RNA polymerases (JEV NS5 and SARS-CoV-2 RdRp). Synthesized modified RNA cross-linked during polymerase reaction with JEV NS5 polymerase, which was confirmed by PAGE and immunodetection. Proteomic analysis of polymerase reaction mixtures identified three lysine residues (K269, K462, K463) of the protein cross-linked with squaramate modified RNA.

Abstrakt

V první části dizertační práce jsem vyvinula reaktivní DNA sondu pro selektivní cross-linkování s lysinovými zbytky DNA vazebních proteinů. Připravila jsem 2'-deoxycytidin 5'-*O*-monofosfát a trifosfát nesoucí v poloze 5 squramátovou skupinu navázanou přes propargylaminovou spojku. Monofosfát byl použit jako modelová sloučenina na testování reaktivity tohoto smíšeného squaramátu v cross-linkových reakcích s lysinem a krátkými peptidy obsahujícími lysin. Squaramátem modifikovaný 2'-deoxycytidin 5'-*O*-trifosfát byl dobrým substrátem pro KOD XL polymerázu pro enzymovou syntézu DNA metodou extenze primeru (PEX) a taky metodou polymerázové řetězové reakce (PCR). Squaramátem modifikovaná DNA tvoří stabilní diamidové vazby s primárními aminy. Reaktivitu modifikované DNA jsem testovala v biokonjugačních reakcích se sulfo-Cy5-aminem a s peptidy obsahujícími lysin. Poté byla squaramátem modifikovaná DNA úspěšně kovalentně připojena k histonovým proteinům bohatým na lysin. Reaktivní, squaramátem modifikovaný nukleotid vykazoval potenciál pro další biokonjugační reakce nukleových kyselin s aminy nebo lysin obsahujícími peptidy a proteiny bez potřeby externího činidla.

Na základě pozitivních výsledků experimentů se squaramátem modifikovanou DNA, jsem v druhé části práce vyvinula a připravila squaramátem modifikovaný ribonukleotid pro studium cross-linků s RNA vazebnými proteiny. Po optimalizaci reakčních podmínek byl squaramátem modifikovaný cytidin 5'-*O*-trifosfát inkorporován do RNA pomocí T7 RNA polymerázy v *in vitro* transkripční reakci. Dále byla modifikovaná RNA použita pro post syntetické značení sulfo-Cy5-aminem. Nakonec jsem testovala squaramátem modifikovanou RNA sondu v reakci s různými modelovými proteiny (tj. JEV a YFV NS5, SARS-CoV-2 RdRp, SARS-CoV-2 nukleoprotein a HIV-rt). Modifikovaný ribonukleotid byl také použit v RNA extenzních reakcích katalyzovaných RNA dependentními RNA polymerázami (JEV NS5 a SARS-CoV-2 RdRp). Modifikovaná RNA se kovalentně vázala na protein během polymerázové reakce s JEV NS5 polymerázou, což bylo potvrzeno gelovou separací (PAGE) a imunodetekcí. Proteomická analýza polymerázových reakčních směsí identifikovala tři lysinové zbytky proteinu (K269, K462, K463) kovalentně připojených ke squaramátem modifikované RNA.

List of abbreviations

6-FAM	6-Carboxyfluorescein
AA	amino acid
Ac	acetyl
βME	β-mercaptoethanol
bp	base pair
BSA	bovine serum albumine
Cy-5	Cyanine-5
DCM	dichloromethane
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dNMP	2'-deoxynucleoside-5'-*O*-monohosphate
dNTP	2'-deoxynucleoside-5'-*O*-triphosphate
DPC	DNA-protein cross-link
ds	double stranded
ESQ	ethoxy squarate
EMSA	electrophoretic mobility shift assay
FG	functional group

HIV-rt	Human immunodeficiency virus reverse transcriptase
HPLC	high performance liquid chromatography
IVT	*in vitro* transcription
JEV	Japanese encephalitis virus
MALDI-TOF	matrix-assisted laser desorption/ionisation-time of flight
mRNA	messenger RNA
MTase	Methyl transferase
MS	mass spectrometry
NA	nucleic acid
NBS	N-Bromo succinimide
NEAR	nicking enzyme amplification reaction
NMR	nuclear magnetic resonance
NPC	nucleic acid-protein cross-linking
NTP	nucleoside triphosphate
ON	oligonucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEX	primer extension
PhotoCAX	photocatalytic cross-linking

RBP	RNA binding protein
RCA	rolling circle amplification
RCAP	reversible cross-linking affinity
RdRp	RNA dependent RNA polymerase
RNA	ribonucleic acid
RNAp	RNA polymerase
rpm	rounds per minute
rRNA	ribosomal RNA
r.t.	room temperature
rt	reverse transcriptase
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
ss	single stranded
SSB	single strand binding protein
TEAB	triethylammonium bicarbonate
TPPTS	3,3',3''-phosphanetriyltris trisodium salt
UV	ultraviolet
YFV	Yellow fever virus

Contents

1. Introduction

1.1 Nucleic acids: structure and function

Nucleic acids (NA; deoxyribonucleic acid – DNA and ribonucleic acid – RNA) are two of the five most important biomolecules, besides proteins, lipids, and carbohydrates, which are natural building blocks crucial for the existence of living organisms.

DNA and RNA are polyanionic biomolecules built by a chain of nucleotide building blocks. The structure of each nucleotide consists of a (deoxy)ribose sugar moiety, a three phosphate groups attached at the C-5 position (two are lost during NA synthesis, so there is one phosphate group per nucleotide in the nucleic acid strand) and one nitrogenous base attached via N-glycosidic bond at the anomeric C-1 position of the pentose in β-configuration. There are two modifications in the chemical structure of nucleic acids in which DNA and RNA vary: the sugar and the nucleobases of each nucleotide. In DNA 2-deoxyribose saccharide and adenine (A), guanine (G), thymine (T) and cytosine (C) nucleobases are found naturally. RNA structure differs by the presence of 2'-OH group on the sugar and thymine is replaced by uracil (U; **Figure 1**). Nevertheless, apart from the similarity in the chemical structure, DNA and RNA holds immense structural and functional variety.[1]

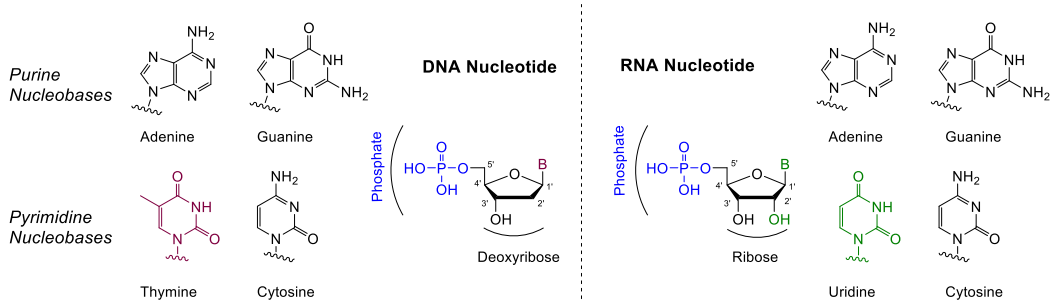


Figure 1. General structures of DNA and RNA nucleotides.

Nucleic acids carry biological information in their linear sequence of nucleotides necessary for the protein synthesis, and via their shapes, they direct

their assembly with other cellular biomolecules. DNA and RNA structure is organized on different levels.

Primary structure of both NA is composed as a linear chain of nucleotides connected through 5', 3'-phosphodiester bond (**Figure 2-a**).

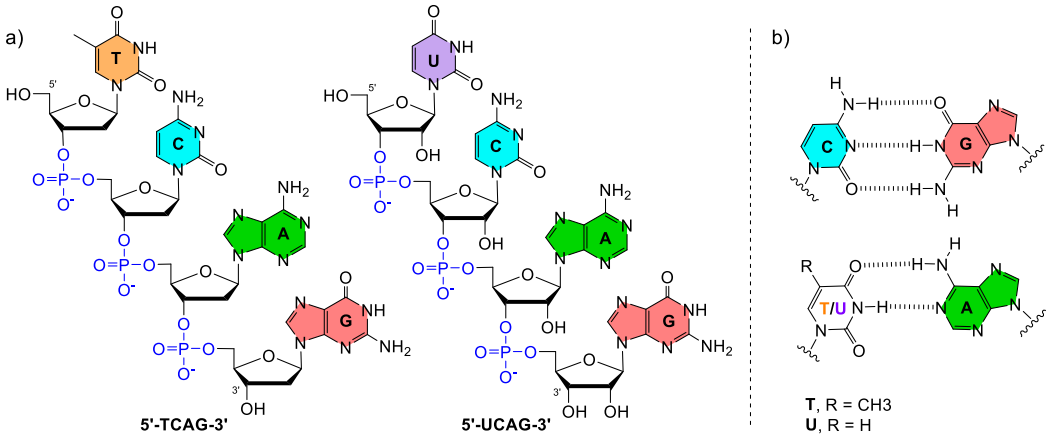


Figure 2. a) Primary structure of DNA/RNA. b) Watson-Crick base pairing.

Secondary structures are the set of interactions between the nucleobases. The two antiparallel strands of the DNA double helix are held together by Watson-Crick hydrogen bonds in the canonical complementarity: pyrimidine pairs with purine (T/U : A; C : G; **Figure 2-b**).[2] There are also additional π-π stacking interactions of the nucleobases that contribute to the stabilization of the molecule structure.[3] In the secondary structure of RNA, base pairs form planar structures of the stem-loop (**Figure 3-b**).[4]

The major standard DNA secondary structures are A-DNA and B-DNA with right-handed helix twist and Watson-Crick base pairing. A left-handed Z-DNA double helix is stabilized by high salt concentrations. Its phosphate backbone appears in a zig-zag fashion (**Figure 3-a**).[5] Unlike DNA, RNA is usually right-handed single stranded polynucleotide. Complementary regions within the strand can base pair together and form diverse three-dimensional structures, which include helices, bulges, hairpins, kissing complexes, knots, and others, highly depended on metal ions (**Figure 3-b**).[6,7]

When interacting with other NA molecules or proteins, nucleic acids form higher-order tertiary structures such as nucleosome in the case of the binding of DNA to histones[8] or ribosome which consists of one or more rRNAs and several ribosomal proteins (**Figure 3-c**).[9]

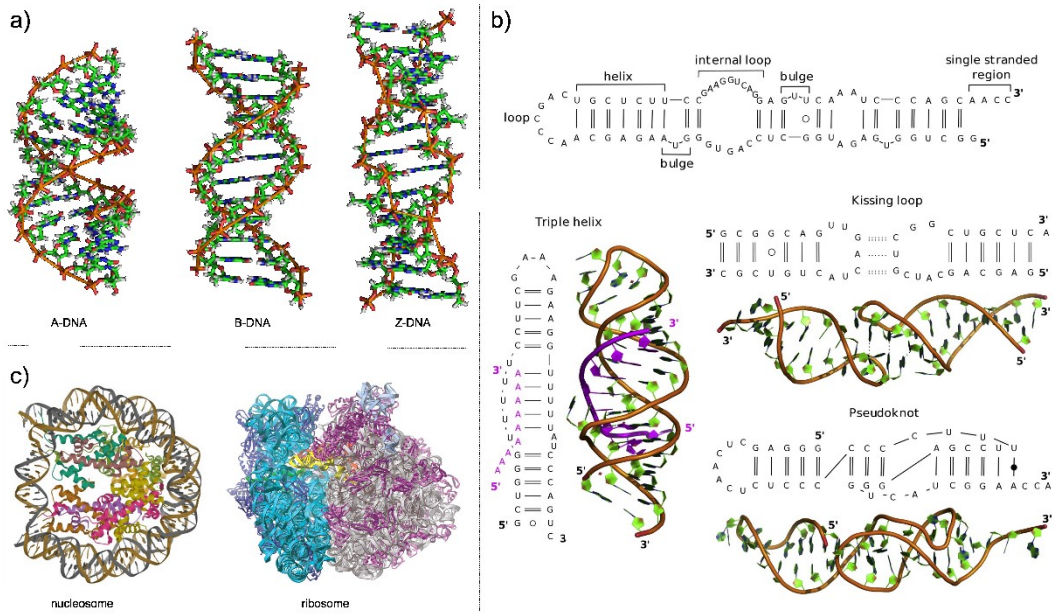


Figure 3. Common structures of nucleic acids: a) DNA secondary structures (picture source: Wikimedia commons). b) Examples of secondary RNA structures[10] c) examples of DNA (PDB 3LZ1) and RNA[9] tertiary structures.

DNA stores the genetic information of an organism and RNA was so far considered only as a mediator in translating these instructions to protein synthesis. Nowadays it is obvious that nucleic acids, particularly RNA, can play many other roles like regulation of gene expression[11–13] or catalysis[14–17] based on their structural abundance. On the contrary, NA have been recognized as a toxic factor involved in the development of numerous human diseases.[18–21] Finally, nucleic acids became a tool applied in biotechnology[15,22,23] and medicine[24–26] for example as a NA-based vaccines.[27–30]

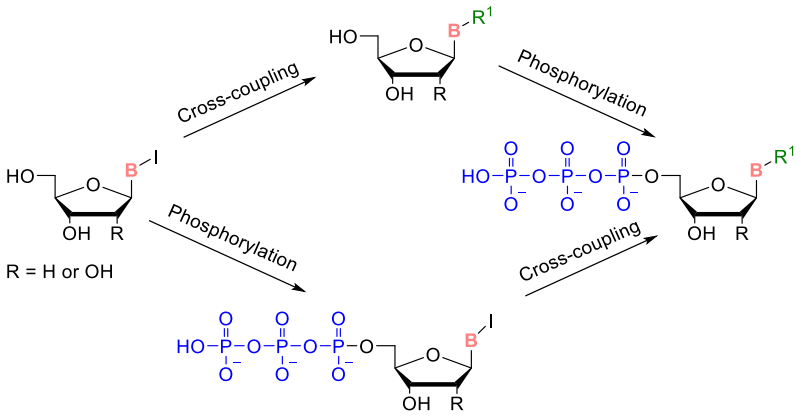
1.2 Modified nucleotides: building blocks for the modification of nucleic acids

Nucleoside triphosphates (NTP) play crucial role in many biological systems, for example adenosine 5'-triphosphate (ATP) serves as a universal energy currency in cellular processes,[31] UTP and CTP take part in the metabolism of phospholipids and galactose,[32,33] while GTP is secondary messenger in growth factors signalling machinery.[34] Their main role is, above all, the construction of nucleic acids. (2'-deoxy)ribonucleoside triphosphates [(d)NTP] are elemental building blocks that serve as a substrate in the synthesis of nucleic acids catalysed by DNA/RNA polymerases.

(d)NTPs can be easily functionalized at various positions.[35,36] The polymerization of modified (d)NTPs catalysed by DNA or RNA polymerases represents flexible way to generate functionalized nucleic acids with many various applications.[37]

1.2.1 Synthesis of modified (d)NTPs

Chemical modifications at the C-5 position of pyrimidine and C-7 position of 7-deazapurine are well tolerated by polymerases.[38,39] Modified (d)NTPs can be prepared in two different ways. It is either direct phosphorylation of modified nucleoside, or the halogenated nucleoside triphosphate is functionalized under mild conditions (**Scheme 1**).[40]

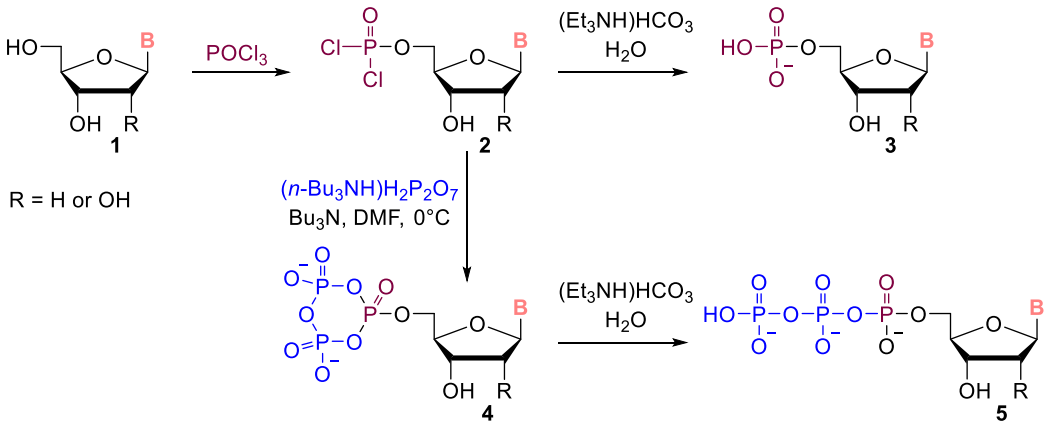


Scheme 1. Two different approaches to prepare modified (d)NTPs (B = nucleobase).

1.2.1.1 Phosphorylation of base-modified nucleosides

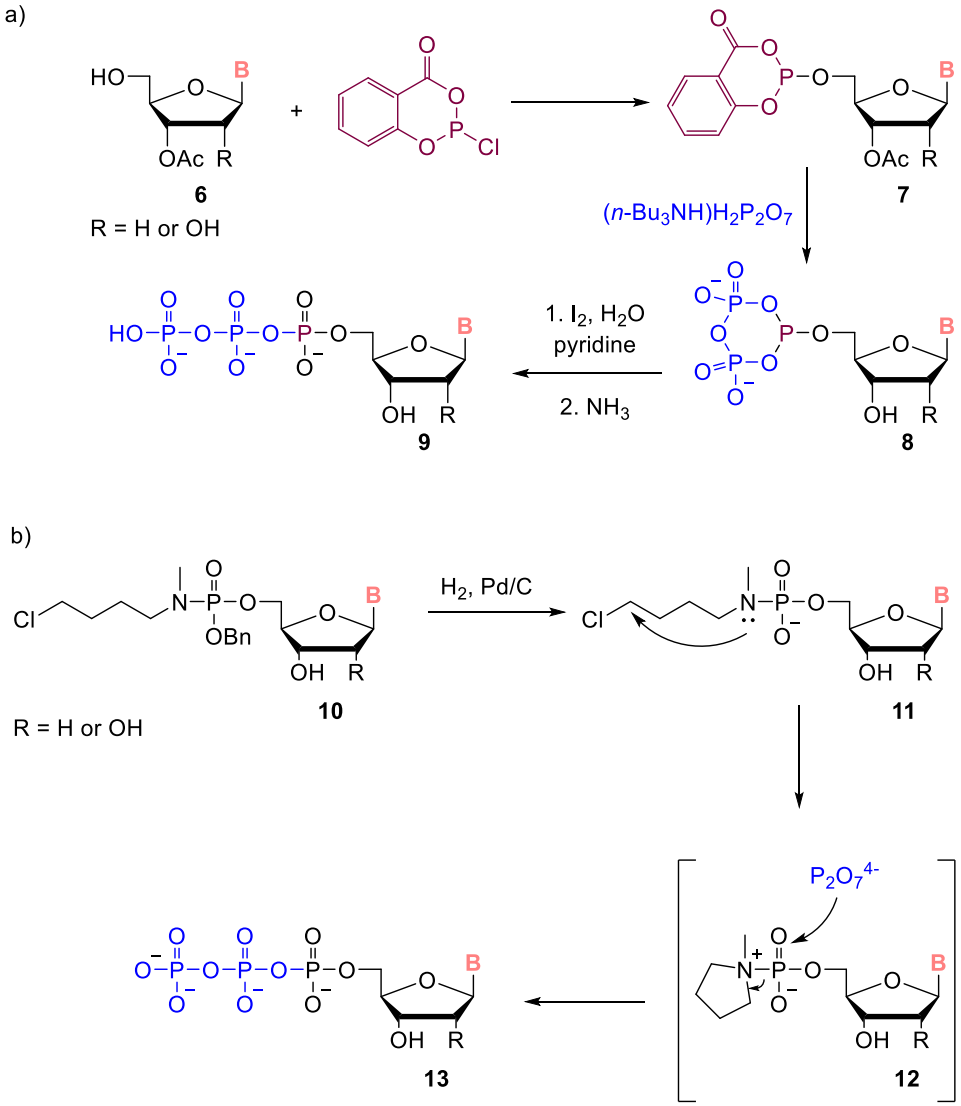
There is still no general high-yielding method for the synthesis of (deoxy)nucleoside triphosphates. The phosphorylation method must be optimized for each individual case and the purifications are often laborious.

Two-step, one-pot Yoshikawa protocol[41] is the oldest and most used method for the synthesis of nucleoside triphosphates (**Scheme 2**). This simple 5'-regioselective[42,43] phosphorylation starts with the reaction of an unprotected nucleoside **1** with $POCl_3$ (phosphorus oxychloride) to give highly reactive phosphorodichlorate **2** which is subsequently reacted in situ with pyrophosphate to yield a cyclic triphosphate **4**. This cyclic triphosphate is hydrolysed to the desired linear compound **5**. To obtain 5'-monophosphorylated nucleoside **3**, intermediate **2** is directly hydrolysed. The primary OH group at 5' position is usually more reactive than secondary hydroxyls (2', 3'-OH), therefore there is no need for the protection in most cases. The final (d)NTPs are usually isolated by preparative HPLC and/or ionex exchange chromatography.



Scheme 2. Yoshikawa phosphorylation (B = modified/natural nucleobase).

The two alternative triphosphorylation protocols, Ludwig-Eckstein[44] and Borch[45] (**Scheme 3**), based on different reagents (e.g., salicyl phosphorochlorite) are barely used for the synthesis of base-modified nucleic acids.



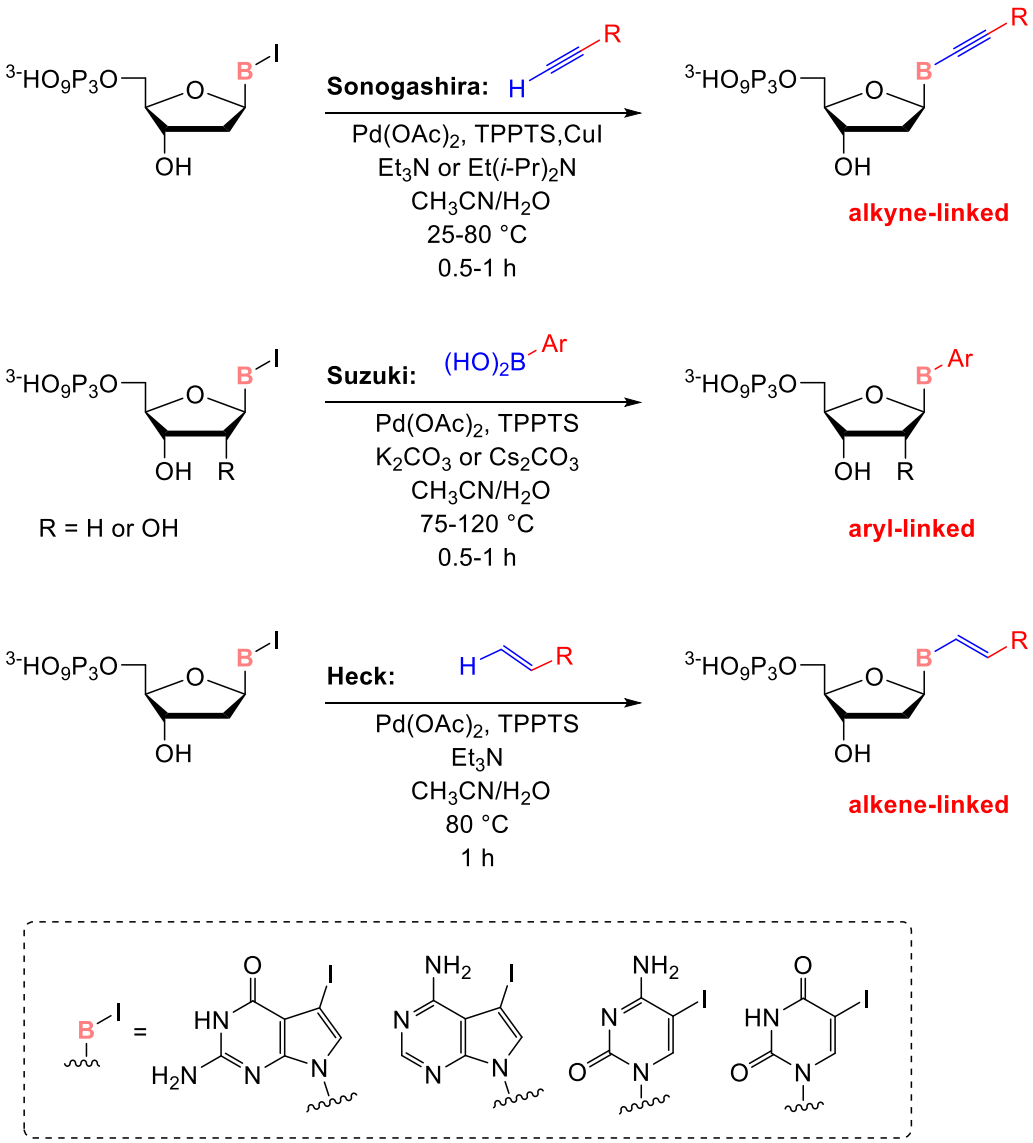
Scheme 3. Alternative triphosphorylation methods: a) Ludwig-Eckstein phosphorylation. b) Borch phosphorylation. (B = modified/natural nucleobase)

1.2.1.2 Synthesis of base-modified NTPs by direct aqueous cross-coupling reactions

The use of Pd-catalysed reactions has achieved wide application in the synthesis of base modified nucleosides. It is a convenient method for the formation of C-C bond and can be used in not only functionalization of pyrimidine and purine nucleosides, but also their mono- and triphosphates once water-soluble catalytic system is used.[46]

Aqueous-phase cross-coupling reactions are ideal for direct functionalization of nucleotides in the case the modification is unstable under phosphorylation conditions. This method overcomes the multistep procedures without the need of protecting groups. A number of methods for direct modification of 5-iodopyrimidine and 7-iodo-7-deazapurine triphosphates via Pd-catalysed aqueous cross-coupling reactions have been developed in the Hocek group and others.[47] All coupling reactions anticipate the synthesis of iodo-modified (d)NTPs and their reaction with boronic acids or trifluoroborates to introduce aryl or alkenyl functional groups (FG; Suzuki reaction),[48–50] terminal alkynes to attach alkynyl-linked moieties (Sonogashira reaction).[51] Lately, Heck[51,52] reaction was developed to attach alkenyl FG. Until now, there is no direct alkylation via cross-coupling reaction available, since no relevant organometallic reagents (alkylmagnesium halides, zinc halides and cuprates) are compatible with water.

Aqueous cross-coupling reactions generally proceed in the presence of palladium catalyst ($Pd(OAc)_2$), water-soluble ligand (TPPTS) and base in aqueous acetonitrile at higher temperatures. Shorter reaction times ($\leq$ 1h) are required to minimise the hydrolysis of the triphosphate moiety (**Scheme 4**).



Scheme 4. General scheme of cross-coupling reactions of iodinated (d)NTP.

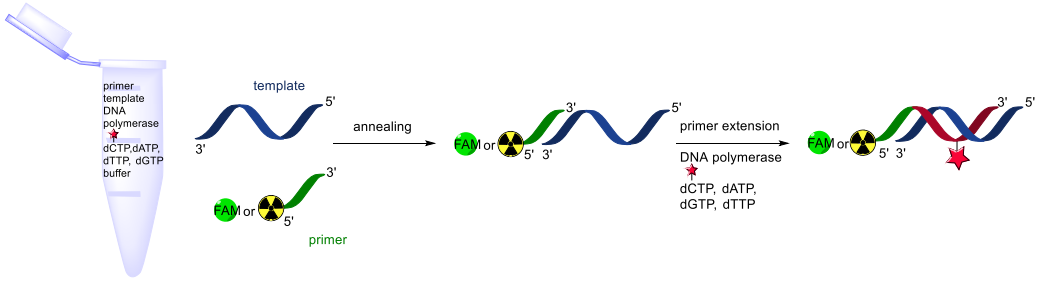
1.3 Enzymatic synthesis of base-modified nucleic acids

Since the first successful enzymatic incorporation of biotin modified (2'-deoxy)uridine triphosphate by DNA and RNA polymerases,[53] many ways of enzymatic incorporation to generate functionalized nucleic acids have been developed, and modified NAs became of interest of chemical biology, material science and diagnostics.[37]

1.3.1 Enzymatic synthesis of base-modified DNA

As was mentioned earlier, polymerization of modified nucleoside triphosphates catalysed by DNA polymerases is an excellent way to generate functionalized nucleic acids.[47] It is favoured for synthesis of longer DNA strands because of 100% efficacy of incorporation of all nucleotides. Primer extension (PEX) and polymerase chain reaction (PCR) are two main enzymatic methods used to synthesize DNA.

PEX is mostly used to synthesize shorter DNA (up to 80 base pairs) with modifications in one strand. The reaction mixture contains of natural or modified 2'-deoxynucleoside triphosphates, DNA polymerase and a template annealed to a short primer, which is extended at its 3'-end. PEX products are analysed by polyacrylamide gel electrophoresis (PAGE). Labelling of primers at 5'-end by ^{32}P-phosphate or various fluorescent probes enables the visualization of the reaction on the gel (**Scheme 5**).

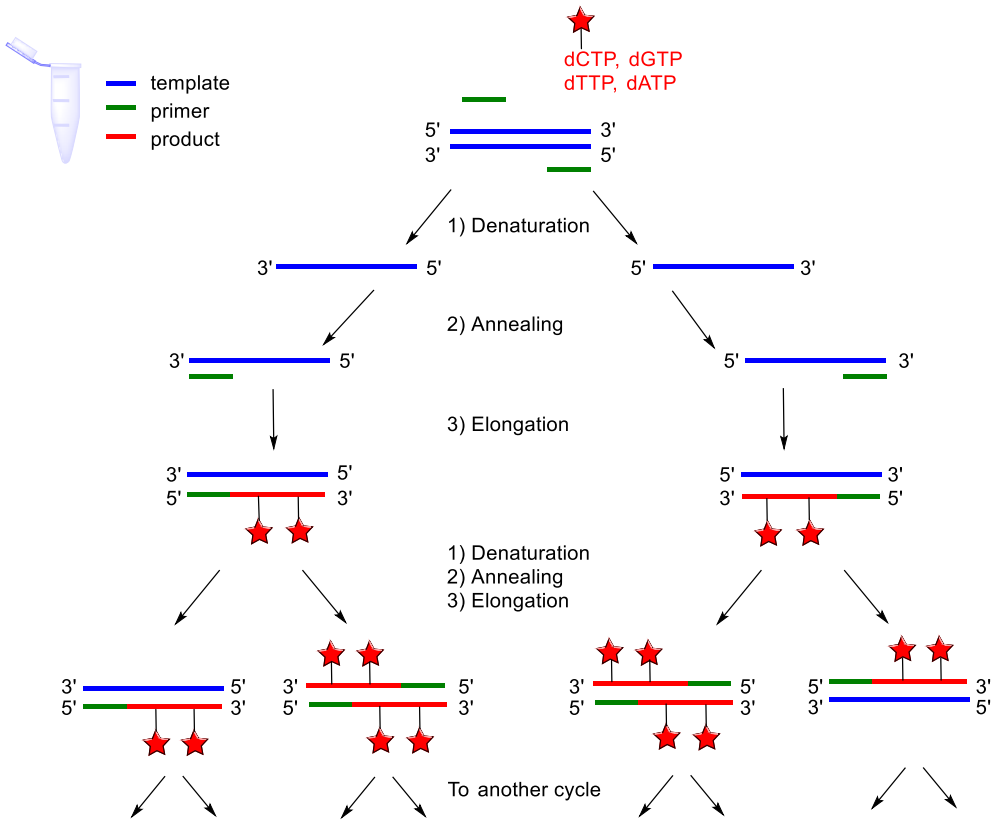


Scheme 5. Example of PEX reaction of modified 2'-deoxycytidine triphosphate and labelled primer. One modification is incorporated into the DNA.

DNA longer than 80 bp is mostly prepared by PCR method.[54] PCR reaction (**Scheme 6**) is a sequence of repeating cycles (usually 30-40), each of three steps:

1. denaturation of long double stranded DNA (dsDNA) template at 95 °C,
2. annealing of, now single stranded, template with excess of two complementary primers (annealing temperature depends on the melting temperature of the primers, usually is around 50-55 °C)
3. and finally, primer extension with thermostable DNA polymerase (usually 72 °C).

The amplification of a DNA is exponential. In the case that modified dNTPs are used, resulting dsDNA is heavily modified in both strands. This feature makes PCR more challenging for the DNA polymerase as it must not only incorporate modified dNTPs, but to read the modified template in the following cycle as well.



Scheme 6. PCR reaction scheme.

Both methods, PEX and PCR, generate double stranded DNA, however, there are several ways how to generate or synthesize single stranded DNA (ssDNA). One of them is to use phosphorylated primers in the reaction followed by λ-exonuclease treatment. Lambda exonuclease is DNA exonuclease which selectively degrades phosphorylated strain of DNA duplex.[55] Next method is based on biotin streptavidin separation. Biotin is attached to a 5'-end of the primer and after incubation of the newly synthesized dsDNA with streptavidin beads, due to the affinity to streptavidin on primer, ssDNA is isolated.[56] Recently, few other synthetic methods to generate ssDNA have emerged, such as nicking enzyme amplification reaction (NEAR)[57] or rolling circle amplification (RCA).[58]

It is important to mention here also the limitations of the enzymatic synthesis. Even though it tolerates large number of modifications, the position of the attached functional group is ultimate for successful incorporation. As was already mentioned, DNA polymerases tolerate chemical modifications at the C-5 position of pyrimidine and C-7 position of 7-deazapurine as they are not vital for base pairing.[38,39] Although C-8 position of purines is not essential for base pairing as well, bulky modifications in this position destabilize secondary structure of dsDNA.[59] Some modifications at the N^4-position of cytosine have been reported as tolerated substrates for phi29 DNA polymerase.[60] Position 5 of pyrimidine and 7 of 7-deazapurines are pointing out to the major groove, where even bulky modifications, for example fluorescent dyes, are well tolerated if they are attached to the nucleobase via long linker.[61] Enzymatic synthesis of DNA modified in minor groove has also been reported (**Figure 4**).[62]

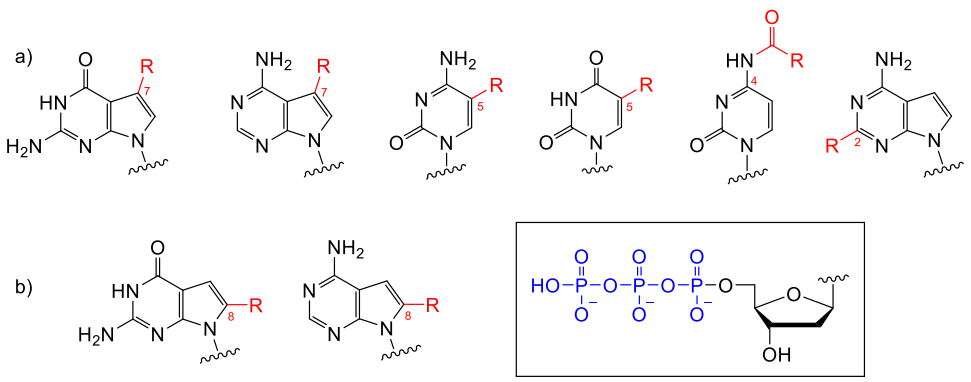


Figure 4. General structures of modified dNTPs a) tolerated and b) not tolerated by DNA polymerases.

1.3.2 Enzymatic synthesis of base-modified RNA

RNA, in any of its types, touches almost every process in a cell. It is essential in broad range of biological roles (coding, decoding, regulation, gene expression).

The first step of gene expression pathway is the transcription, the copying of the genetic information stored in DNA into RNA molecules. This process is controlled by RNA polymerase (RNAp), which assists in both separation of the two DNA strands and the polymerization of RNA nucleotides. The resulting RNA strand is complementary to the ssDNA template and identical to the coding strand (only U is replacing T). Transcription reaction produces precursors for functional RNA or mRNA which codes the protein synthesis (**Figure 5**).[63]

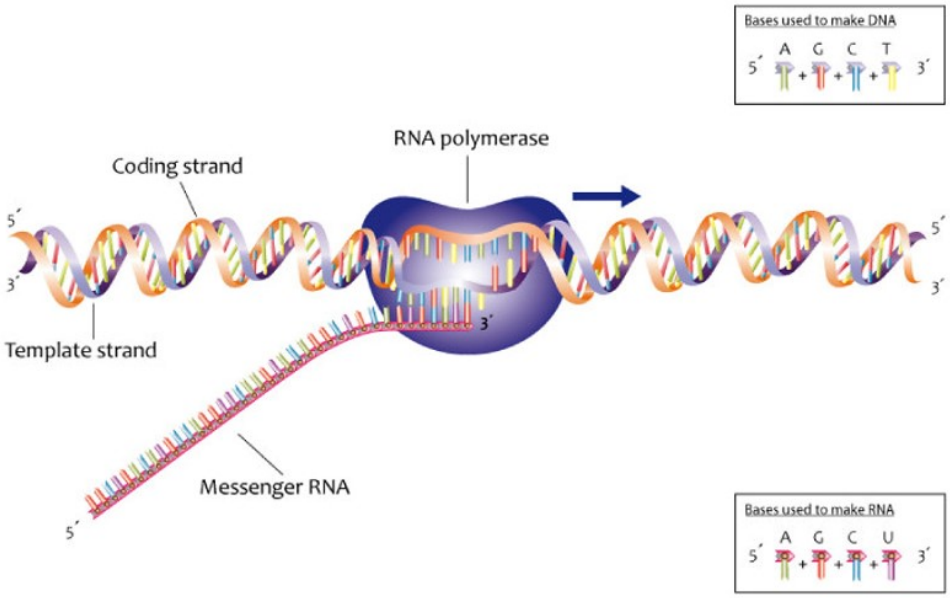


Figure 5. Transcription reaction (picture source: Wikimedia commons).

The most used method for the enzymatic synthesis of long modified RNA is the *in vitro* transcription (IVT) reaction catalysed by bacteriophage T7 RNA polymerase. As opposed to DNA synthesis by DNA polymerases, primer is not essential for the synthesis of RNA by DNA dependent T7 RNAp. For the successful initiation of the transcription, the presence of guanosines in the position + 1 and/or + 2 and specific promoter sequence recognized by the polymerase is required (**Figure 6**).[64–68] Once polymerase is bound to the promoter, repeated

cycles of initiation and product release lead to the presence of short abortive products. Only after new RNA is 8-12 nucleotides long, the T7 RNAp changes its conformation and starts elongation phase. Termination is caused by reaching the terminator sequence, however for *in vitro* transcripts it is by "run off" when polymerase falls of the 5' end of a linear template.[69]

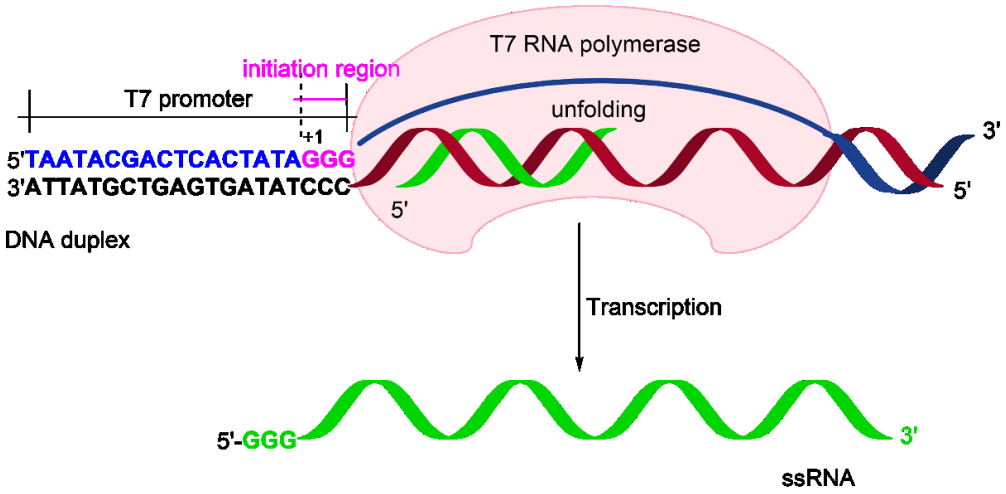


Figure 6. *In vitro* transcription reaction catalysed by T7 RNAp with highlighted promoter sequence.

The T7 RNA polymerase can transcribe diverse DNA templates with T7 promotor regions *in vitro*, therefore T7 RNAp catalysed transcription reaction has become standard procedure used to synthesize RNA for many biochemical applications (analytical techniques, structural, biochemical, and genetic studies).[70]

Wild type T7 RNA polymerase accepts many different nucleotide modifications (**Figure 7**): amino acid residue,[71] ethynyl,[72] azido,[73] vinyl[74] and diazirine[75] groups as well as different, even bulky, fluorophores.[76,77] Recently, a systematic study of enzymatic synthesis of base modified RNA was reported by Milisavljevič et al.[78]

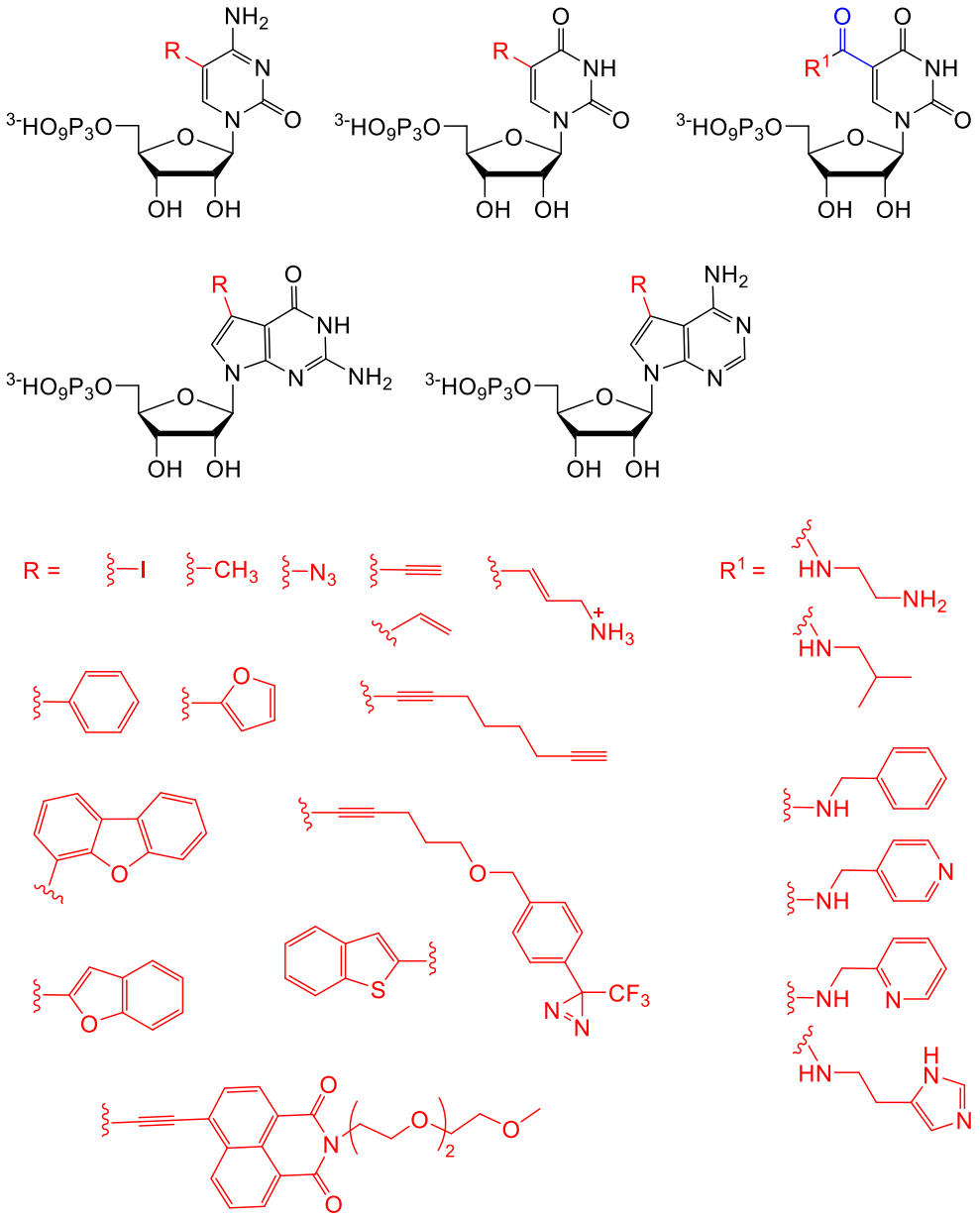


Figure 7. Examples of structures of modified triphosphates accepted by wild type T7 RNAp.

1.4 Nucleic acids – protein cross-linking

Cross-linking is the process of generating the covalent bond between two molecules (peptides, proteins, nucleic acids, drugs, small particles) with affinity among them (they aggregate or associate under certain conditions). It can be induced photochemically, or chemical cross-linking reagents are used.[79] The main challenge in this approach is to design and develop the functional group (cross-linking reagent/crosslinker) that is biorthogonal - it will not interfere with cellular processes, it is site-specific, adequately reactive under physiological conditions and nontoxic.

The cross-linking methods can be applied in diverse areas, spanned from classical protein biochemistry[80] to bioengineering,[81] medicine or immunology.[82–84]

1.4.1 Site-selective modifications of native proteins

Proteins are polymers consisting of 20 amino acids (AA) playing various essential roles in biological systems. They possess many nucleophilic side-chain functionalities that are tempting targets for chemical modification to study, confer, alter, or improve protein functions for diverse applications in biology or medicine.[85]

Chemical modification of proteins is not an easy task. The reactions must proceed at mild conditions while maintaining structure and function of the protein. However, many methods have taken advantage of the reactivity of endogenous amino acid side chains. Only few of total 20 amino acids involved in protein structure are suitable targets for bioconjugations. The traditional methods for chemical modification of natural amino acids use the nucleophilicity, solvent accessibility and high distribution frequency of lysine or cysteine residues. *N*-hydroxysuccinimide esters, maleimides, iodoacetamides are widely used as electrophiles for chemical modification.[86] AA residues of lysine, cysteine, tryptophan, tyrosine, methionine, and arginine are the most common targets as well as N and C termini of the proteins are often accessible and have different chemical environment than rest of the protein what makes them potentially relevant.[87–90] Selective targeting of AA residues in a protein chain is important

for various biological applications, such as inhibition or modulation of enzymatic activity.[91]

The plethora of methods available for selective reactions with primary amines and its high natural abundance (~ 6%) makes ε-amino group of lysine (Lys, K) a popular target for conjugations.[92] Lys can efficiently react with isocyanates and isothiocyanates, activated esters, sulfonyl halides and sulfonates. For conjugation of protein with a molecule with an aldehyde group, reductive amination with $NaBH_3CN$ (reference[93]) or iridium catalyst[94] or 6π-aza electron cyclic reaction[95] is often used (**Figure 8**). All these conventional amine reactive methods are, regardless of their name, not absolutely amine selective. The problem is caused by hydrolysis of activated reagents or if other stronger nucleophiles are present on protein surface (namely cysteine, serine, tyrosine, and threonine). These side reactions relay on the substrate, conjugated molecule, pH, buffer composition and temperature. Particularly buffers containing free amines must be avoided.

The thiol group of cysteine (Cys, C) is the most suitable target for protein bioconjugation due to its high nucleophilicity and low occurrence rate in proteins (1 - 2%).[96] The thiolate group can be directly labelled by alkylation with α-halocarbonyls (iodoacetamide), nucleophilic addition to maleimides or vinyl sulfones, chemically- or photoinduced thiol-ene/yne coupling or displacement reactions to form mixed disulphides (**Figure 8**).[86] Nucleophilic cysteine can be reversed via β-elimination reaction of thiolate to electrophilic dehydroalanine for wide variety of other post-translational modifications.[97]

Aromatic amino acids are rare on protein surfaces. Tryptophane (Trp, W) abundance in proteins is about 1%, but there is usually at least one in 90% of protein sequences.[98] Foettinger *et al.*[99] described the reaction of Trp containing octapeptide with malondialdehydes. In their work, they have solved the selectivity issues (side reactions with arginine); however, the reaction required strongly acidic conditions, which are not suitable for susceptible protein targets, yet applicable in proteomics. Another approach is to use metallocarbenoids, in particular $Rh_2(OAc)_4$ (**Figure 8**),[100] or transition-metal free method using keto-ABNO

(9-azabicyclo[3.3.1]nonane-3-one-*N*-oxyl] *N*-oxylradical[101] as an electrophile precursor, although there is the same high pH (pH 1.5-3.5) disadvantage.

Davis group recently reported radical protein trifluoromethylation approach to Trp modification using sodium trifluoromethanesulfinate (NaTFMS), tert-butyl hydroperoxide (TBHP) and a radical initiator. This method has preferred tryptophane over other canonical AA.[102]

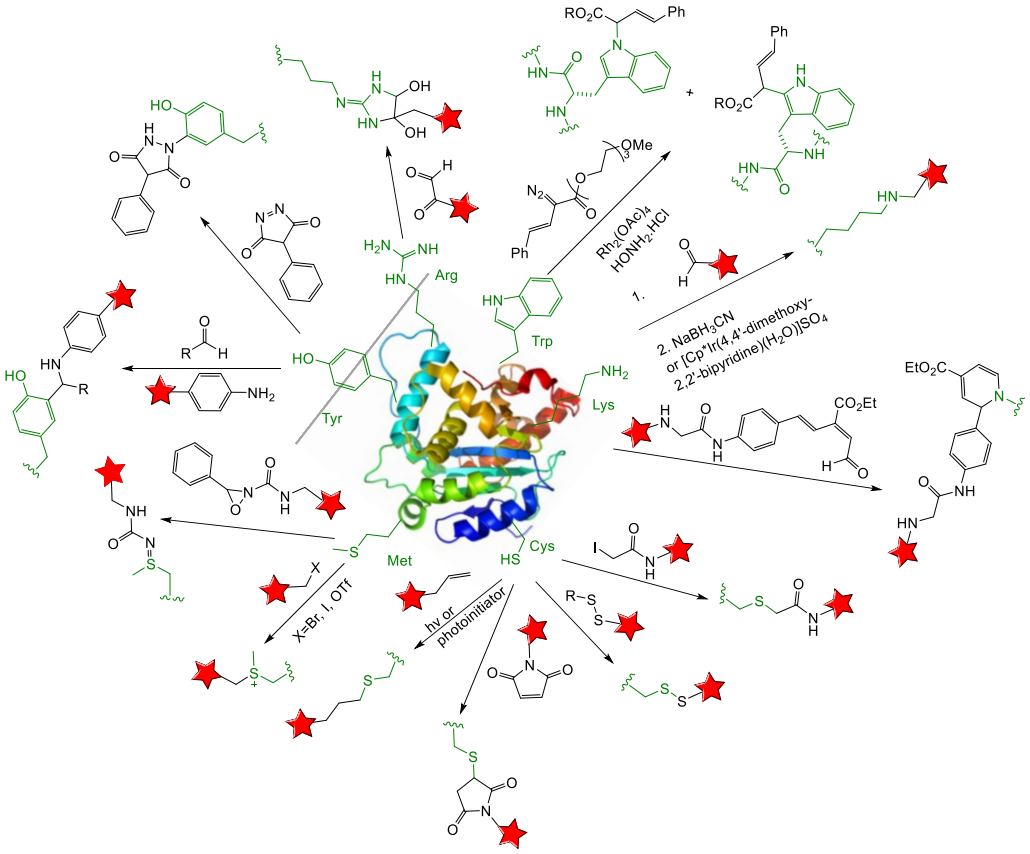


Figure 8. Examples of bioconjugation reactions of native protein AA sidechains. Protein model 2P51 generated from PDB.

The amphiphilic phenyl group of Tyrosine (Tyr, Y) is often hidden in the protein, but it is still interesting target for bioconjugation due to its electron rich aromatic ring and easily deprotonable phenolic group. Many methods for selective tyrosine modification have been developed. Namely, the Mannich reaction[103] of Tyr with aldehyde and aniline, azo coupling with diazonium salts,[104] allylic alkylation of the phenolic hydroxyl of Tyr with π–allyl palladium

complexes,[105] and click-like ene-type reaction with cyclic diazodicarboxamides (**Figure 8**).[106]

Methionine (Met, M) is weakly nucleophilic, redox-sensitive, second of only two sulphur containing AA with low abundance in proteins (~ 2%).[96,107] Despite its weak nucleophilicity at neutral pH, methionine is the only AA that can be alkylated at low pH. However, most proteins aggregate or lose activity in acidic environment. Epoxides, hypervalent iodine or oxaziridine are among other methionine modification reagents (**Figure 8**).[108]

The guanidium side chain of arginine (Arg, R) is the cationic group with lowest acidity among all amino acids. Majority of methodologies to label arginine use the α-dicarbonyl compounds chemistry[109] and germinal diones (**Figure 8**).[110,111] Recently, Gauthier and Klok reported site-selective PEGylation of proteins at arginine residues via α-oxo-aldehyde moiety.[112]

1.4.2 Site-selective cross-linking reactions of nucleic acids with proteins and their applications

Comprehension of the complex interaction motif of nucleic acids with proteins is essential for understanding of biological processes and disease mechanisms. Transcription factors,[113] reapair[114] and epigenetic modification[115] enzymes are just few of the DNA-binding proteins that are involved in regulation of gene expression. RNA-binding proteins play crucial role in post-transcriptional events, such as polyadenylation, splicing, RNA editing, mRNA stabilization, localization, and translation.[116] These cell processes are often disturbed by naturally occurring NA-protein cross-linking (NPC) which forms bulky lesions of covalently trapped biomolecules upon exposure to various endogenous, environmental, or chemotherapeutic agents.[117] If NPC is not repaired, it interferes with the replication and transcription processes which can result in cell mutations or apoptosis.[118]

Although cross-linking may be fatal for the cell, it has found many applications in research. X-ray crystallography, NMR, mass spectrometry, gene knockout or small molecular inhibitors are powerful tools to analyse NA-protein interactions, but chemical cross-linking makes possible to capture non-covalent interactions *in vivo*.[119]

Protein modifications (described in chapter 1.4.1) inspired many scientists studying NA-protein interactions to introduce electrophile warheads in the oligonucleotides. The cross-linker can be introduced at three feasible sites of the oligonucleotide at the 5'- or 3'-end, phosphodiester linkage, or sugar part of NA backbone or at the nucleobase (**Figure 9**). Nucleobase modification offers better programmability of the position of cross-linker in oligonucleotide sequence.

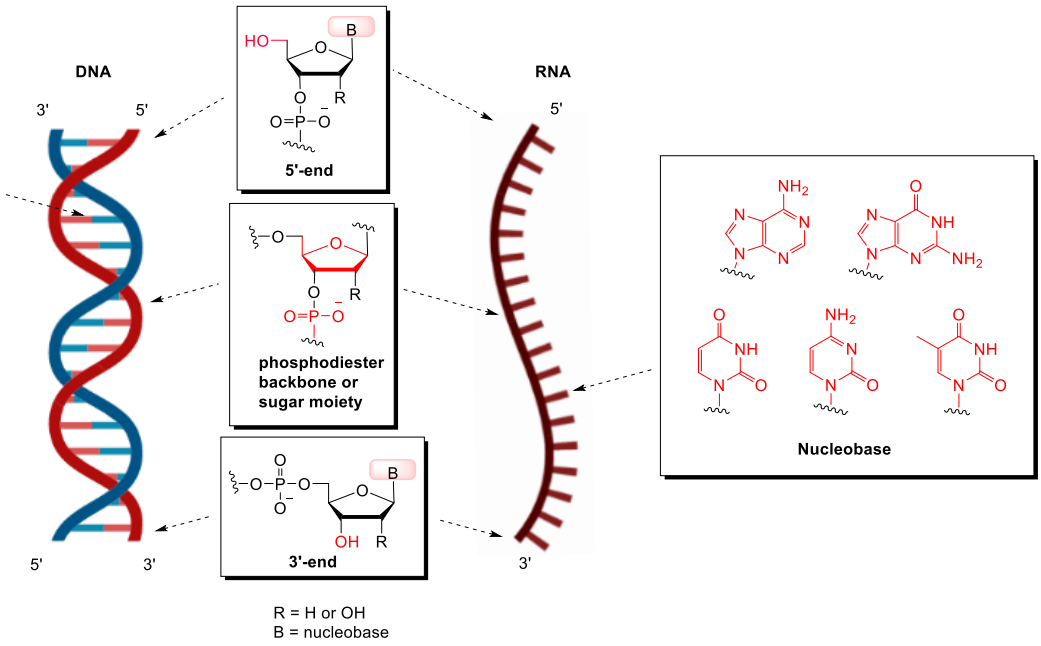


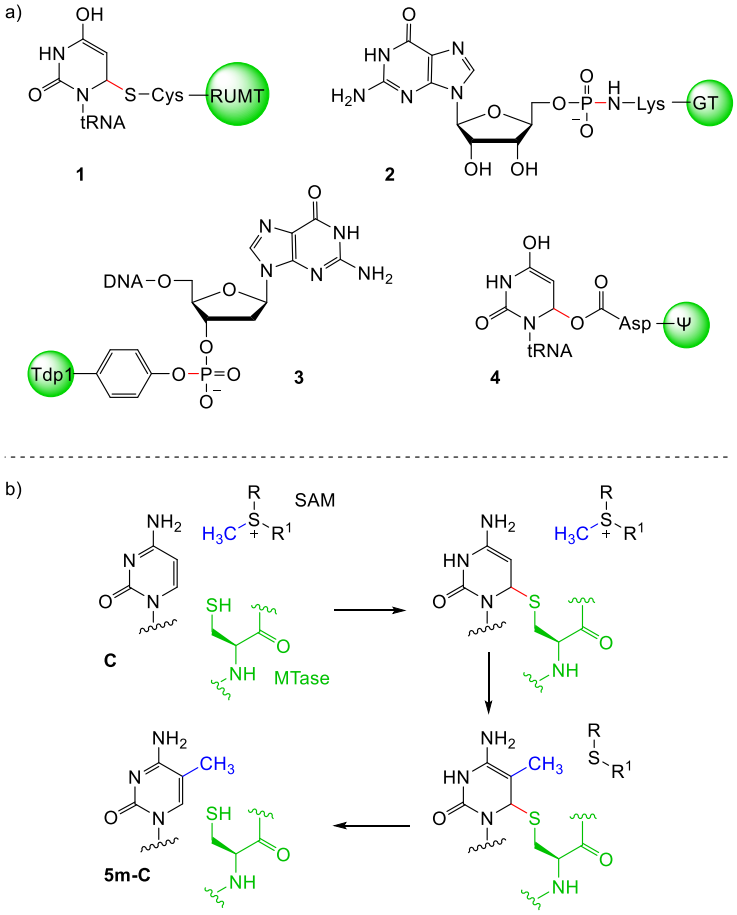
Figure 9. Feasible sites of chemical nucleic acids modifications.

1.4.2.1 Covalent complexes of nucleic acids and proteins

Many enzymatic reactions involve protein association with either DNA or RNA, some of them end up covalently linked to form stable covalent complexes. NA- protein covalent bonds are quite common in viral pathogens.[120]

Proteins involved in NA-protein complexes formation can be split into two classes. The first class is bound to the NA via nucleophilic attack of hydroxyl amino acid residue on a phosphate group of internucleotide. These proteins are polynucleotide transferases combining cleavage and restoration of internucleotide bond.[121] Type I topoisomerases form either 3' (eukaryotic Topo I) or 5' (bacterial Topo I) bound relaxed covalent complexes without the need of external energy source (e.g., ATP hydrolysis).[122] For example, covalent complex of RNA and poliovirus VPg RNA has cleavage-joining activity. It is attached to the 5' end of positive strand viral RNA and acts as a primer in RNA synthesis.[123] The second class of proteins involved in covalent bonding with NA does not catalyse formation and cleavage of covalent complex. Other enzymes (e.g., polio RNA polymerase catalyses covalent binding of class II proteins to 5' end of NA; unlinking enzymes catalyse hydrolysis of phosphodiester bonds between RNA and VPg or DNA and Topo I) do this. Class II proteins belong to the terminal proteins that initiate

reactions during replication of nucleic acids and need energy rich suppliers.[121] **Scheme 7** shows structures of some other examples of specific natural covalent complexes between RNA/DNA and enzymes.

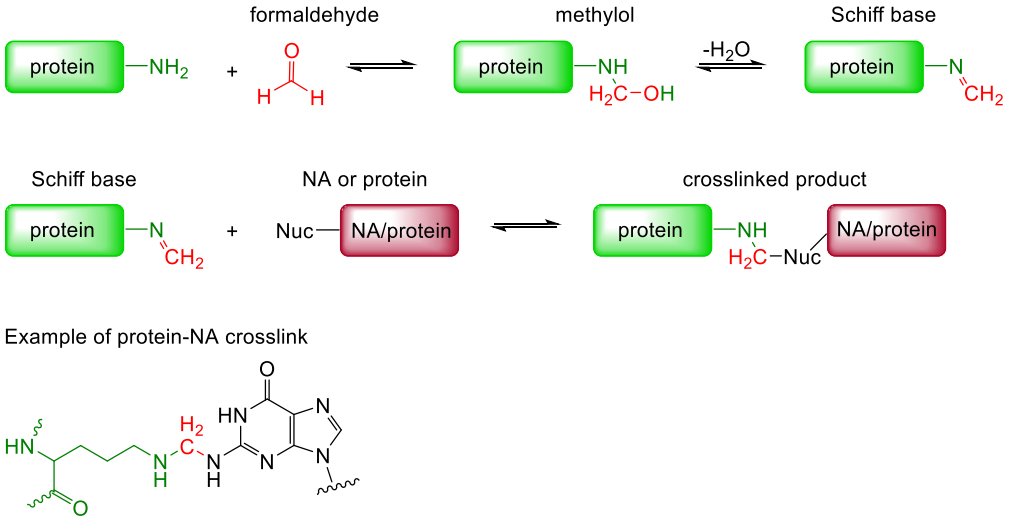


Scheme 7. Structures of RNA-enzyme covalent complexes (shared bond is highlighted in red). a) tRNA U54 methyltransferase (RUMT);[124] Guanylyl transferase (GT);[125] Tyrosyl-DNA-phosphodiestherase 1 (TdP1);[126] Pseudouridine synthase (Ψ).[127] b) Mechanism of DNA methylation by (cytosine-5)-methyltransferase.[128]

Various external reactive agents such as aldehydes, reactive oxygen species (ROS), metal ions or radiation can generate nonspecific cross-linking of proteins to nucleic acids.[118]

Aldehydes are generated naturally from histone demetylation (formaldehyde),[129] or as a product of ethanol oxidation (acetaldehyde).[130] Formaldehyde cross-linking method has been employed for many decades to trap nucleic acids-protein complexes for structure analysis since it is reactive towards

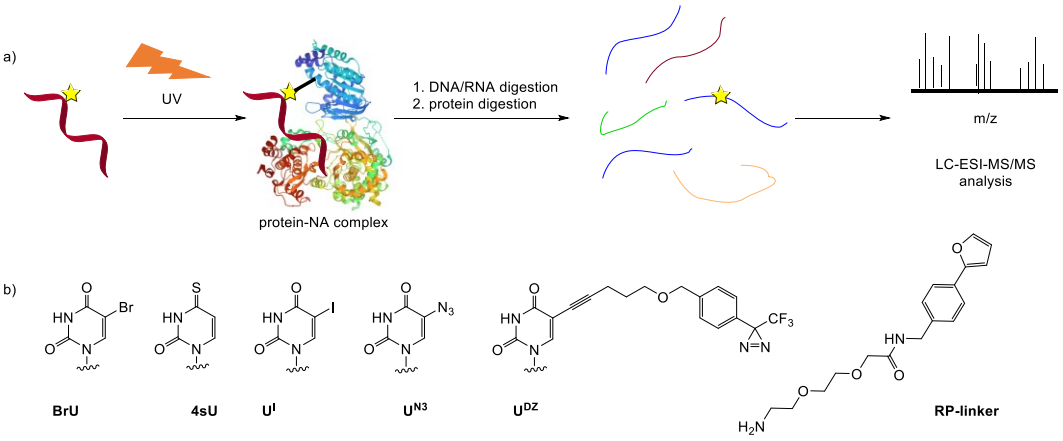
both macromolecules.[131] The reaction of formaldehyde with NA and proteins proceeds in a few steps (**Scheme 8**). It includes formation of covalent methylol adduct between nucleophilic group of AA (e.g. Lys, Arg) or DNA/RNA base (e.g. cytosine), which is then converted to a Schiff base. Methylols and Schiff base can be stabilized by formation of a methylene bridge[132] with another amino group of second molecule. This reversible cross-linking approach has been widely used for *in vivo* studying of DNA-protein interactions by chromatin immunoprecipitation,[133,134] which can be coupled with mass spectrometry (MS) for better definition of binding sites.[135–137] Reversible cross-linking affinity purification (RCAP) was developed as a modification of chromatin immunoprecipitation to map the RNA-protein binding regions *in vitro*.[138,139]



Scheme 8. Formaldehyde cross-linking reaction of biomolecules.

UV cross-linking method (**Scheme 9**) has been widely used by biochemists and molecular biologists to study structure of NA-protein complexes such as ribosomes or chromatin since early 1960s.[140] While the natural binding is maintained, the UV wavelengths that induce NA-protein cross-linking (~ 250 - 280 nm) provide low yields and can damage the molecules (photocleavage or oxidation) what makes this approach limited.[141] UV light generates covalent bond between NA and protein that are in close proximity during light activation. Many photoreactive groups that absorb at wavelengths > 300 nm was introduced into DNA or RNA to overcome damage caused by UV light of harmful wavelength during *in vivo* experiments. Halides,[142] sulphur atoms,[143] aryl azides[144] and

diazirines[145,146] have been used as photoreactive functional groups in bioconjugation chemistry of DNA or RNA with proteins. The drawback of these moieties is the lack of selectivity for the particular amino acid. Very recently, photocatalytic cross-linking strategy (PhotoCAX) was reported for studying RNA-protein interactions in living cells by Chen et al.[147]



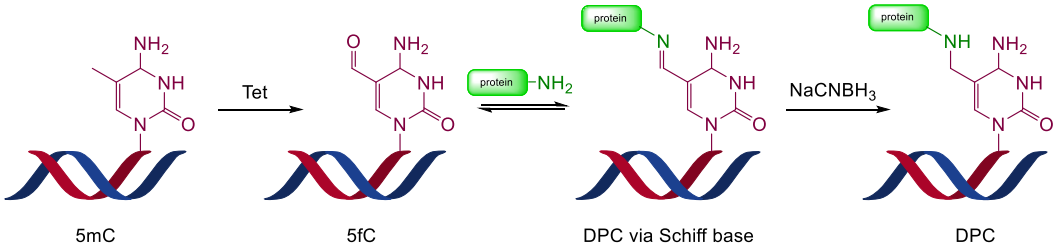
Scheme 9. a) General scheme for photo-induced NA-protein crosslinking for MS analysis (protein image is generated from PDB 4K6M). b) Examples of photoreactive functional groups (BrU = 5-bromouridine; 4sU = 4-thiouridine; U^{I} = 5-iodouridine; U^{N3} = 5-azidouridine; U^{DZ} = Uridine analogue with diazirine moiety; RP-linker for PhotoCAX strategy).

Ribo- or deoxyoligonucleotides modified with reactive groups effectively cross-link with proteins due to the proximity effect with the binding proteins. Diol[148] and furan[149] reactive groups activated by oxidation react specifically with Lys or Arg side chains of peptides or proteins.

Hocek lab has developed polymerase synthesis of reactive DNA probes for specific cross-linking with natural side chains of AA of protein. Functional groups targeting cysteine (vinylsulfonamide[150] and chloroacetamide[151] moiety), histidine and arginine (1,3-diketone[152] and glyoxal[153] functional group) were developed (**Scheme 11**). Matyašovský et. all reported that oligodeoxynucleotides with 2-substituted 2'-deoxyinosine 5'-triphosphates react with SH-containing minor groove binding peptide via proximity effect.[154] Lysine was approached through reductive amination with aldehyde[155] and diol[156] modified DNA. In 2020, Krömer et. all[157] synthesized 2-formyl-2'-deoxyadenosine modified oligonucleotide used

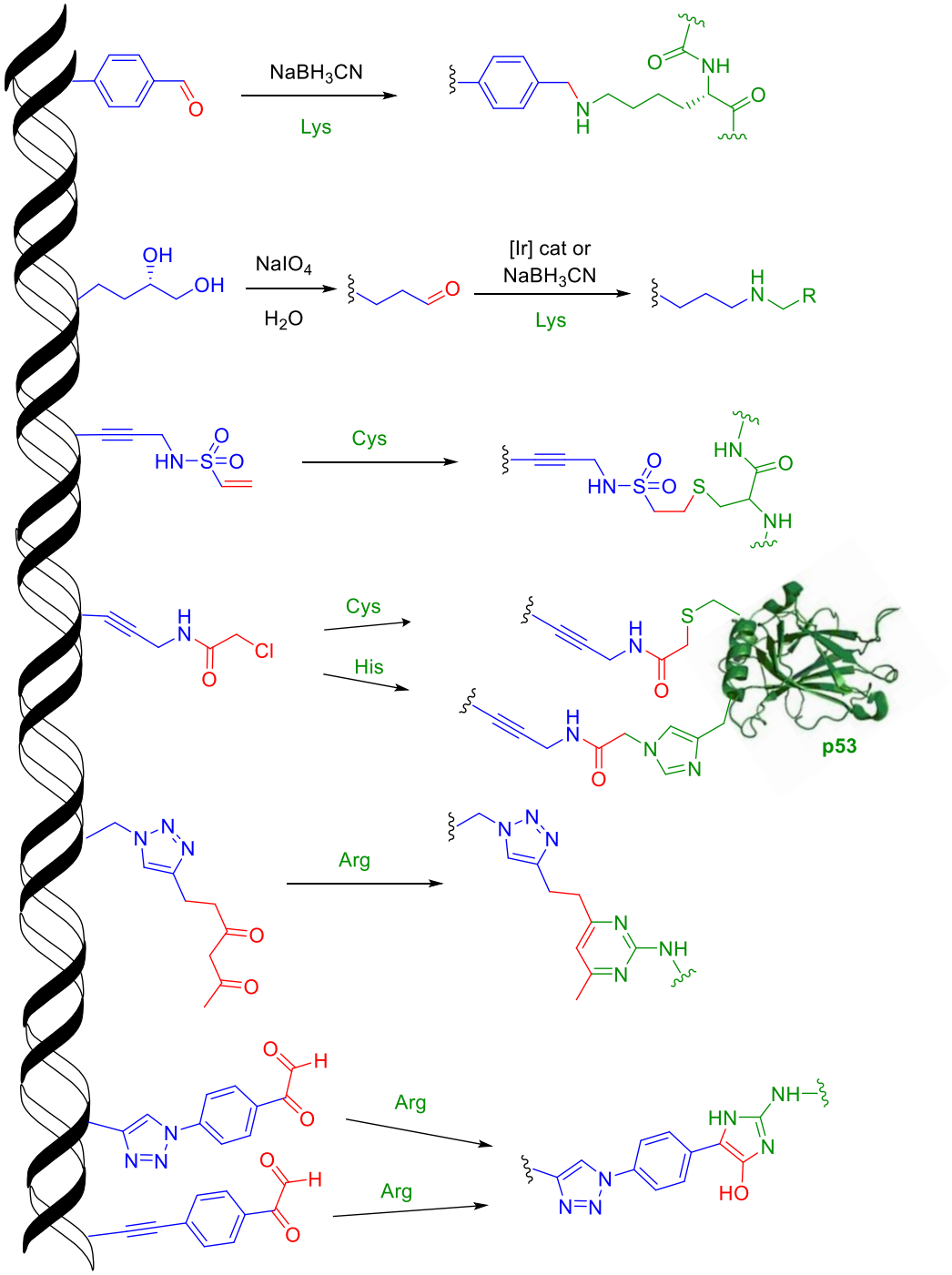
for cross-linking with proteins via reductive amination and for fluorescent labelling with AlexaFluor647 hydroxylamine through oxime formation.

The group of Tretyakova has been intensively studying DNA-protein cross-links (DPCs) of endogenous 5-formylcytosine base (5fC) lysine of histones via reductive amination (**Scheme 10**).[148,158,159] 5fC plays crucial role in epigenetic control of gene expression and its formation was reported as well in nucleosome core particles by Li et al.[160]



Scheme 10. Formation of Schiff-base between 5fC of DNA and Lys side chain of protein and its stabilization via reductive amination using $NaBH_3CN$.

So far, all attempts to target natural Lysine residue in peptides and proteins require further reductive ($NaBH_3CN$, iridium catalysts) or oxidative (NBS) reagents to stabilize the cross-linked complex. These reagents are usually toxic, therefore incompatible with cell experiments. It is desired to develop alternative methodology without the need of external reagents.

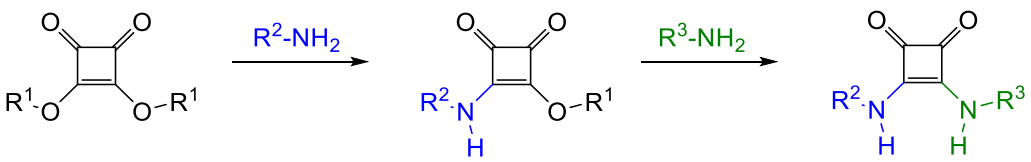


Scheme 11. Major groove DNA-protein crosslinking strategies developed in Hocek group.

1.5 Squaramide: a tool for bioconjugation

Squaramides are exceptional, conformationally rigid four-member ring derivatives of squaric acid are widely used in many areas of the chemical and biological sciences.[161] Special physical and chemical properties (hydrogen bonding, aromatic switching, stability, structural rigidity)[162] and facile synthesis make this small molecule useful in various areas such as supramolecular (self-assembly) chemistry,[163–165] catalysis,[166] anion recognition (anion receptors,[167] sensors,[168–170] transporters[171,172]) and medicinal chemistry.[173–175] Asymmetric squaramides can be synthesized under relatively mild or aqueous conditions, making them ideal for bioconjugation of small molecules, biomolecules, or polymers.[162]

N-hydroxysuccinimidyl esters, aldehydes, isocyanates, and sulfonyl chlorides are often used in post-modification reactions of proteins targeting the NH_2 group of lysine residues. These strategies often suffer from side reactions, reversibility, low hydrolytic stability of the protein or need of additional reducing agents.[176] An alternative approach is two-step coupling of squaric acid diester with two amines (**Scheme 12**). The mono-squaramate after first step has reduced reactivity for the further nucleophilic attack due to higher electron density in the cyclobut-3-ene-1,2-dione ring, allowing generation of unsymmetric squaramides with high yields and selectivity. These reactions are selective towards amines and tolerate many functional groups (hydroxyl, arginine, histidine).[177,178] N-terminal cysteine residues can form reversible thioesters, otherwise are thiol groups tolerated.[179]



Scheme 12. General scheme of consecutive amidation of squarates leading to asymmetrical bisamides of squaric acid.

Squarate moiety was used in nucleic acid conjugation reactions as a phosphate replacement at the 5'-end of oligonucleotide.[180,181] Etheve-Quelquejeu and her co-workers introduced squaramide monoester as an electrophilic site at the 3'-end of RNA. They chemically synthesized 2'-sugar connected squaramate-RNA that was crosslinked to a UDP-MurNAc-pentapeptide (peptidoglycan precursor of aminoacyl-transferase $FemX_{wv}$). The peptidyl-RNA inhibited $FemX_{wv}$ by cross-linking with catalytic Lys 305 residue. This was so far the first example of squaramate-nucleic acids conjugation (**Figure 10**).[182]

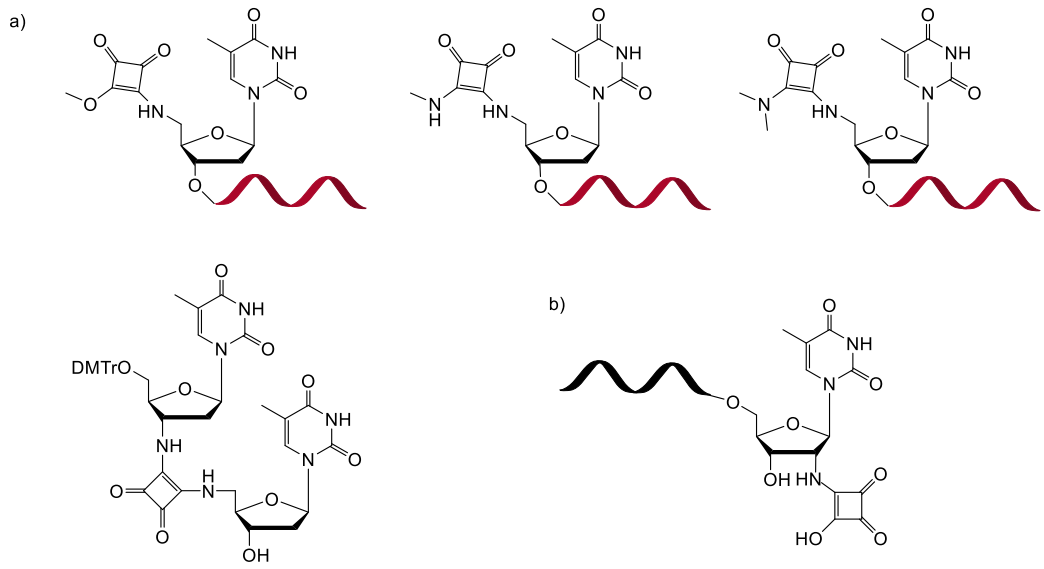


Figure 10. Structures of modified oligonucleotides with squaramate/squaramide a) instead of phosphate group at 5'-end and b) 2'-sugar linked squaramate-RNA.

2. Specific aims of the thesis

1. Design and enzymatic synthesis of DNA probes bearing reactive modifications for amine-selective bioconjugation with DNA binding proteins.
2. Testing of cross-linking reactions of reactive DNA probes with DNA binding proteins.
3. Design and enzymatic synthesis of RNA probes bearing reactive modifications for amine-selective bioconjugation with RNA binding proteins.
4. Testing of cross-linking reactions of reactive RNA probes with RNA binding proteins.

2.1 Rationale of the specific aims

DNA-protein interactions play crucial role in DNA packaging, replication, transcription, epigenetic modifications, and repair.[183] RNA-binding proteins are involved in almost all vital processes of cellular biology (modulation of RNA metabolism, RNA editing, regulation of gene expression).[116] Although, there are many methods how to study nucleic acids-protein interactions and to identify DNA/RNA-binding proteins, there are still alternative methods required, notably for weakly binding proteins. Covalent cross-linking is useful tool for identification of NA-binding proteins as well as for other applications in medicine, immunology,[184] chemical biology[185] or biosensing.[186]

Many photoreactive functional groups (5-halouracil,[187] diazirines[145]) are used in bioconjugation chemistry, but they bind non-specifically to neighbouring AA. To target specifically one or several AA sidechains is useful approach, but only few DNA-protein cross-linking reactions have been reported (see chapter 1.4.2). Reductive amination is prevalent method used for targeting lysine in proteins. Aldehyde-containing DNA is cross-linked with Lys via reversible Schiff base formation,[188,189] and irreversibly reduced to a secondary amine[190,148,156] using stoichiometric reductant $NaBH_3CN$ which is incompatible with *in vivo* or *in cellulo* applications. Until now, there has been no reactive nucleobase modification in NA reported for irreversible cross-linking with lysine without the need for external reagent. Therefore, the aim of this thesis was to develop

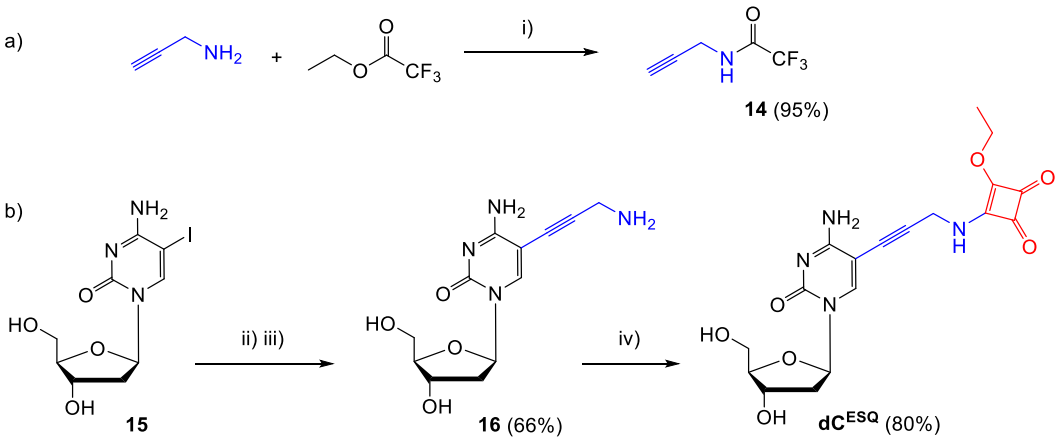
reactive DNA and RNA probe, which will react selectively with lysine without use of any further reagents. For this purpose, we have chosen squaramate moiety as it is known to be selective towards primary amines, reacts without any external additives and it tolerates many functional groups.[177,178] Research in our group is aimed at base-modified nucleic acids for application in chemical biology.[191] Within this framework, we designed new squaramate modified (2'-deoxy)cytidine triphosphate which can be used for enzymatic synthesis of modified nucleic acid probes for cross-linking with proteins.

3. Results and discussion

3.1 Enzymatic synthesis of reactive DNA probes for cross-linking with proteins

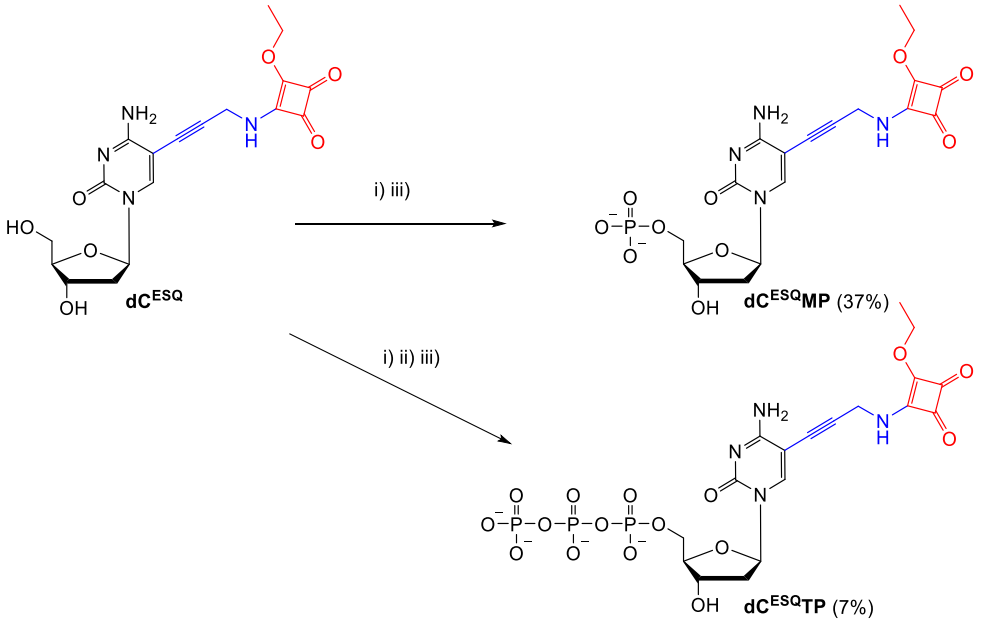
3.1.1 Synthesis of squaramate-modified deoxynucleoside (dC^{ESQ}) and its mono- ($dC^{ESQ}MP$) and triphosphate ($dC^{ESQ}TP$). Model reactions using $dC^{ESQ}MP$ as a substrate

The squaramate moiety was designed to be tethered to the position 5 of cytosine via propargylamine linker. First, protected propargylamine (**14**)[192] was prepared by treatment with ethyl trifluoroacetate (95%, **Scheme 13-a**) and then used in Sonogashira cross-coupling reaction with 5-iodo-2'-deoxycytidine (**15**) in the presence of $Pd(PPh_3)_4$ and Amberlite Ira-67 resin.[193] Subsequent deprotection of known trifluoroacetamide[194] in aqueous ammonium hydroxide gave 5-(3-aminopropynyl)-2'-deoxycytidine (**16**) in acceptable yield 66%. The reaction of amine **16** with 2 equiv of commercially available diethyl squarate in ethanol proceeded smoothly and gave the mixed squaramate nucleoside **dC^{ESQ}** in 80% yield (**Scheme 13-b**). As expected,[162] there was no formation of symmetric bis-squaramide side product.



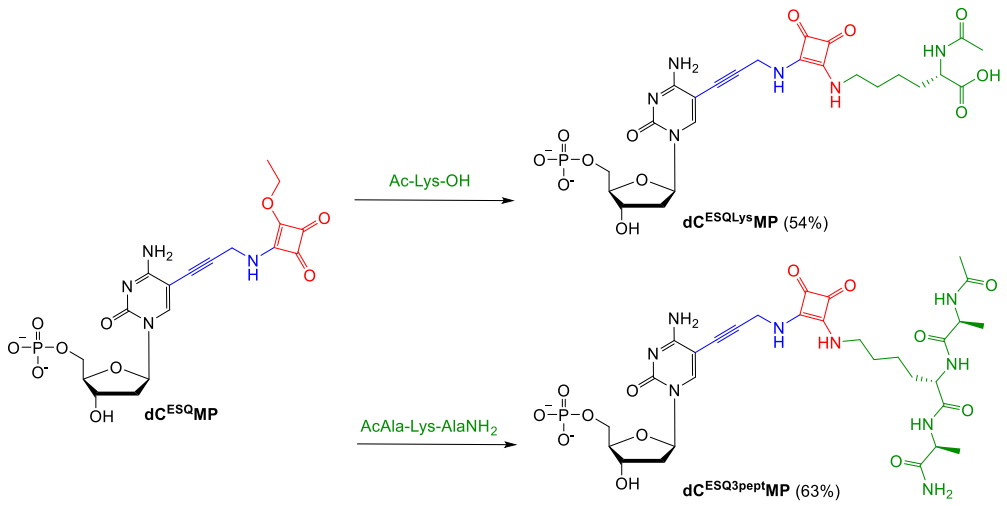
Scheme 13. Synthesis of a) trifluoroacetamide **14** and b) squaramate modified 2'-deoxycytidine (**dC^{ESQ}**). Conditions: i) EtOH, r.t., dark, 1h; ii) amide **14**, $Pd(PPh_3)_4$ (10%), CuI (20%), Amberlite Ira-67, DMF, r.t., 24h; iii) H_2O/NH_4OH, r.t., 24h; iv) diethyl squarate (2equiv), EtOH, r.t., 1h.

Monophosphate **dC^{ESQ}MP** was prepared according to standard Yoshikawa phosphorylation[41] with $POCl_3$ in 37% yield, while the reaction with $POCl_3$ followed by pyrophosphate[195] and triethylammonium bicarbonate (TEAB) yielded the triphosphate **dC^{ESQ}TP** (7%, **Scheme 14**).



Scheme 14. Synthesis of squaramate modified 2'-deoxycytidine mono- (**dC^{ESQ}MP**) and triphosphate (**dC^{ESQ}TP**). Conditions: i) $POCl_3$, $PO(OMe)_3$, 1-2h, 0 °C; ii) $(NHBu_3)_2H_2P_2O_7$, DMF, Bu_3N, 45 min, 0 °C; iii) 2 M TEAB, H_2O.

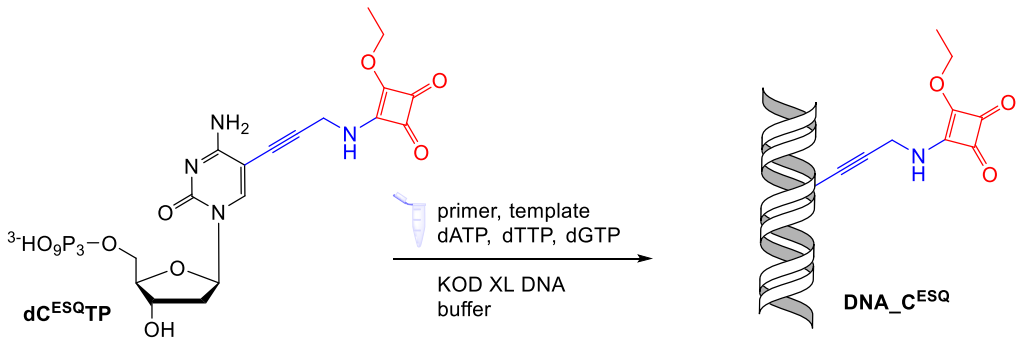
The modified **dC^{ESQ}MP** was used as a surrogate of DNA in model reaction to evaluate the reactivity of this mixed squaramate towards Lys and Lys-containing peptide. The reaction of **dC^{ESQ}MP** was performed with 2-fold excess of N_α-acetyl-L-lysine and tripeptide (for sequence see **Table 3**) at room temperature overnight (**Scheme 15**). As it is known from the literature, in the second amidation step the addition of a base or alkaline aqueous conditions are needed,[196] that is why the reaction was performed in borate buffer at pH 9. The desired conjugates **dC^{ESQLys}MP** and **dC^{ESQ3pept}MP** were isolated by reverse-phase HPLC in 54 and 63% yields, respectively. Characterization by NMR spectroscopy and mass spectrometry (MS) confirmed the formation of amide bond with NH_2 of lysine.



Scheme 15. Model reactions of **dC^{ESQ}MP** with N_α-acetyl-L-lysine and tripeptide. Conditions: borate buffer pH 9 (0.5 M), r.t., overnight.

3.1.2 Incorporation of dC^{ESQ}TP into DNA by PEX and PCR

The squaramate modified dNTP (**dC^{ESQ}TP**) was then assessed as a substrate for synthesis of modified DNA catalysed by KOD XL DNA polymerase (**Scheme 16**).



Scheme 16. Scheme of the enzymatic synthesis of squaramate modified DNA (**DNA_C^{ESQ}**).

First, the primer extension reaction (PEX) was performed with 19-, 20-, 31- and 98-mer templates (for sequences see **Table 1**). The primers were labelled at 5'-end with 6-carboxyfluorescein (6-FAM) for further visualisation of formed products. In all cases the polymerase gave full length oligonucleotides (ON) containing 1, 4 or 18 **dC^{ESQ}** modifications. Formation of all ONs was characterised by denaturing polyacrylamide gel analysis PAGE (**Figure 11**).

Table 1. List of oligonucleotides used or synthesized in this study.[a]

ON	Sequence (5' → 3')	Length
prim^{A}	5′-AGACATGCCTAGA-3′	13
prim^{B}	5′-CATGGGCGGCATGGG-3′	15
prim$^{REV_LT25\text{-}TH}$	5′- CAAGGACAAAATACCTGTATTCCTT-3′	25
temp^{19_1C}	5′-CCCGCCCATGCCGCCCATG-3′	19
temp^{20_1C}	5′-AAACATGTCTAGGCATGTCT-3′	20
temp^{31_4C}	5′-CTAGCATGAGCTCAGTCCCATGCCGCCCATG-3′	31
temp$^{FVL\text{-}A}$	5′-GACATCATGAGAGACATCGCCTCTGGGCTAAT	98
ON_C^{ESQ_19}	5′-CATGGGCGGCATGGG**C**GGG-3′	19
ON_C^{ESQ}	5′-AGACATGCCTAGA**C**ATGTTT-3′	20
ON_4C^{ESQ}	5′-CATGGGCGGCATGGA**C**TGAG**C**T**C**ATG**C**TAG -3′	31
DNA_C^{ESQ_19}	3′-GTACCCGCCGTACCCGCCC-5′	19
DNA_C^{ESQ}	3′-TCTGTACGGATCTGTACAAA-5′	20
DNA_4C^{ESQ}	3′-GTACCCGCCGTACCCTGACTCGAGTACGATC-5′	31
DNA_C^{ESQ_98}	3′-GTTCCTGTTTTATGGACATAAGGAACGGACAGG TCCCTAGACGAGAATGTCTAATCTTCATCAGGATAA TCGGGTCTCCGCTACAGAGAGTACTACAG-5′ 5′-CAAGGACAAAATACCTGTATTCCTTG**C**C**C**TGT**C**C AGGGAT**C**TG**C**T**C**TTA**C**AGATTAGAAGTAGT**C**C**C**TAT TAG**C**C**C**AGAGG**C**GATGT**C**T**C**T**C**ATGATGT**C**-3′	98

[a]primers are underlined, nucleotides bearing modification are in bold red

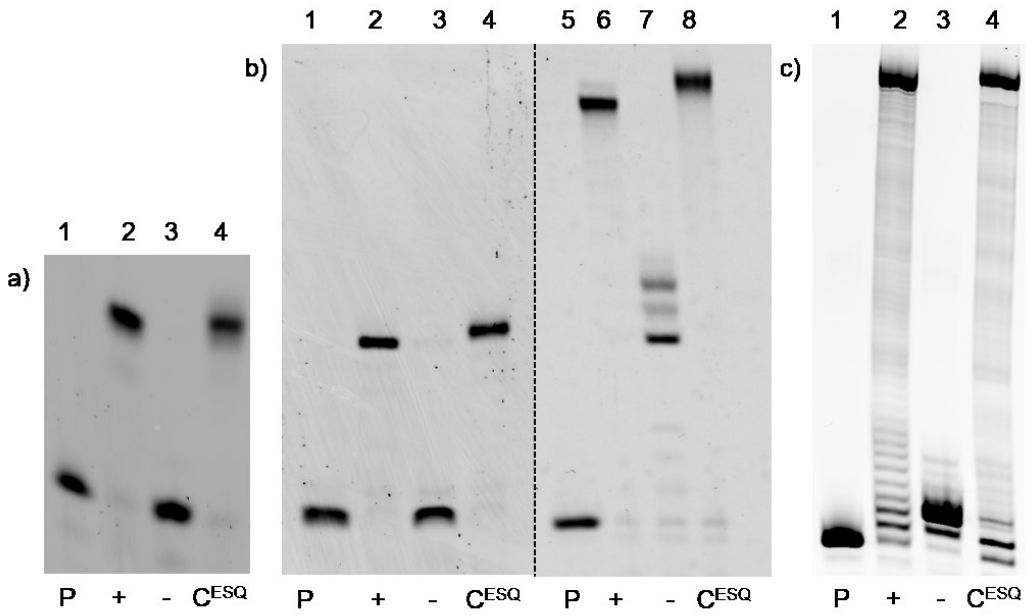


Figure 11. Primer extension using KOD XL polymerase and temp^{19_1C} (a; lines 2-4), temp^{20_1C} (b; lines 2-4), temp^{31_4C} (b; lanes 6-8) and temp$^{FVL\text{-}A}$ (c; lanes 2-4). a) Lane 1 (P): primer^{B}; lane 2 (+): natural dNTPs; lane 3 (-): negative control without dCTP; lane 4 (C^{ESQ}): **dC^{ESQ}TP**, dGTP. b) Lane 1 (P): primer^{A}; lane 5 (P): primer^{B}; lanes 2,6 (+): natural dNTPs; lanes 3,7 (-): negative control without dCTP; lanes 4,8 (C^{ESQ}): **dC^{ESQ}TP**, dGTP, dTTP, dATP. c) Lane 1 (P): prim$^{REV_LT25\text{-}TH}$; lane 2 (+): natural dNTPs; lane 3 (-): negative control without dCTP; lane 4 (C^{ESQ}): **dC^{ESQ}TP**, dGTP, dTTP, dATP.

From our experience, it is sometimes problematic to obtain MALDI-TOF mass spectra of ONs longer than 30 nucleotides. Only short ONs were confirmed by MALDI-TOF analysis (**Table 4**). The modified DNA for mass analysis was prepared by PEX using 5'-end biotinylated templates followed by magnetoseparation[197] of **ON_C^{ESQ}_19** and **ON_C^{ESQ}**.

As the next step, I tested the **dC^{ESQ}TP** in more demanding polymerase chain reaction (PCR). In the PCR, the enzyme must not only accept the modified nucleotide for the synthesis of modified strand (as it is in PEX), but it must tolerate the modification in the template strand as well. PCR allows enzymatic synthesis of long dsDNA molecules highly modified in both strands. PCR amplification of 98- and 235-bp templates (**Table 2**) in the presence of KOD XL polymerase proceeded well, giving full length dsDNA products in both cases (**Figure 12**).

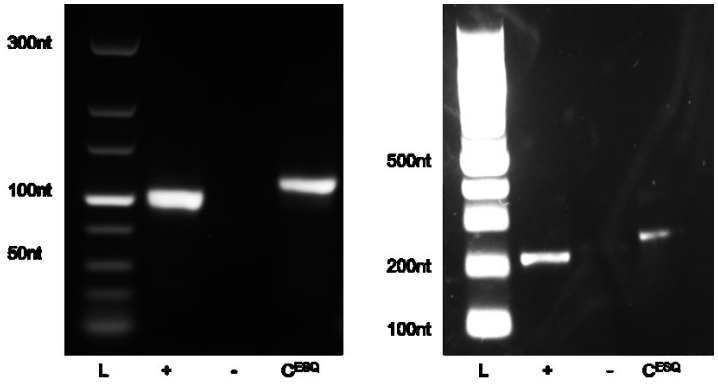


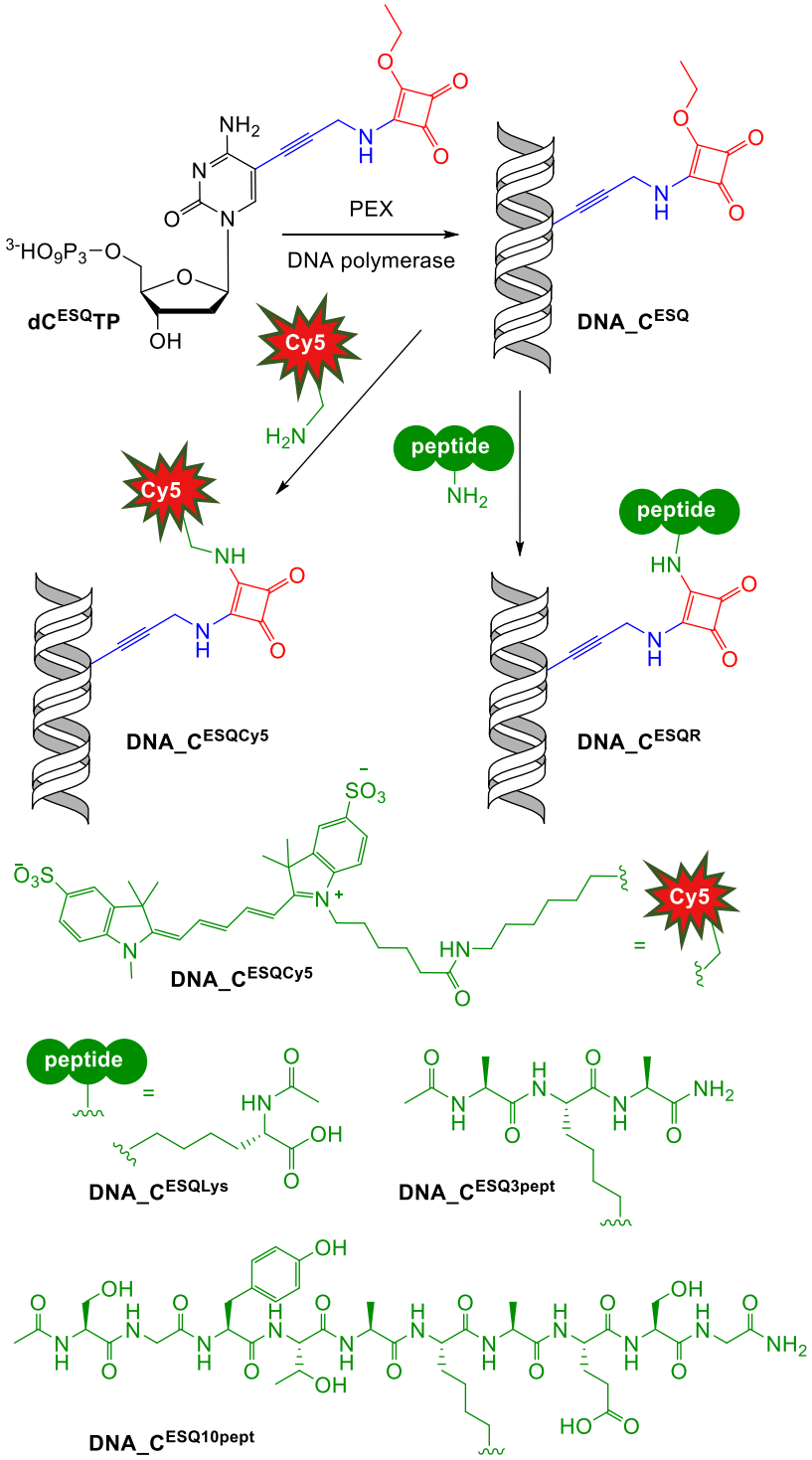
Figure 12. Agarose gel analysis of PCR product amplified by KOD XL DNA polymerase; (L): ladder; (+): natural dNTPs; (-): negative control without dCTP; (C^{ESQ}): **$dC^{ESQ}TP$,** dATP, dGTP, dTT; 2% (1.3%) agarose gel for 98-mer (235-mer) stained with GelRed.

Table 2. List of oligonucleotides used in PCR experiments.

ON	Sequence (5' → 3')	Length
$prim^{FOR\_L20}$	5′- GACATCATGAGAGACATCGC-3′	20
$prim^{REV\_LT25\text{-}TH}$	5′- CAAGGACAAAATACCTGTATTCCTT-3′	25
$prim^{FOR\_235}$	5′-CGTCTTCAAGAATTCTATTTGACA-3′	24
$prim^{REV\_235}$	5′-GGAGAGGTTCACCGACA-3′	18
$temp^{FVL\text{-}A}$	5′-GACATCATGAGAGACATCGCCTCTGG GCTAATAGGACTACTTCTAATCTGTAAGAGCAGAT CCCTGGACAGGCAAGGAATACAGGTATTTTGTCCTT G-3′	98
$temp^{PCR\_235}$	5′-CGTCTTCAAGAATTCTATTTGACAAAAATGGGC TCGTGTTGTACAATAAATGTGTCTAAGCTTGGGTCC CACCTGACCCCATGCCGAACTCAGAAGTGAAACGC CGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGC GAGAGTAGGGAACTGCCAGGCATCAAATAAAACG AAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT CTGTTGTTTGTCGGTGAACGCTCTCC-3′	235

Successful PEX and PCR experiments confirm that ESQ-modification does not react with DNA polymerase during the reaction and that the **$dC^{ESQ}TP$** is a good substrate and building block for the enzymatic synthesis of reactive modified DNA.

3.1.3 Reaction of DNA_C^{ESQ} with sulfo-Cy-5-amine, lysine and lysine containing model peptides



Scheme 17. Synthesis of squaramate-modified DNA (**DNA_C^{ESQ}**) and its cross-linking reactions with sulfo-Cy-5-NH_2, N_α-acetyl-L-lysine, tri- and decapeptide. Conditions of cross-linking: borate buffer (pH 9, 0.5 M), 37 °C, 36 h.

Double stranded squaramate-modified DNA bearing one modification (**DNA_C^{ESQ}**) was prepared by PEX in semipreparative scale and tested in conjugation reactions with sulfo-Cy-5-NH_2 dye, N_α-acetyl-L-lysine, lysine containing tri- and decapeptide (**Scheme 17**; for peptide sequences see **Table 3**).

All reactions were carried out in borate buffer (pH 9) at 37 °C for 36 h. Large excess of amines (100 equiv of the dye and 2500 equiv of AA and peptides) was used because there was no proximity effect expected in this non-DNA-binding molecules.

The reaction of natural 20-bp DNA and **DNA_C^{ESQ}** (both non-labelled) with sulfo-Cy-5-NH_2 was analysed by measuring the fluorescence spectra of the reaction mixtures purified after incubation with the dye. Only in the case when **DNA_C^{ESQ}** was used, the fluorescence signal of Cy-5 fluorophore was detected (**Figure 13**). The reaction was analysed also on denaturing PAGE. For this, 5'-6-FAM natural and squaramate modified DNA was prepared and incubated with sulfo-Cy-5-NH_2 at the same conditions. After incubation, the reaction mixtures were purified to get rid of excess of unreacted fluorescent dye, loaded on the polyacrylamide gel and visualised by fluorescence imaging. Newly formed spot with lower mobility visible both in FAM and Cy-5 channels was observed only in the case of **DNA_C^{ESQ}** (C^{ESQ} line in **Figure 13**).

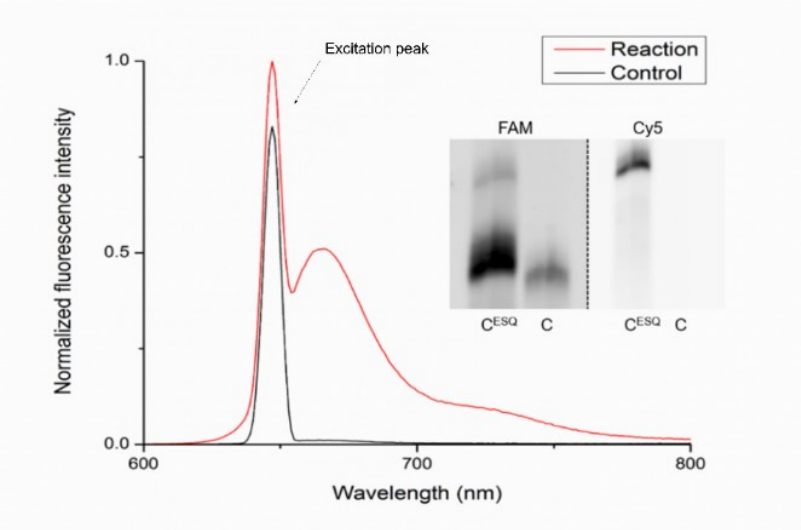


Figure 13. Conjugation of natural and ethoxy squarate modified **DNA_C^{ESQ}** with sulfo-Cy-5-amine. Conditions: natural DNA (Control line in black) or **DNA_C^{ESQ}** (Reaction line in red), borate buffer (0.5 M, pH 9), 37 °C, and 36 h. Fluorescence of the products was measured on spectrofluorometer (water; λ_{ex} = 646 nm, λ_{em} = 662 nm). PAGE analysis: C line: FAM labelled natural DNA + sulfo-Cy-5-NH_2; C^{ESQ} line: FAM labelled **DNA_C^{ESQ}** + sulfo-Cy-5-NH_2. Excited with 473 nm for FAM and 635 nm for Cy-5.

The reactions of the **DNA_C^{ESQ}** with N_α-acetyl-L-lysine and Lys-containing peptides proceeded with moderate efficacy and gave the desired conjugates **DNA_C^{ESQLys}**, **DNA_C^{ESQ3pept}** and **DNA_C^{ESQ10pept}** with 50, 43 and 20% conversions, respectively (**Table 3**). 5'-6-FAM products were identified as bands with lover mobility on PAGE (**Figure 14**) and non-labelled **ON_C^{ESQLys}**, **ON_C^{ESQ3pept}** and **ON_C^{ESQ10pept}** conjugates were confirmed by MALDI-TOF (**Table 4**).

Table 3. List of amino acids and synthetic peptides used to test cross-linking abilities of **DNA_C^{ESQ}**.

	Peptide sequence	Molecular weight	Conversion
Acetyl-lysine		188.23	50%
tripeptide	Ac-AKA-NH_2	329.2	43%
decapeptide	Ac-SGYTAKAQSG-NH_2	1011.47	20%

Note: Conversions were determined from the gel using ImageJ Quantificator.

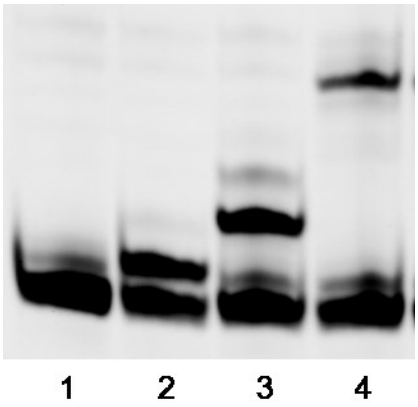


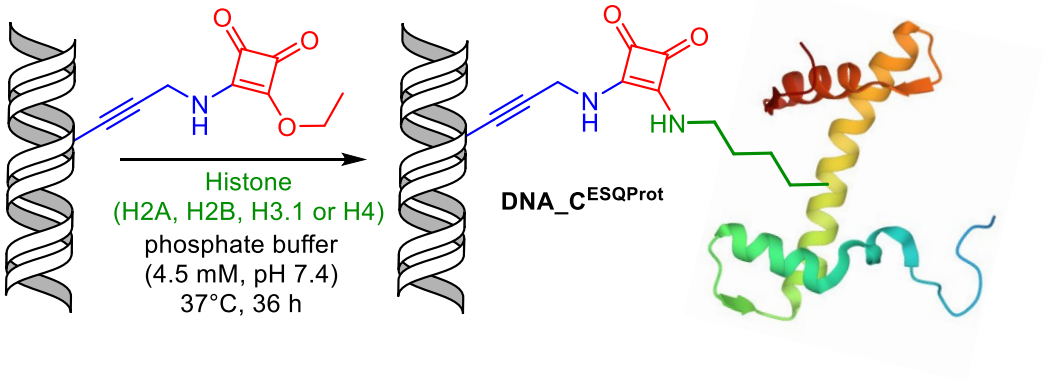
Figure 14. Conjugation of **DNA_C^{ESQ}** (4.3 μM; line 1) with Ac-lysine (11 mM; line 2), tripeptide (11 mM; lane 3) and decapeptide (11 mM; line 4). Conditions: borate buffer (0.5 M, pH 9), 37 °C, 36 h.

Table 4. MALDI-TOF data of modified oligonucleotide **ON_C^{ESQ_19}**, **ON_C^{ESQ}** and adducts of **DNA_C^{ESQ}** with Ac-lysine, tripeptide and decapeptide (for MALDI spectrum see **Figures A1-A3**). Single stranded conjugates were obtained after magnetoseparation.

ON	M (calc.) [Da]	M (found) $[M + H]^+$ [Da]
ON_C^{ESQ_19}	6127.9	6128.1
ON_C^{ESQ}	6293.3	6294.3
ON_C^{ESQLys}	6435.5	6436.4
ON_C^{ESQ3pept}	6576.5	6576.6
ON_C^{ESQ10pept}	7257.8	7258.0

3.1.4 Conjugation of ESQ modified DNA with proteins

For the bioconjugation experiments, GSTp53CD was kind donation from Dr. Marie Brázdová from the group of Doc. Miroslav Fojta at the Institute of Biophysics AS CR in Brno.



Scheme 18. General scheme of cross-linking of squaramate modified DNA (**DNA_C^{ESQ}**) with histone recombinant proteins (H4 protein representative from PDB 1EQZ chain F).

Having the reactive **DNA_C^{ESQ}** and suitable conjugation procedures in hand, I shifted focus to the cross-linking of squaramate modified DNA with a set of recombinant histones (H2A, H2B, H3.1, H4) as model Lys-rich DNA-binding proteins. Bovine serum albumin (BSA) contains 60 lysine residues, of which 30 - 35 have primary amines available for post-modification and does not interact with DNA, therefore it was used as a negative control. The core domain of p53 (GSTp53CD) was also used as negative control as it is sequence-specific DNA binding protein that contains lysines, but not in the proximity to the modification in binding site.

All cross-linking reactions were performed with the DNA containing one modification and 2 equiv of the proteins. Since the histones usually dimerize, in this case the ratio can be assumed as equimolar. The reactions were carried out at physiological conditions at 37 °C in phosphate buffer pH 7.4.

First, I performed a kinetic study of the cross-linking reaction of **DNA_C^{ESQ}** with H3.1 protein. The reaction reached the maximum conversion in 16-24 h as shown in **Figure 15**. Based on this result, I have used the 36 hours incubation in all other experiments to ensure ample conversions.

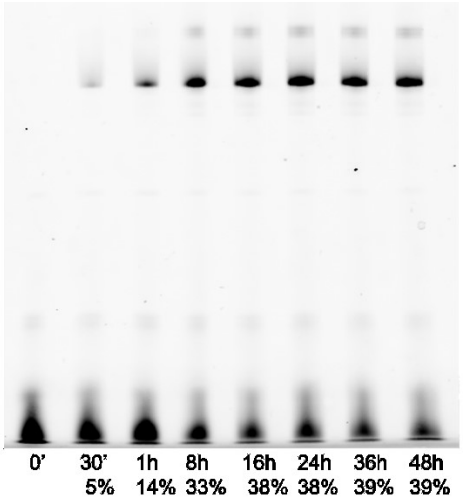


Figure 15. Effect of reaction time on cross-link formation between **DNA_C^{ESQ}** and H3.1. Conditions: 2 equiv of protein, phosphate buffer (4.5 mM, pH 7.4), 37 °C. SDS-PAGE analysis on 17.5% SDS gel. Conversions were determined from the gel using ImageJ Quantificator.

In addition, I also examined the effect of DNA/protein molar ratio, temperature and various buffers on ESQ mediated cross-linking efficacy. The highest conversions were achieved at 37 °C and 10 equiv of protein to DNA (**Figure 16-a, b**). TRIS and HEPES buffers were tested to see if the reaction also proceeds in the presence of tertiary amines (TRIS) or free OH groups (HEPES). These buffers did not interfere with the reaction and the conversions (34% for HEPES and 29% for TRIS) were comparable to the ones achieved in phosphate buffer (**Figure 16-c**).

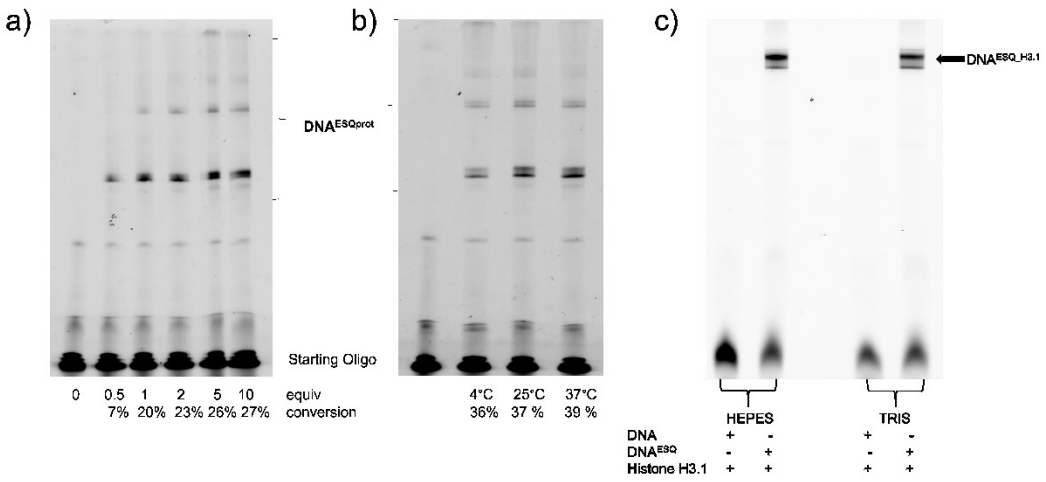


Figure 16. a) SDS-PAGE analysis of DNA-protein cross-link between **DNA_C^{ESQ}** and different amounts of H2.A protein. Conditions: phosphate buffer (4.5 mM, pH 7.4), 36 h. SDS-PAGE analysis on 17.5% SDS gel. b) SDS-PAGE analysis of DNA-protein cross-link between **DNA_C^{ESQ}** and H2.A protein at various temperatures. Conditions: 2 equiv of protein, phosphate buffer (4.5 mM, pH 7.4), 36 h. SDS-PAGE analysis on 17.5% SDS gel. c) Reaction of the FAM labelled **DNA_C^{ESQ}** with histone H3.1 using different buffers. Conditions: 2 equiv. of protein, HEPES buffer (100 mM, pH 7.4) or TRIS buffer (100 mM, pH 7.4), 37 °C, 36 h. SDS-PAGE analysis on 17.5% SDS gel. Conversions: HEPES: 34%, TRIS: 29%. All conversions were determined from the gel using ImageJ Quantificator.

Figure 17 shows the results of ESQ-mediated DNA-protein cross-linking experiments. The formation of covalent **DNA_C^{ESQ}**-protein conjugates (**DNA_C^{ESQprot}**) with all four histone monomers was detected as reduced mobility bands on denaturing SDS-PAGE gel. Strong **DNA_C^{ESQprot}** bands were observed when **DNA_C^{ESQ}** was incubated with histone proteins (**Figure 17-a**, lanes 6,8,10,12), but none were spotted in control experiment when natural DNA and BSA was used or in the absence of proteins (**Figure 17-a**, lanes 1-5, 7, 9, 11). The conversions of the reactions were 31 – 34% (**Table 5**). The bands with lower mobility were confirmed to be DNA_**C^{ESQprot}** also by SDS-PAGE with PageBlue^{TM} protein staining (**Figure 17-b**) and by HPLC-MS analysis using electrospray ionization (ESI, **Figure A4 - A6**, **Table 6**).

So far, all these experiments were performed with **DNA_C^{ESQ}** bearing one modification, therefore I evaluated also longer 98-bp PEX product containing 18 squaramate groups (**DNA_C^{ESQ_98}**) in cross-linking with histone H3.1, however mixture of cross-linked products was obtained (**Figure 17-c**).

To conclude, squaramate modified DNA did not cross-link with BSA or GSTp53CD (**Figure 17-d**) nor with DNA polymerase during PEX or PCR reactions, showing that proximity effect is vital for effective cross-linking in the absence of large excess of the peptide or the protein.

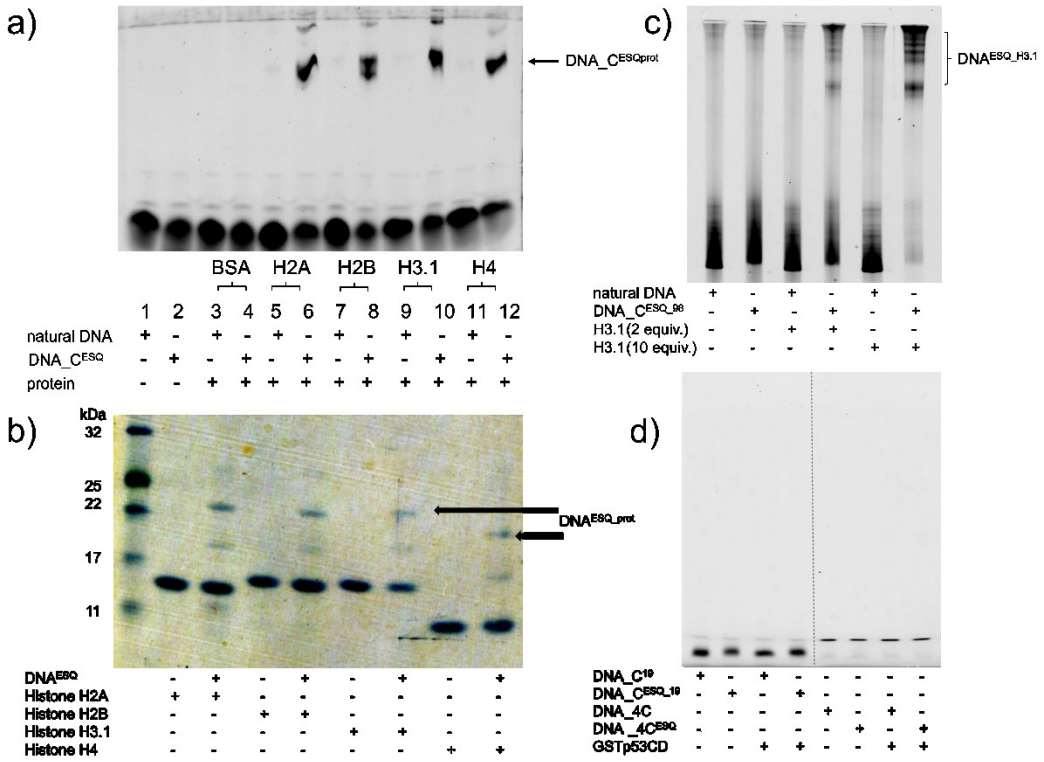


Figure 17. a) SDS analysis of control experiment with FAM labelled natural or modified DNA and various recombinant proteins. b) SDS-PAGE analysis of DNA-protein cross-link between **DNA_C^{ESQ}** and recombinant proteins followed by PageBlue^{TM} staining. Conditions: 2 equiv of protein, phosphate buffer (4.5 mM, pH 7.4), 37 °C, 36 h. SDS-PAGE analysis on 17.5% SDS gel. c) SDS analysis of cross-links of natural or modified 98-bp PEX product and H3.1 protein. Conditions: 2 resp. 10 equiv of protein, phosphate buffer (4.5 mM, pH 7.4), 37 °C, 36 h. SDS-PAGE analysis on 4 - 12% SDS gel. d) SDS analysis of control experiment with natural or modified DNA and GST_p53_CD protein. Conditions: HEPES (2.5 mM, pH 7.6), 37 °C, 36 h. SDS-PAGE analysis on 10% SDS gel.

Table 5. DNA-protein cross-link conversions between **DNA_C^{ESQ}** and recombinant proteins.

Protein	Molecular weight [kDa]	DNA binding	No of Lys	Conversion %
BSA	69.3	no	60	0%
GST_p53_CD	78.5	yes	21	0%
H2A	13.99	yes	13	34%
H2B	13.79	yes	20	34%
H3.1	15.27	yes	13	31%
H4	11.24	yes	11	34%

Note: Conversions were determined from the gel using ImageJ Quantificator.

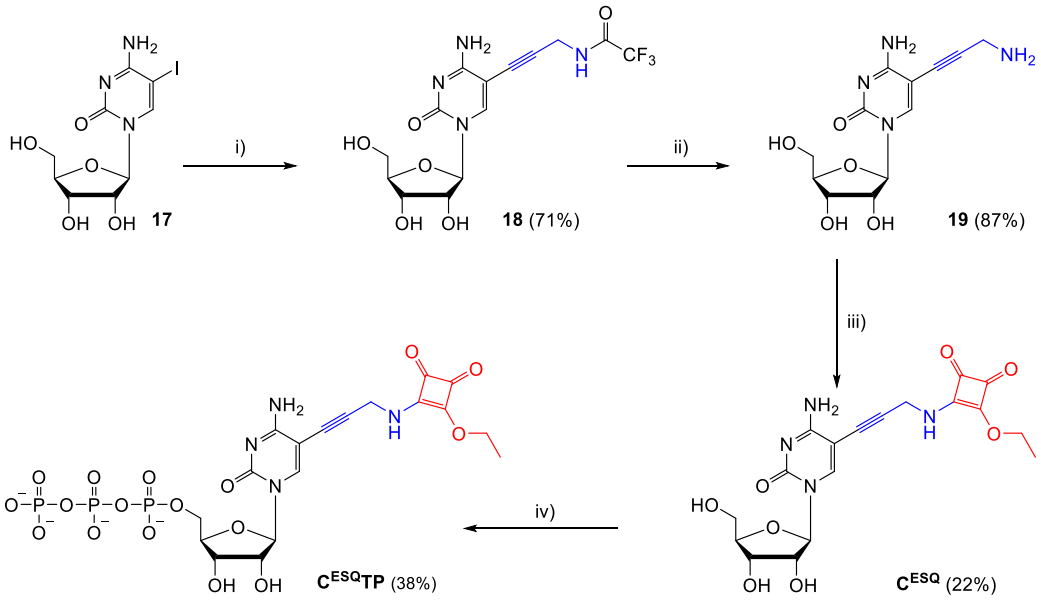
Table 6. MS data of DNA-protein conjugates (for MS spectrum see **Figures A4-A6**).

ON	M (calc.) [Da]	M (found) [Da]
ON_C^{ESQH2B}	20036.78	19905
ON_C$^{ESQH3.1}$	21520.63	21388
ON_C^{ESQH4}	17484.11	17351

3.2 Enzymatic synthesis of reactive RNA probes for cross-linking with proteins

3.2.1 Synthesis of squaramate-modified ribonucleoside (C^{ESQ}) and its triphosphate ($C^{ESQ}TP$)

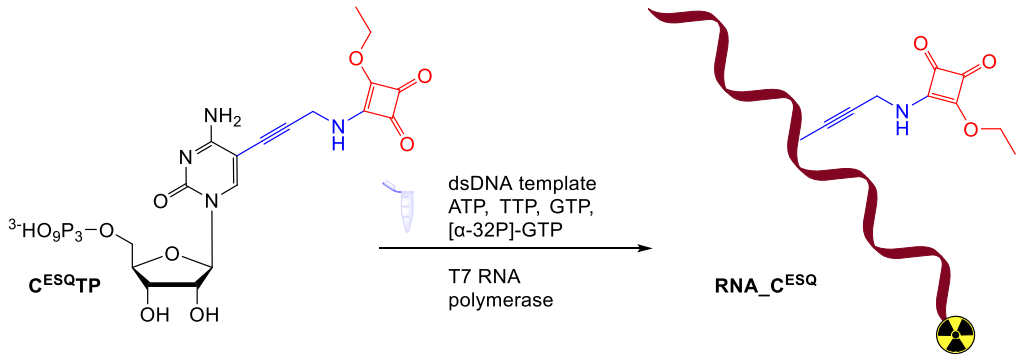
The squaramate modified ribonucleotide was synthesized in a similar way as squaramate deoxycytidine triphosphate (section 3.1.1), starting from Sonogashira reaction of 5-iodo-cytidine (**17**) with *N*-propargyl trifluoroacetamide (**14**; **Scheme 19**). Deacylation of trifluoroacetamide **18** gave 5-(3-aminopropynyl)-cytidine (**19**) in excellent yield (87%). The reaction of amine **19** with 2 equiv of diethyl squarate gave squaramate-linked ribonucleoside (**C^{ESQ}**) in acceptable yield (22%). The triphosphorylation of **C^{ESQ}** with $POCl_3$ in trimethyl phosphate followed by treating with tributyl ammonium pyrophosphate gave the corresponding **C^{ESQ}TP** (38%) after reverse-phase HPLC purification.



Scheme 19. Synthesis of squaramate modified cytidine (**C^{ESQ}**) and cytidine triphosphate (**C^{ESQ}TP**). Conditions: i) amide **14**, $Pd(PPh_3)_4$ (10%), CuI (20%), Amberlite Ira-67, DMF, r.t., 24h; ii) H_2O/NH_4OH, r.t., overnight; iii) diethyl squarate (2equiv), EtOH, r.t., 1h; iv) 1. $POCl_3$, $PO(OMe)_3$, 90 min, 0 °C; 2. $(NHBu_3)_2H_2P_2O_7$, ACN, Bu_3N, 1 h, 0 °C; 3. 2 M TEAB, H_2O.

3.2.2 Incorporation of synthesized $C^{ESQ}TP$ into RNA

In vitro transcription experiments were carried out to investigate the ability of T7 RNA polymerase to recognise the **$C^{ESQ}TP$** and site-specifically incorporate it into sequence of interest (**Scheme 20**).



Scheme 20. General scheme of the enzymatic synthesis of squaramate modified RNA (**RNA_C^{ESQ}**) by *in vitro* transcription reaction.

3.2.2.1 Design of dsDNA templates, transcription sequences and optimisation of conditions

For initial experiments, to test my triphosphate as a substrate for T7 RNAp, I used the DNA templates already available in our group which would result in 35mer RNA transcripts with one, three and seven modifications in the strand (**DNA_1C^{ESQ}**, **DNA_3C^{ESQ}**, **DNA_7C^{ESQ}**; for sequences see **Table 7**).[78] dsDNA templates were prepared by annealing of two complementary DNA strands (52mer) which contained T7 promoter from the positions - 17 to + 1 and guanosines at positions + 1 to + 3 to effectively start the transcription.[64] The last two 5'-terminal nucleotides in the antisense strand have 2'-methoxy group (MeO) instead of 2'-OH to minimize non-templated nucleotide addition.[198]

Table 7. Sequences of ds DNA template and RNA transcripts.[a]

	Sequence (5' → 3')	Length
DNA_1C^{ESQ}	5′-TAATACGACTCACTATAGGGAGGGACTGTGAGTGG AGATTGTAGGATTGAGG-3′	52
	3′-ATTATGCTGAGTGATAT**C**CCTCCCT**G**ACACTCACC TCTAACATCCTAACT[mC][mC]-5′	52
RNA_1C^{ESQ}	5′-pppGGGAGGGA**C**UGUGAGUGGAGAUUGUAGGAUU GAGG-3′	35
DNA_3C^{ESQ}	5′-TAATACGACTCACTATAGGGAGGATCAGTACAGAG GTATGCTGGGATAGGGA-3′	52
	3′-ATTATGCTGAGTGATAT**C**CCTCCTA**G**TCAT**G**TCTC CATAC**G**ACCCTATCCT[mC][mU]-5′	52
RNA_3C^{ESQ}	5′-pppGGGAGGAU**C**AGUA**C**AGAGGUAUG**C**UGGGAUA GGGA-3′	35
DNA_7C^{ESQ}	5′-TAATACGACTCACTATAGGGGATGTCTGAGCCTAA GGATGCCAACTGTTGTC-3′	52
	3′-ATTATGCTGAGTGATAT**C**CCCTACA**G**ACTC**GG**ATT CCTAC**GG**TT**G**ACAAC[mA][m**G**]-5′	52
RNA_7C^{ESQ}	5′-pppGGGGAUGU**C**UGAG**CC**UAAGGAUG**CC**AA**C**UGUU GU**C**-3′	35
DNA_C8d^{ESQ}	5′-TAATACGACTCACTATAGATACTAAGCCAAGAAG TTCACACAGATAAACTTCT-3′	53
	3′-ATTATGCTGAGTGATAT**C**TAT**G**ATTC**GG**TTCTTC AA**G**T**G**T**G**TCTATTT**G**AA[m**G**][mA]-5′	53
RNA_C8d^{ESQ}	5′-pppGAUA**C**UAAG**CC**AAGAAGUU**C**A**C**A**C**AGAUA AA**C**UU**C**U-3′	36
DNA_C1e^{ESQ}	5′-TAATACGACTCACTATAGATAGTAAGGCATAAA TTGAGAGAGATAAATTTAT-3′	52
	3′-ATTATGCTGAGTGATAT**C**TATCATTCC**G**TATTT AACTCTCTCTATTTAAA[mU][mA]-5′	52
RNA_C1e^{ESQ}	5′-pppGAUAGUAAGG**C**AUAAAUUGAGAGAGAUAAA UUUAU-3′	35

DNA_C3f^{ESQ}	5′-TAATACGACTCACTATAGATACTAAGCCAATAAA TTGAGAGAGATAAATTTAT-3′	53
	3′-<u>ATTATGCTGAGTGATAT</u>**C**TAT**G**ATTC**GG**TTATTT AACTCTCTCTATTTAAA[mU][mA]-5′	53
RNA_C3f^{ESQ}	5′-pppGAUA**C**UAAG**CC**AAUAAAUUGAGAGAGAUAAA UUUAU-3′	36
DNA_C1g^{ESQ}	5′-TAATACGACTCACTATAGATACTAAGGGAATAAA TTGAGAGAGATAAATTTAT-3′	53
	3′-<u>ATTATGCTGAGTGATAT</u>**C**TAT**G**ATTCCCTTATTT AACTCTCTCTATTTAAA[mU][mA]-5′	53
RNA_C1g^{ESQ}	5′-pppGAUA**C**UAAGGGAAUAAAUUGAGAGAGAUAAA UUUAU-3′	36

[a] Promoter region in the antisense strand is underlined. The two 5'-terminal nucleotides in the antisense strand were 2'-MeO ribonucleotides (m) to minimize non-templated nucleotide addition. The first base after the promoter region is marked bold in magenta, while those that are directing transcription of modified bases are marked bold in blue, whereas the modified bases in the transcripts are marked bold in red.

Transcription reactions were performed with templates **DNA_1/3/7C^{ESQ}** in the presence of ATP, UTP, GTP, CTP/modified CTP (**C^{ESQ}TP**) and spiked with radioactive [α-^{32}P]-GTP. Water was used instead of the **C^{ESQ}TP** in the negative control experiment. The products were resolved by 12.5% PAGE and visualized by autoradiography (**Figure 18-a**). The full-length products were observed in all cases, although the negative control in the case of the template for RNA with one modification (**RNA_1C^{ESQ}**) revealed the product of misincorporation.

To fully confirm the formation of modified RNA, the purified transcripts were submitted to MALDI-TOF MS analysis, however there was possibility that RNA might hydrolyse due to the measurement conditions. Although that was not the case and I could observe the peaks in MS spectrum of **RNA_1C^{ESQ}**, the mass 12052.1 Da, however, it did not correspond to the calculated one (11954 Da; **Figure 18-c**). After excluding possible misincorporations, I found out that the peak at *m/z* = 12052.1 can be assigned to the product of the reaction

of **RNA_1C^{ESQ}** with spermidine (primary amine, M = 145.25 Da; **Figure 18-d**) from the reaction buffer supplied with T7 RNAp by manufacturer. Since the squaramate moiety is known to prefer primary amines to the more substituted ones, to overcome this undesired reaction, I substituted the spermidine with quaternary amine betaine (**Figure 18-d**). To compare the reaction effectivity, I prepared the same transcription buffers as the supplied one (1×; 40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl) with 2mM betaine (buffer B) instead of 2 mM spermidine (buffer A) and with absence of amine (buffer C). All transcription reactions, analysed on 20% denaturing PAGE, gave the full-length products (**Figure 18-b**), although the misincorporation in all negative controls was observed as well. In the case of buffer A the + 1 product is observed, which is the RNA-spermidine conjugate. There seemed to be no difference in intensity of the product spots between buffer with betaine and without amine. The product of the transcription reaction B was confirmed by MALDI to be the squaramate modified **RNA_1C^{ESQ}** only (**Figure 18-e**).

As it is known, the presence of stimulatory natural polyamines, especially spermidine, affects the in vitro T7 RNAp transcription system[64,199] and since there was no significant difference between buffer B and C, I have chosen to proceed further with the buffer containing betaine (buffer B).

All transcription reactions of templates **DNA_1/3/7C^{ESQ}** were repeated with the conditions described above but with the home-made reaction buffer with 2 mM betaine. The products were resolved by 12.5% PAGE and visualized by autoradiography (**Figure 19-a**). The full-length products were successfully confirmed by MALDI-TOF for the **RNA_1C^{ESQ}** and **RNA_3C^{ESQ}** but not in the case of **RNA_7C^{ESQ}** (**Table 8**).

Table 8. MALDI-TOF data of modified **RNA_1C^{ESQ}** and **RNA_3C^{ESQ}**.

Transcript	M (calc.) [Da]	M (found) [Da]
RNA_1C^{ESQ}	11954	11955.1 $[M + H]^+$
RNA_3C^{ESQ}	12277	12279.3 $[M + 2H]^+$

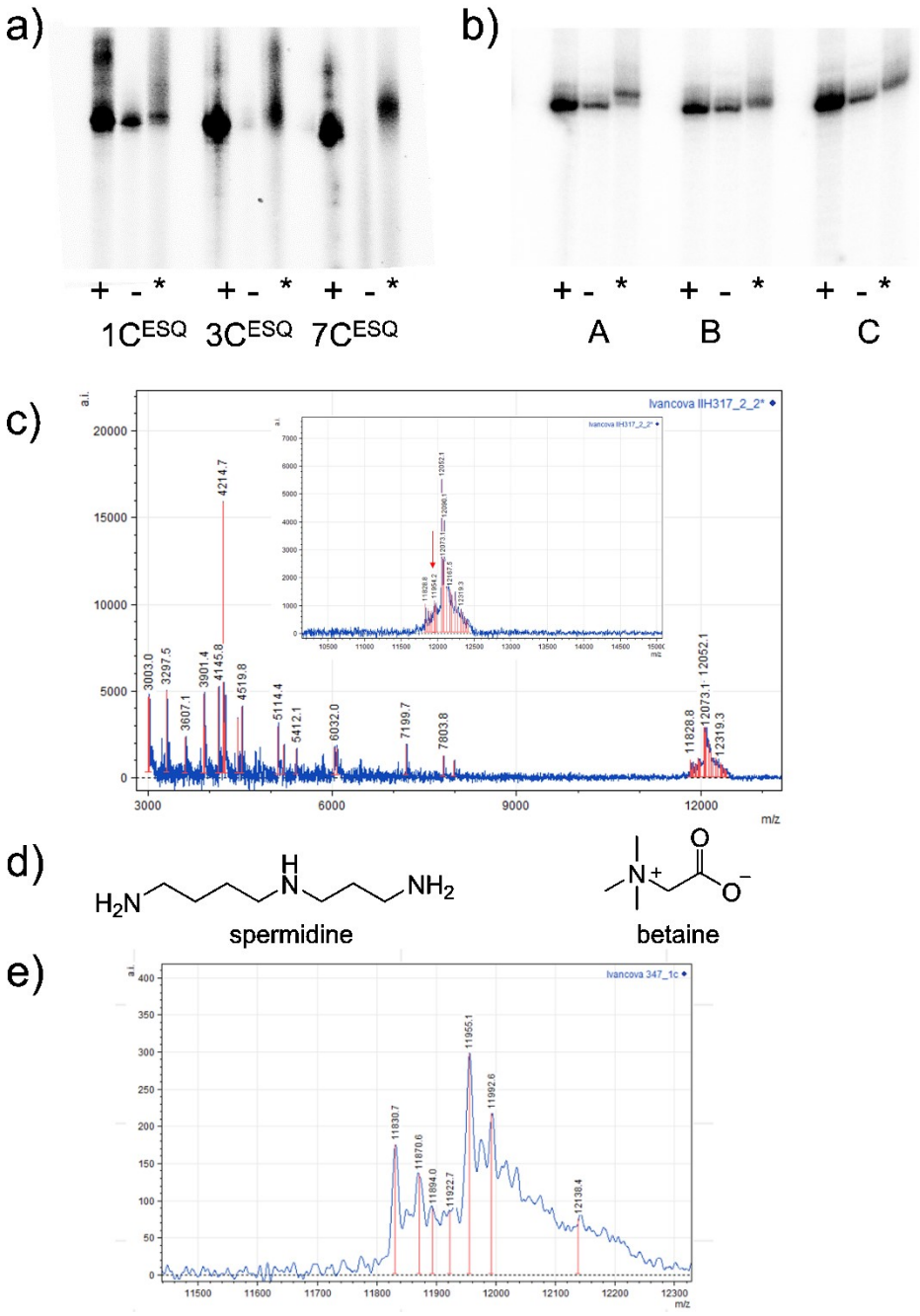


Figure 18. a) Phosphor image of transcripts obtained by *in vitro* transcription in the presence of CTP (+), **CTP^{ESQ}** (*) and in their absence for negative control (-) using the commercial T7RNAp buffer. b) Phosphor image of transcripts obtained by *in vitro* transcription using buffer supplied by manufacturer containing spermidine (A), buffer with betaine (B) and without any amine (C). c) MALDI-TOF MS spectrum of **RNA_1C^{ESQ}** after transcription reaction in commercial T7 RNAP reaction buffer. M (calc.) = 11954 Da, M (found) = 11954.2 Da. The peak at *m/z* = 12052.1 can be assigned to the **RNA_1C^{ESQ}** reacted with spermidine (M = 145.25 Da) from reaction buffer. d) Structures of amines used in transcription buffers. e) MALDI-TOF MS spectrum of **RNA_1C^{ESQ}** obtained after transcription reaction with buffer B. M (calc.) = 11954 Da, M (found) = 11955.1 Da $[M + H]^+$.

After having the right transcription conditions and knowing that the T7 RNAp accepts **C^{ESQ}TP**, I designed additional DNA templates that contain T7 polymerase promoter sequence linked to the sequence of Japanese encephalitis virus (JEV) genome derived RNA templates that represent the 3'-end of JEV genome of the c5'-UTR (the region complementary to the c5'-UTR).[200] These DNA templates would result in 35 and 36mer RNA transcripts with one, three and eight modifications in various positions in the strand (**DNA_C1e^{ESQ}**, **DNA_C1g^{ESQ}**, **DNA_C3f^{ESQ}**, **DNA_C8d^{ESQ}**; **Table 7**).

The RNA templates were synthesized by *in vitro* transcription as described above by using transcription buffer B. The products were resolved by 12.5% PAGE and visualized by autoradiography (**Figure 19-b**). The full-length products were observed in all cases, although **DNA_C3f^{ESQ}** and **DNA_C8d^{ESQ}** templates were transcribed less efficiently most probably due to the nature of the viral RNA sequence. Natural uridines decrease the transcription yield when close to initiation region[201] as well as might the number and position of the modified nucleotides. In the case of the template **DNA_C1e^{ESQ}** the negative control revealed the product of misincorporation. The full-length products were successfully confirmed by MALDI-TOF except of the **RNA_C3f^{ESQ}** (**Table 9**). Fragmentation of **RNA_C1g^{ESQ}** was observed in the MALDI spectra.[202]

Table 9. MALDI-TOF data of modified JEV genome derived RNA templates.

Transcript	M (calc.) [Da]	M (found) [Da]
RNA_C8d^{ESQ}	13150.3	13457.4 $[M + UMP]^+$
RNA_C1e^{ESQ}	11748	11749.6 $[M + H]^+$
RNA_C1g^{ESQ}	12080	11950.5 $[M - U - H_2O]^+$

For the purpose of this project, there was no need to further improve the transcription condition since I was able to obtain decent amounts of purified modified RNA needed for following experiments.

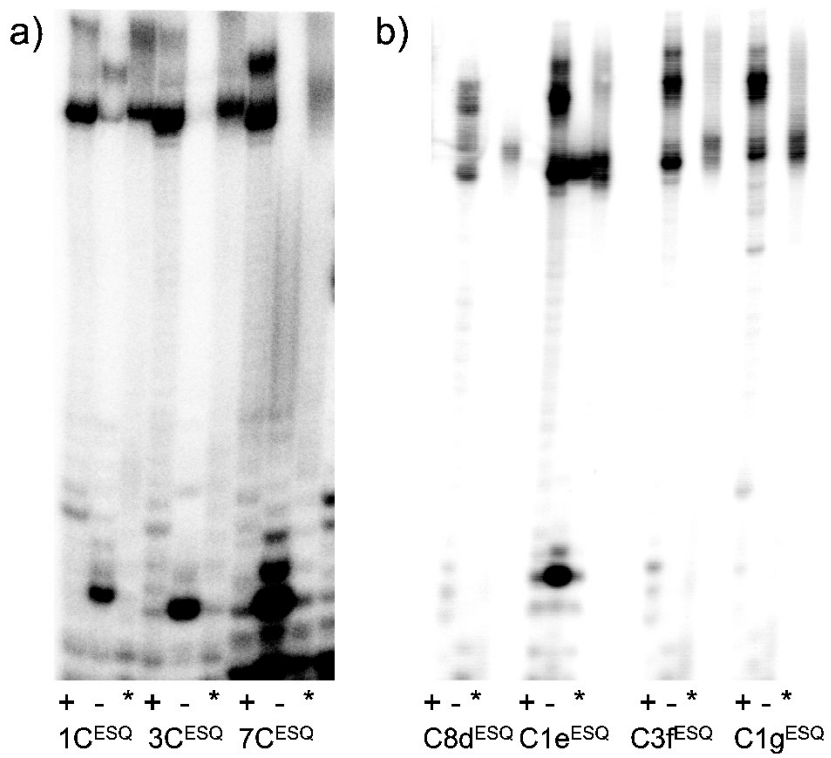
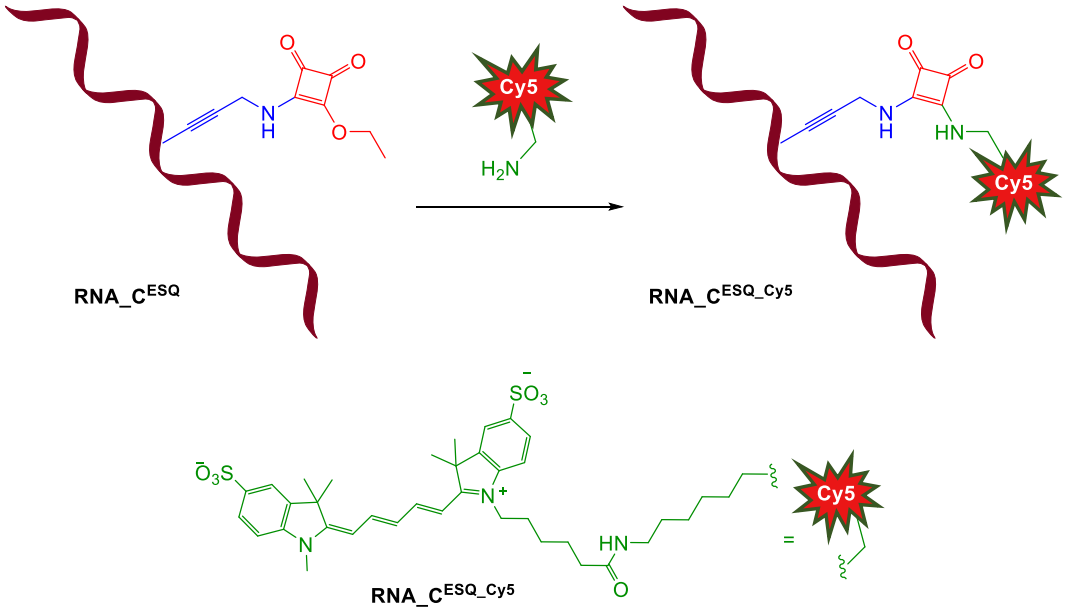


Figure 19. Phosphor image of transcripts obtained by *in vitro* transcription in the presence of CTP (+), **CTP^{ESQ}** (*) and in their absence for negative control (-) using transcription buffer with betaine and templates **DNA_1/3/7C^{ESQ}** (a) or templates designed for JEV 3'-end region of the c5'-UTR (b).

3.2.3 Post-synthetic labelling of ESQ modified RNA

The next step after successful incorporation of **C^{ESQ}TP** into the RNA was to try site-specific RNA labelling. I used the same sulfo-Cy-5-NH_2 dye as for the **DNA_C^{ESQ}** (see chapter 3.1.3; **Scheme 21**).



Scheme 21. General scheme of the cross-linking reaction of squaramate-modified RNA (**RNA_C^{ESQ}**) with sulfo-Cy-5-NH_2.

Non-labelled natural RNA and squaramate modified RNA (**RNA_1C^{ESQ}**) was prepared by *in vitro* transcription followed by DNAse I treatment to digest the DNA templates and purified by spin column method. Unlike DNA, RNA is prone to hydrolysis in alkaline conditions, therefore the RNA was incubated with sulfo-Cy-5-amine in buffer of pH 7.4-7.9 (PBS or TRIS) at 37 °C. After reaction completion, the Cy-5-labelled RNA was purified again to get rid of the excess dye and analysed by fluorescence spectroscopy (**Figure 19-a**) and by fluorescence imaging of the denaturing gel (**Figure 20-b**).

Figure 20-a shows the fluorescence spectra of natural RNA (black line) and **RNA_1C^{ESQ}** (red line) after incubation with Cy-5-amine. Strong fluorescence signal was observed ony for modified RNA suggesting the formation of the cross-linking product **RNA_C^{ESQCy5}**.

The fluorescence imaging of the denaturing gel, shown in **Figure 20-b**, supports the conclusion of fluorescence measurements. Fluorescent band after excitation with 635 nm for Cy-5 is observed only in the lane

where **RNA_C^{ESQCy5}** was formed, while all other lanes are fluorescently silent. SYBR Gold staining visualizes all RNA sequences, however the Cy-5-RNA conjugate was not visible most probably due to the quenching of SYBR Gold by Cy-5 fluorophore.

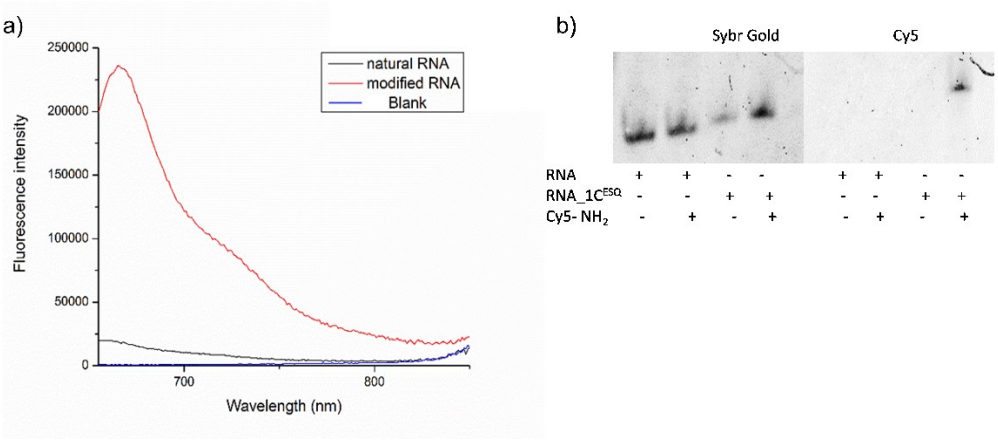
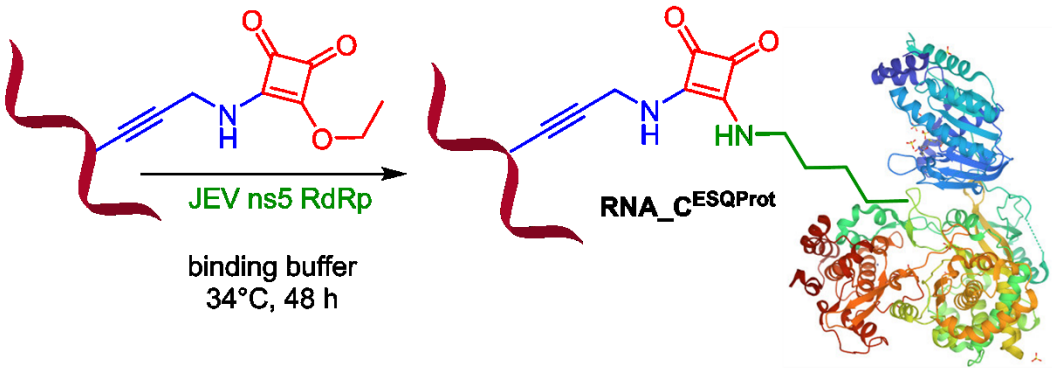


Figure 20. Conjugation of natural and ethoxy squarate modified **RNA_1C^{ESQ}** with sulfo-Cy-5-amine. a) Fluorescence spectra. Conditions: unlabelled natural RNA (Control line in black) or unlabelled **RNA_1C^{ESQ}** (Reaction line in red), sulfo-Cy-5-amine (400 equiv), PBS buffer (pH 7.4), 37 °C, overnight. Fluorescence of the products was measured on spectrofluorometer (water; λ_{ex} = 646 nm, λ_{em} = 662 nm). b) Denaturing PAGE analysis. Conditions: unlabelled natural RNA or unlabelled **RNA_1C^{ESQ}**, sulfo-Cy-5-amine (100 equiv), Tris HCl buffer (40 mM, pH 7.9), 37 °C, overnight. Excited with 473 nm for SYBR Gold and 635 nm for Cy-5.

3.2.4 Conjugation of ESQ modified RNA with proteins

For the bioconjugation experiments, JEV and YFV NS5 RdRp and SARS-CoV-2 RdRp was kind donation from Mgr. Eva Konkoľová, Ph.D. from the group of E. Bouřa at the IOCB AS in Prague.



Scheme 22. General scheme of cross-linking of squaramate modified RNA (**RNA_C^{ESQ}**) with JEV NS5 RdRp protein (PDB structure 4K6M).

For conjugation studies of squaramate modified RNA with RNA binding proteins (RBP) I have selected three RNA dependent RNA polymerases (RdRp; Japanese encephalitis virus NS5, Yellow fever virus NS5 and SARS-CoV-2 RdRp), SARS-CoV-2 nucleoprotein (SC_NP), and HIV reverse transcriptase (HIV-rt) as a biologically relevant target.[203–206]

However, as a first experiment, the T7 RNAp was incubated with **RNA_C3f^{ESQ}** and **RNA_C1g^{ESQ}** to support the results from *in vitro* transcription assays that the modification does not interfere with the enzyme during transcription. The incubation was carried out at the same conditions as the transcription reaction and analysed by the 10% protein SDS PAGE (**Figure 21**). No cross-linking product was observed (**Table 11-12)**.

Here I would like to note that all cross-linking experiments were done with radioactively labelled natural or modified RNA prepared by *in vitro* transcription reaction with T7 RNA polymerase (as described in chapter 3.2.2.1).

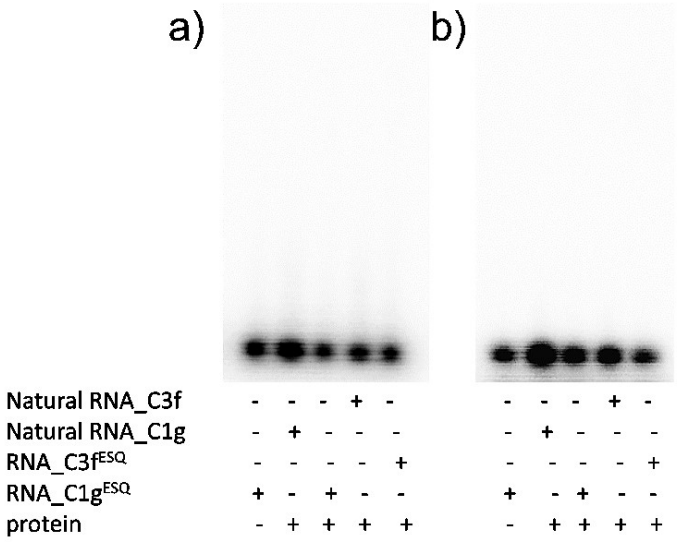


Figure 21. Phosphor image of 10% SDS analysis of control experiment with natural or modified RNA and T7 RNA polymerase. Conditions: RNA (1 μM), transcription buffer B, 37°C, 30 min for EMSA (a) and 2h for SDS (b).

Next, control experiments with RNA nonbinding proteins were performed. For this purpose, I used BSA and Single strand binding protein (SSB). BSA does not bind to RNA but contains many lysines accessible for conjugation. SSB is DNA binding protein with 10-fold lower affinity for ssRNA[207,208] and is useful as a "proof of concept". No binding of BSA to **RNA_C1g^{ESQ}** was observed neither on native PAGE (**Figure 22-a**), nor after denaturation (**Figure 22-b**). On the other hand, there was appearance of a bands with lower mobility for SSB, with higher intensity of the band where modified RNA was incubated with the protein compared to natural RNA (**Figure 22-a**). The denaturing PAGE revealed formation of covalent cross-link (**RNA_C1g^{ESQ_SSB}**; conversion 9%; **Figure 22-b, Table 11**). Moreover, the identity of the product was confirmed by mass spectrometry (**Table 10**; **Figure A7**).

Table 10. MS data of RNA-SSB protein conjugate (for MS spectrum see **Figure A7**).

RNA	M (calc.) [Da]	M (found) [Da]
RNA_C1g^{ESQ_SSB}	30920.52	30943.57 $[M + Na]^+$

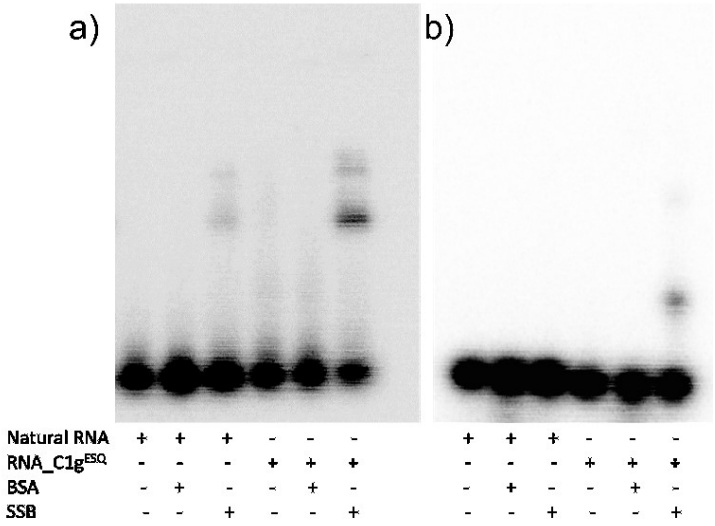


Figure 22. Phosphor image of SDS analysis of control experiment with natural or modified RNA and various RNA nonbinding recombinant proteins (20 equiv. of BSA or SSB to RNA): 10% protein gel. Conditions: binding buffer D, 37°C, 1h for EMSA (a) and 24h for SDS (b).

Viral RNA dependent RNA polymerases form a replication complex along with genomic RNA.[209] For conjugation reactions, genome derived RNA templates with the sequence of Japanese encephalitis virus (JEV) were designed and prepared by *in vitro* transcription (chapter 3.2.2.1).

Kinetic study of the cross-linking reaction of **RNA_C1g^{ESQ}** (**Figure 23-a**) and **RNA_C3f^{ESQ}** (**Figure 23-b**) with JEV NS5 protein shows increasing conversion with the time. After 3 days, conversions reached 30% for modified RNA bearing one modification and 38% for RNA with three ESQ groups. For further experiments I have used more convenient 48h incubation time which gave already sufficient conversions (26% and 37% respectively).

Figure 23-c shows, that 4 equivalents of JEV NS5 protein to 1 equivalent of **RNA_C1g^{ESQ}** is the most efficient ratio for cross-link formation.

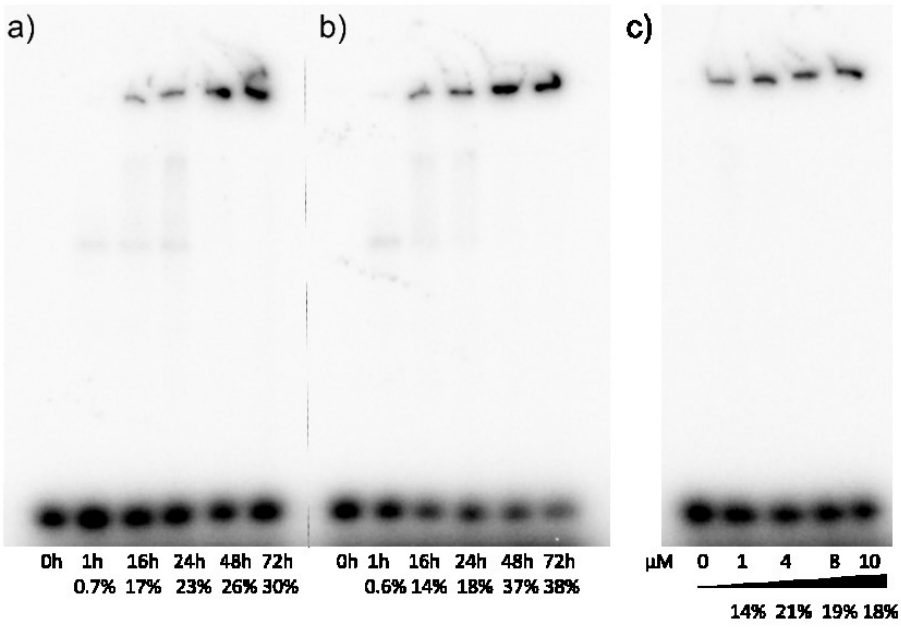


Figure 23. Effect of reaction time on cross-link formation between JEV NS5 RdRp and a) **RNA_C1g^{ESQ}** b) **RNA_C3f^{ESQ}**; phosphor image of 10% SDS gel. Conditions: 4 equiv. of protein to RNA, binding buffer E, 34 °C. c) Effect of the RNA/protein ratio on cross-link formation between **RNA_C1g^{ESQ}** and JEV NS5 RdRp; phosphor image of 10% SDS gel. Conditions: JEV NS5 RdRp (0, 1, 4, 8, 10 µM), **RNA_C1g^{ESQ}** (1 µM), binding buffer E, 34 °C/48 h. All conversions were determined from the gel using ImageJ Quantificator.

The ability of Japanese encephalitis virus and Yellow fewer NS5 proteins to bind to the RNA target was monitored by 7% native PAGE (**Figure 24-a**) confirming that the squaramate modification does not prevent protein binding. The denaturing PAGE reveals formation of covalent cross-links represented by the bands with slower mobility (**Figure 24-b**). The conversion of JEV NS5 conjugate with the **RNA_C1g^{ESQ}** was comparable with that of YFV NSP5 (43% and 40% respectively; **Table 11**). The same trend was observed for the **RNA_C3f^{ESQ}** with three modifications in the sequence (JEV NSP5 41%, YFV NSP5 42%; **Table 12**).

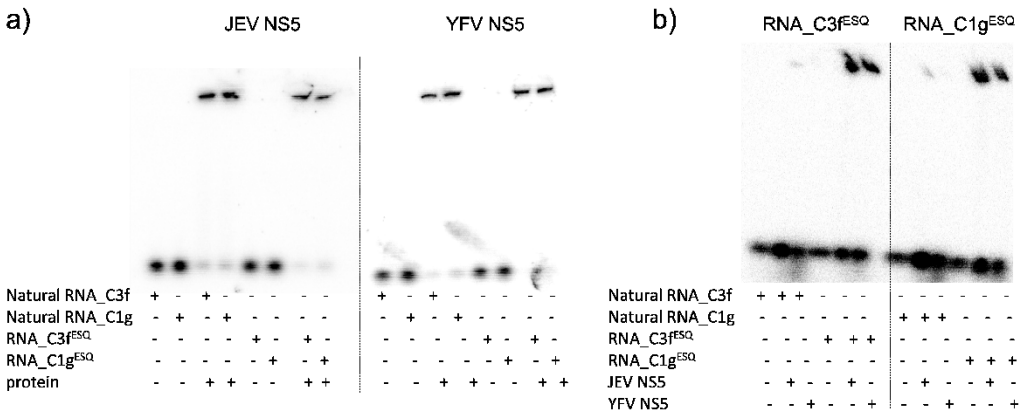


Figure 24. Phosphor image of protein gel from cross-linking reaction between JEV and YFV NS5 RdRp and RNA with 3 or 1 modification. Conditions: 4 equiv. of protein to RNA (0.5 μM), binding buffer E, 34 °C, 1h for EMSA (a) and 48h for SDS (b).

RNA dependent RNA polymerase of SARS-CoV-2 (SC-RdRp) virus is a tetrameric complex of SARS-CoV-2 non-structural proteins nsp7 (one protein; ~ 10 kDa), nsp8 (two proteins; ~ 20 kDa each) and catalytic subunit nsp12 (one protein; ~ 108 kDa).[210,211]

A gel mobility shift assay was performed to detect binding of modified RNA by the RdRp complex. Native PAGE showed only weak binding of the RdRp complex to both native and modified RNA (**Figure 25-a**). After denaturation, only traces of covalent cross-link were detected (**Figure 25-b**). More modifications in the RNA strand did not lead to better binding to the protein. The small shift of the product band suggests that the squaramate modified RNA cross-linked to the one of the smaller proteins of RdRp complex, most probably nsp8 (shift of the conjugate ~ 32 kDa compared to ~ 59 kDa **RNA_C^{ESQ_SC_NP}**; **Figure 25-b**; **Table 11-12**).

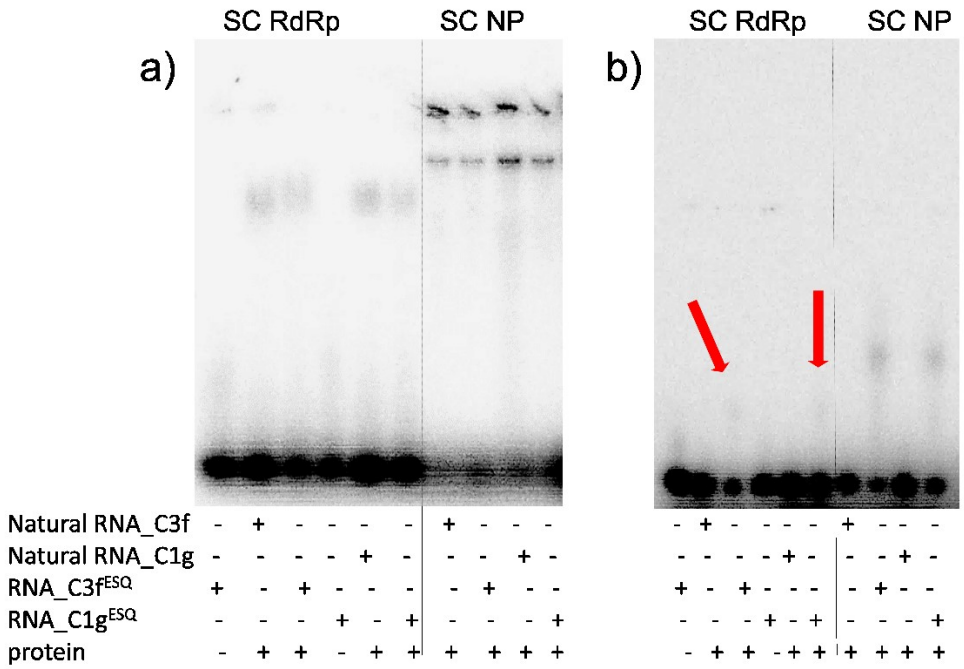


Figure 25. Phosphor image of gel analysis of cross-linking experiment with natural or modified RNA and SARS CoV2 RdRp and SARS CoV2 nucleoprotein (8 equiv. of protein to 1 μM RNA): 10% gel. Conditions: binding buffer D, 30 °C, 1h for EMSA (a) and overnight for SDS (b).

Nucleoproteins are proteins structurally associated with nucleic acids. Viral nucleoprotein (NP) binds to the genomic RNA and forms virion core containing a ribonucleoprotein complex essential for replication and transcription. SARS-CoV-2 NP (SC_NP; 47 kDa) is not only involved in viral assembly, but as well regulates immune response and represses host antiviral response, which makes nucleoproteins interesting target for potential antivirals.[212]

Electrophoretic mobility shift assay (EMSA) of modified RNA and SARS-CoV-2 NP complex showed the formation of the complex of both native and modified RNA and SC_NP. The shift of the RNA - SC_NP complex is slower than expected (~ 200 kDa to expected 59 kDa of monomeric complex) which indicates formation of homooligomers of SARS-CoV-2 NP (**Figure 25-a**).[213] However, only traces of covalent cross-link were detected after denaturation (**RNA_C^{ESQ_SC_NP}**; **Figure 25-b**; **Table 11-12**). More modifications in the RNA strand did not improve covalent binding to the protein.

HIV reverse transcriptase is an asymmetric heterodimeric protein that consists of two subunits, p66 (66 kDa) and p51 (51 kDa). It has two functions, a DNA polymerase activity that copies either DNA or an RNA template, and RNase H activity that cleaves RNA if it is a part of an DNA/RNA duplex with the final aim to convert RNA into linear dsDNA.[214]

The squaramate modified **RNA_C1g^{ESQ}** was incubated with HIV-rt. Native PAGE revealed formation of protein- **RNA_C1g^{ESQ}** complex and after denaturation the bands with slower mobility showed covalent bond formation between HIV-rt and squaramate modified **RNA_C1g^{ESQ}** (15% conversion; **Table 11**).

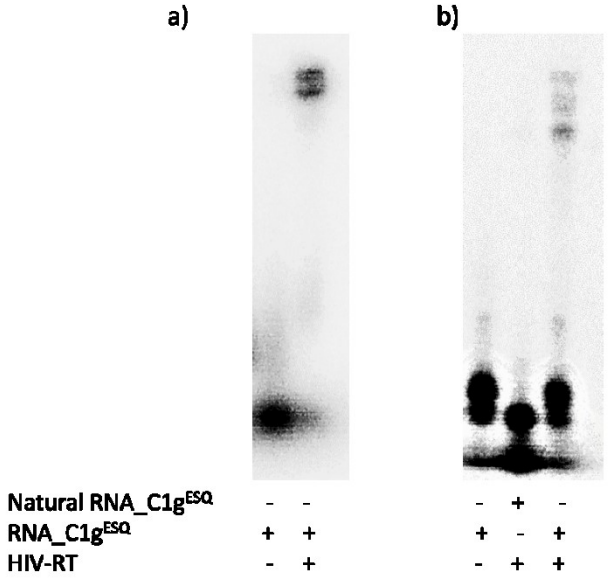


Figure 26. Phosphor image of gel analysis of cross-linking experiment with natural or modified RNA and HIV-rt (2 equiv. of protein to 0.5 μM RNA): 10% gel. Conditions: transcription buffer B 37 °C, incubation 1h for EMSA (a) and overnight for SDS (b).

To sum up, **RNA_C^{ESQ}** did not cross-link with T7 RNA polymerase, which confirms that **C^{ESQ}TP** does not interfere with the polymerase activity during transcription reaction and as well did not react in control experiment with BSA. On the other hand, squaramate modified RNA cross-linked with SSB and all other targeted proteins at different levels (JEV, YFV, SARS-CoV-2 RdRp, SARS-CoV-2 NP and HIV-rt; for conversions see **Table 11 and 12**).

Table 11. RNA-protein cross-link conversions between **RNA_C1g^{ESQ}** and recombinant proteins.

Protein	Molecular weight [kDa]	RNA binding	No of Lys	Conversion %
BSA	69.3	no	60	0%
T7 RNAP	99	no	66	0%
SSB	18.9	weakly	6	9%
HIV rt	117	yes	108	15%
JEV NS5	103.3	yes	58	43%
YFV NS5	103.3	yes	60	40%
SC RdRp	62.2	yes	72	traces
SC NP	45.6	yes	31	traces

Note: Conversions were determined from the gel using ImageJ Quantificator.

Table 12. RNA-protein cross-link conversions between **RNA_C3f^{ESQ}** and recombinant proteins.

Protein	Molecular weight [kDa]	RNA binding	No of Lys	Conversion %
T7 RNAP	99	no	66	0%
JEV NS5	103.3	yes	58	41%
YFV NS5	103.3	yes	60	42%
SC RdRp	62.2	yes	72	traces
SC NP	45.6	yes	31	traces

Note: Conversions were determined from the gel using ImageJ Quantificator.

3.2.5 In vitro RdRp assay with squaramate-modified ribonucleoside triphosphate ($C^{ESQ}TP$)

Viral RNA dependent RNA polymerases are essential for viral replication and therefore are interesting targets for antiviral drug development.[203] *Flaviviridae* RdRp duplicates the ssRNA genome from 3'-end and copies the whole RNA molecule in a primer-independent way (**Figure 27-a**).[215] Full-length NS5 protein has methyltransferase (MTase; involved in RNA cap formation) activity in the N-terminal and RNA dependent RNA polymerase domain at C-terminal of the protein.[216]

Following the cross-linking studies in the previous chapter (chapter 3.2.4) that have showed that the *Flaviviridae* RdRps readily cross-linked with squaramate modified RNA, I decided to investigate how modified triphosphate (**$C^{ESQ}TP$**) will affect the polymerase activity of the JEV recombinant RdRp (JEV NS5) protein *in vitro*.

In this study, I used non-modified templates that represent the 3'-end of JEV genome of the c5'-UTR, that fold into a hairpin with 5'-end FAM labelled overhang as a template for primer extension (**Table 13**).[217]

Table 13. Sequences of RNA templates and RNA products used in *in vitro* RdRp polymerase assay.[a]

	Sequence (5' → 3')	Length
RNA_tHP^{C4}	5′-[6FAM]AUACUAA**G**CCA**AGAAGUUCACACAGAUAAA CUUCU**-3′	35
RNA_HP^{C4}	5′-[6FAM]AUACUAAGCCA**AGAAGUUCACACAGAUAAA CUUCU**UGG**C**UUAGUAU-3′	46
RNA_tSC^{C1}	5′-[6FAM]AACAAAACAU**GCGCGUAGUUUUCUACGC G**-3′	29
RNA_SC^{C1}	5′-[6FAM]AACAAAACAUG**CGCGUAGUUUUCUACGC GC**AUGUUUUGUU-3′	40
RNA_tSC^{C4}	5′-[6FAM]AACAAAA**G**AUC**CGCGUAGUUUUCUACGC G**-3′	29
RNA_SC^{C4}	5′-[6FAM]AACAAAA**G**AUC**CGCGUAGUUUUCUACGC G**GAU**C**UUUUGUU-3′	40

[a] Hairpin part is marked bold, base that is directing position of modified base is marked bold in blue, whereas the modified base in the product is marked bold in red.

I used gel based RdRp polymerase assay conditions similar to those published for various flaviviruses.[218] **Figure 27** shows that the enzyme was able to synthesize RNA by *de novo* initiation with RNA template derived from JEV genome. After 1 hour reaction, the full-length product of natural RNA (**Figure 27-b**, lane 3) and only traces of modified RNA (**Figure 27-b**, lane 4) were observed. After treatment of the reaction mixture with Proteinase K (pK) to release all RNA from JEV NS5 protein, both natural and modified full-length RNAs were observed with additional band of slower mobility in the **RNA_C^{ESQ}** lane (**Figure 27-b,** lanes 5 and 6). The intensity of new band increased after prolonged reaction time (**Figure 27-c,** lane 10). I assumed that this band might be the squaramate modified RNA cross-linked to the short peptide resulting from JEV NS5 protein digestion. The presence of both full-length JEV NS5 RdRp and modified RNA adduct was independently proved by comparison of 10% SDS protein gel visualised by FAM and western blotted, followed by immunodetection of JEV NS5 (**Figure 27-d**). Aliquots from reaction mixtures were loaded to

denaturing SDS gel and stained with PageBlue™ protein stain to confirm full protein digestion (**Figure 27-e**).

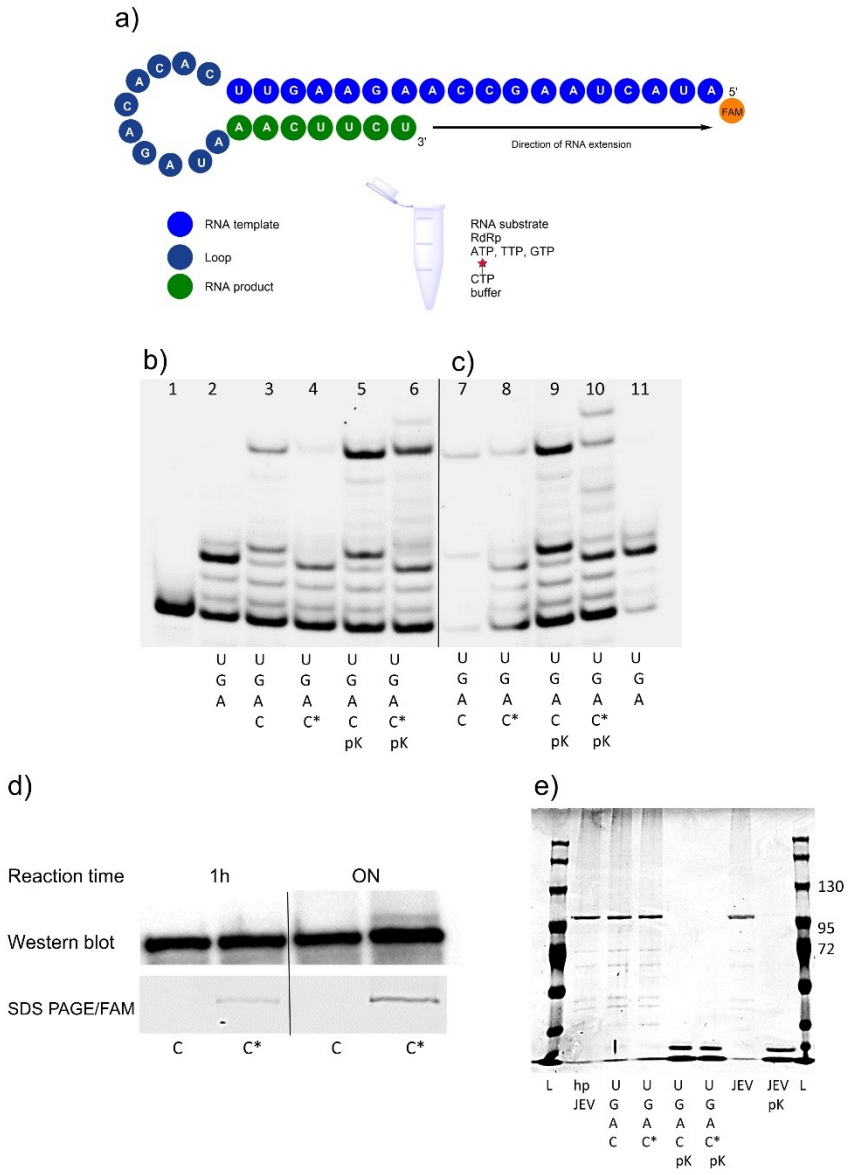


Figure 27. a) Scheme of RdRp polymerase assay. Hairpin RNA substrate with "template" and "product" regions / template-product scaffold. Gel based JEV RdRp polymerase assay using template RNA_tHP^{C4}. b) Reaction time 1 hour c) reaction time overnight. Lane 1: template; lanes 2, 11: negative control without CTP; lanes 3,7: positive control (natural NTPs); lanes 4,8: modified (**C^{ESQ}**, GTP, ATP, UTP); lanes 5,9: positive control followed by Proteinase K treatment (pK); lanes 6,10: modified, followed by Proteinase K (pK) treatment. d) Western blot. Lane C: positive control (natural NTPs); lane C*: modified (**C^{ESQ}**, GTP, ATP, UTP). e) PageBlue™ protein staining of the SDS gel analysis of JEV RdRp assay (hp = template, JEV = protein, pK = Proteinase K, L = ladder).

These results suggest that the JEV RdRp incorporates **C^{ESQ}TP** and further cross-links with produced modified RNA during polymerase reaction. To characterize cross-linked product fully and to determine the RNA-protein binding sites the reaction mixture after polymerase assay was submitted to proteomic analysis, which will be discussed in the following chapter.

The squaramate modified RNA gave only traces of cross-linking adduct with RNA dependent RNA polymerase of SARS-CoV-2 virus (**Figure 25-b**) therefore I was wondering how modified **C^{ESQ}TP** will influence RdRp activity on RNA template-product scaffold (**Scheme 23**). I used two RNA templates to incorporate **C^{ESQ}TP** at the positions + 1 or + 4 respectively (for template sequences see **Table 13**). **C^{ESQ}TP** was readily incorporated into RNA and the reactions gave full-length products for both positive control and modified RNA (**Figure 28-a**, lanes 8,9 and **28-b**, lanes 5,6) without any decreased intensity of the product bands. In the case of template where C is in the position + 4 (template: RNA_tSC^{C4}; **Table 13**; **Figure 28-b**, lane 4) some level of misincorporation was detected in negative control but once natural or modified C was present, polymerase proceeded correctly. The full-length products of the polymerase reaction were confirmed by MALDI (**Table 14**; **Figure A8** and **A9**).

These results show that modification does not interfere with the enzyme during polymerase assay, which is consistent with the cross-linking experiment with the SC RdRp during incubation with squaramate, modified RNA (chapter 3.2.4; **Figure 25-b**).

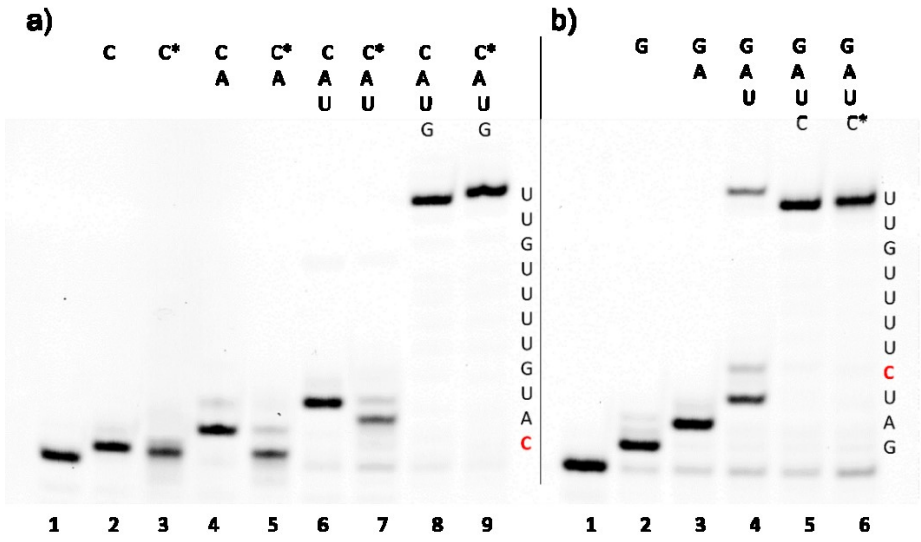


Figure 28. RNA extension in SARS-CoV-2 RdRp polymerase assay a) Template RNA_tSC^{C1} (lane 1). Lanes 2-7: control reactions in the presence of indicated NTPs; lane 8: positive control (natural NTPs); lane 9: modified (**C**^{ESQ}, GTP, ATP, UTP). b) Template RNA_tSC^{C4} (lane 1). Lanes 2-4: control reactions in the presence of indicated NTPs; lane 5: positive control (natural NTPs); lane 6: modified (**C**^{ESQ}, GTP, ATP, UTP).

Table 14. MALDI-TOF data of RNA produced during SARS-CoV-2 RdRp polymerase reaction.

Transcript	M (calc.) [Da]	M (found) [Da]
RNA_C$^{ESQ\_SC1}$	13451.1	13423.3 [M – CH_2CH_3 - H]
RNA_C$^{ESQ\_SC4}$	13451.1	13421.7 [M – CH_2CH_3]

3.2.6 Proteomic analysis of ESQ modified RNA-JEV conjugate

Finally, I attempted to identify lysine residues participating in JEV NS5 RdRp-**RNA_C^{ESQ}** cross-linking during polymerase assay. I used in solution digestion technique. First, the protein was digested to peptides by the tryptic digestion followed by the digestion of RNA to nucleosides by RNase A/T1 mix. Tryptic peptides were analysed by nanoLC-ESI$^+$-MS/MS to determine the AA residues cross-linked to **C^{ESQ}**. The analysis of two independent samples revealed three lysine residues K269 (sample 1; **Figure 29**; **Table 15**), K462 and/or K463 (sample 2; **Figure 30**; **Table 15**) participating in protein cross-linking to **C^{ESQ}** in newly synthesized RNA. There were identified ions for both neighbouring lysine residues in the spectrum of sample 2, and it cannot be determined with certainty whether it is on K462 and/or K463. For sequence coverage of analysed protein samples see **Figures A10-11**.

Examination of the crystal structure (PDB 4K6M) reveals the location of the K269 in 10-residue linker region (residues 266-275) which connects the MTase to its C-terminus (RdRp; **Figure 31**). The sequence of this linker varies highly among genus flavivirus. The lysines K462 and K463 are in ring finger domain of RdRp (residues 453-479; **Figure 31**) that forms the roof of the NTP entry chanel at the - 5 position from the NTP binding motif F (Phe 467). These AA are highly conserved in flaviviruses.[219]

Table 15. NanoLC-ESI$^+$-MS/MS data of tryptic peptides sequences bearing the lysine residue targeted by modification.

Peptides sequence	$[M + H]^{3+}$	$[M + H]^+$	$[M + H]^+$
NH2-AVG**K***GEVHSNQEK-COOH	586.26	1756.79	1756.70
NH2-E**K***KPGEFGK-COOH	465.22	1393.64	1393.84
NH2-EK**K***PGEFGK-COOH	465.22	1393.64	1393.84

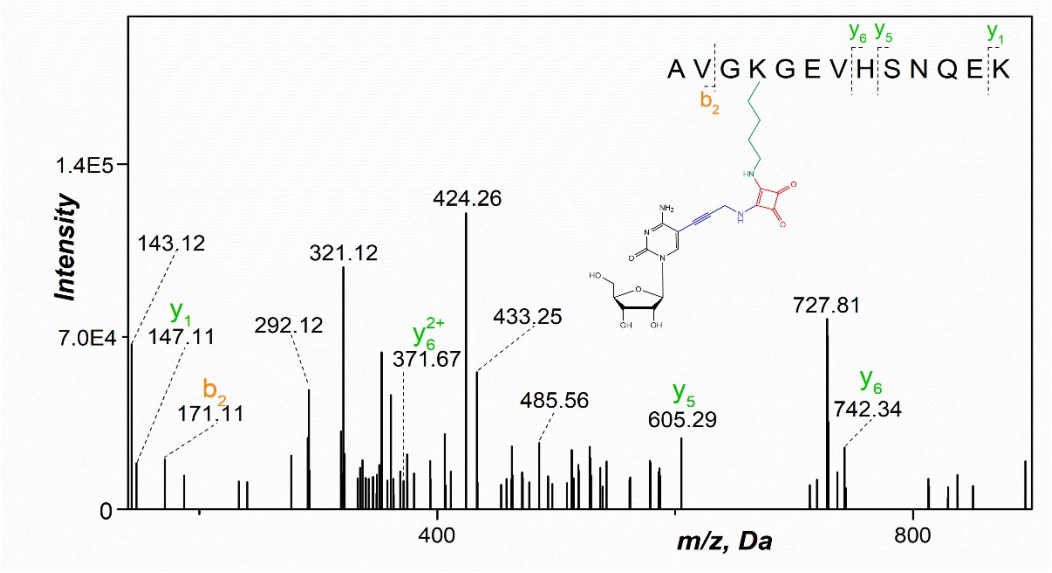


Figure 29. NanoLC-ESI^{+}-MS/MS spectrum of **RNA_C^{ESQ_JEV}** identification of K269 (m/z acquired: 586.26 Da).

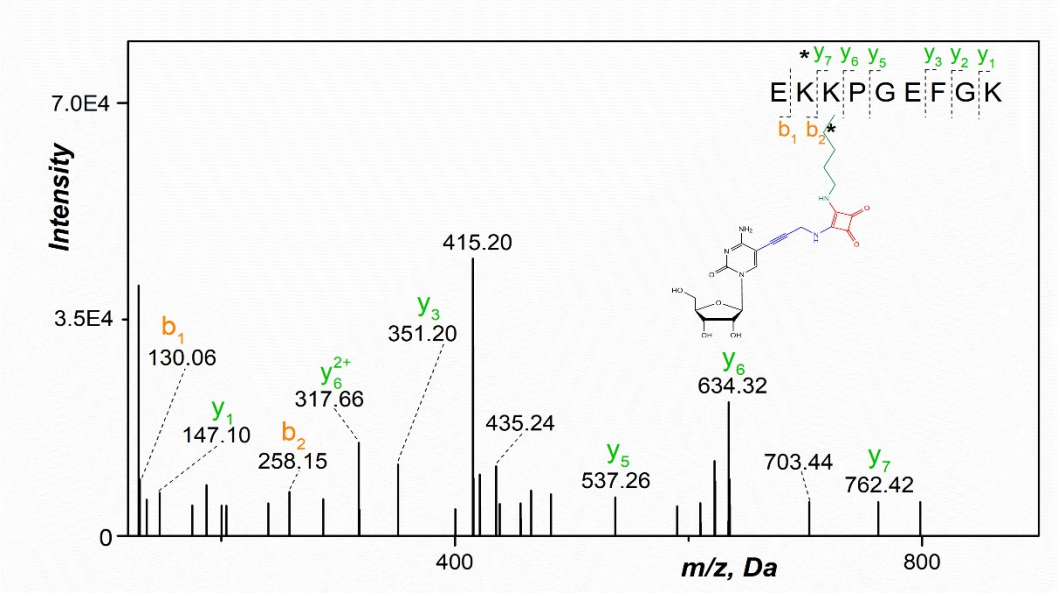


Figure 30. NanoLC-ESI^{+}-MS/MS spectrum of **RNA_C^{ESQ_JEV}** identification of K462/463 (m/z acquired: 465.22 Da).

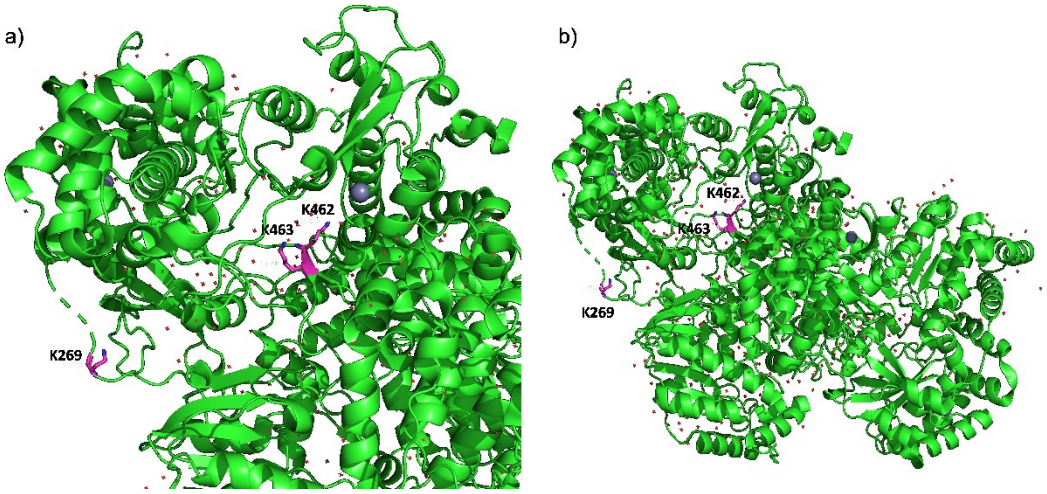


Figure 31. Crystal structure of JEV NS5 RdRp (PDB structure 4K6M) showing position of lysine residues participating in RNA-protein cross-link formation. a) Detail view b) crystal structure of full length JEV NS5. Identified cross-linked lysines are in magenta.

4. Conclusion

In conclusion, I developed reactive DNA and RNA probes, which reacted selectively with lysine without use of any external reagents.

In the first part of the thesis, I prepared squaramate modified dNTP (**dC^{ESQ}TP**) and showed that it was very good substrate for KOD XL DNA polymerase giving the full-length modified DNA in PEX and PCR reactions. The ESQ-modification did not interfere with DNA polymerase. The squaramate modified DNA readily reacted with amino-linked sulfo Cy-5 dye and formed fluorescently labelled DNA. Model reactions with lysine and lysine containing peptides required large excess for successful cross-linking. On the contrary, the squaramate modified DNA cross-linked with histones (lysine containing DNA-binding proteins) under physiological conditions in almost equimolar ratio. The fact, that modified DNA did not cross-link with BSA or GSTp53-CD (lysine containing DNA-nonbinding proteins) nor with DNA polymerase during PEX and PCR, shows that proximity effect is important for effective cross-linking in the absence of large excess of the peptide or protein.

In the second part, squaramate modified ribonucleotide **C^{ESQ}TP** was synthesized and incorporated into the RNA by *in vitro* transcription. I found that the spermidine (primary amine) present in the commercial reaction buffer interfered with the modification during transcription reaction. After optimization of reaction conditions by exchanging the primary amine for quaternary betaine, non-altered full-length RNA transcripts were synthesized and used in post-synthetic labelling with sulfo Cy-5-amine. Next, squaramate modified RNA was incubated with various lysine containing RNA-binding proteins (i.e., JEV and YFV NS5, SARS-CoV-2 RdRp, SARS-CoV-2 nucleoprotein and HIV-rt) and RNA-nonbinding BSA and weakly binding SSB as a control. The cross-linking with flaviviral RdRp polymerases (JEV and YFV NS5) yielded cross-linking products in ~ 40%. On the other hand, reactions with SARS-CoV-2 RdRp and SARS-CoV-2 nucleoprotein gave only traces of cross-linking product. Modified RNA did not crosslink with BSA or T7 RNAp and gave only 9% of SSB conjugate. Incubation of squaramate modified RNA with HIV-rt proceeded with ca. 15% conversion. Finally, I

investigated how modified triphosphate (**C^{ESQ}TP**) affects the polymerase activity of RNA dependent RNA polymerases *in vitro*. Squaramate modification did not interfered with SARS-CoV-2 RdRp and gave full length RNA. On the other hand, experimental results show that newly synthesized RNA with incorporated **C^{ESQ}TP** cross-links with JEV RdRp during polymerase reaction. Proteomic analysis of the reaction mixtures of JEV RdRp RNA extension identified K269, K462, K463 lysine residues modified by **C^{ESQ}** from modified RNA.

The squaramate modification transforms to stable amide under physiological conditions without any external reagent. This property of squaramate modification shows good potential in post-synthetic labelling of nucleic acids, bioconjugation of NA with various biomolecules or in developing irreversible inhibitors of NA-binding enzymes containing free lysine residues (e.g., viral polymerases). Further studies on this topic are under way in Hocek's lab.

5. Experimental part

5.1 General remarks

5.1.1 General remarks for the synthetic part

NMR spectra were recorded on a Bruker Avance-IIIHD 500 (500.0 MHz for ^{1}H, 202.4 MHz for ^{31}P, 125.7 MHz for ^{13}C, and 376.5 MHz for ^{19}F) and Bruker Avance-IIIHD 400 (400.0 MHz for ^{1}H) spectrometer. ^{1}H and ^{13}C resonances were assigned using H, H-COSY, H,C-HSQC and H,C-HMBC experiments. The samples were measured in D_2O, CD_3OD or DMSO-*d*$_6$. Chemical shifts (δ scale, in ppm) were referenced as follows: D_2O (referenced to dioxane as internal standard: 3.75 ppm for ^{1}H NMR and 69.30 ppm ^{13}C NMR); CD_3OD (referenced to solvent signal: 3.31 ppm for ^{1}H NMR and 49.00 ppm for ^{13}C NMR); DMSO-*d*$_6$ (referenced to solvent signal: 2.50 ppm for ^{1}H NMR and 39.70 ppm for ^{13}C NMR). ^{31}P chemical shifts were referenced to H_3PO_4 as an external standard (0 ppm). Coupling constants (*J*) are given in Hz. High resolution mass spectra were measured on LTQ Orbitrap XL (Thermo Fisher Scientific) using ESI ionization technique. All materials were purchased from commercial suppliers and used without further purification unless otherwise stated. $POCl_3$ and $PO(OMe)_3$ were distilled prior to use. The water used in synthetic part was of HPLC quality. Purification of nucleoside triphosphates was performed using HPLC (Waters) on a C18 reversed phase column (Phenomenex, Luna C18 (2) 100 Å).

5.1.2 General remarks for the biochemical part

All gels were analysed by fluorescence imaging using Typhoon FLA 9500 (GE Healthcare). PCR gels were visualized using transilluminator equipped with GBox iChemi-XRQ Bio imaging system (Syngene, Life Technologies). Mass spectra of oligonucleotides were measured on UltrafleXtreme MALDI-TOF/TOF (Bruker) mass spectrometer with 1 kHz smartbeam II laser. UV-Vis spectra were measured at room temperature on NanoDrop1000 (ThermoScientific). Fluorescence was measured on a Fluoromax 4 spectrofluorimeter (HORIBA Scientific). Samples were concentrated on CentriVap Vacuum Concentrator system (Labconco). Synthetic DNA oligonucleotides (primers, templates, and biotinylated templates) were purchased from Generi Biotech (Czech Republic) or Eurofins Genomics (Germany). Synthetic RNA oligonucleotides were purchased from Sigma Aldrich. Sulfo-Cyanine5 amine from Lumiprobe. [α-^{32}P]-GTP from M.G.P. spol. s r.o. Natural nucleoside triphosphates (dATP, dGTP, dTTP, dCTP, ATP, GTP, TTP, CTP), BSA, T7 RNA polymerase, Ribolock RNase inhibitor and histone human recombinant proteins (H2A, H2B, H3.1, H4) were purchased from New England Biolabs. SSB protein frm Sigma Aldrich. DNase I was purchased from Thermoscientific. Sars-Cov-2 nucleoprotein from ProteoGenix and HIV-rt from Merck Millipore. KOD XL DNA polymerase and corresponding polymerase reaction buffer from Merck Millipore, streptavidine magnetic particles from Roche, QIAquick® Nucleotide Removal Kit from Qiagen, NucAway spin columns from Ambion and Bio-Spin 6 from Biorad. Japanese encephalitis virus NS5 polyclonal antibody and goat anti-rabbit IgG (H + L) secondary antibody, HRP were purchased from Invitrogen. Immobilon-P transfer membrane from Millipore. Western blott detection reagent SuperSignal West Femto Chemiluminescent Substrate from Thermo Scientific. PageBlue^{TM} protein staining solution from ThermoFisher Scientific. Milli-Q water was used for all experiments. For RNA experiments DEPC water was used.

For the bioconjugation experiments, GSTp53CD was kind donation from Dr. Marie Brázdová from the group of Doc. Miroslav Fojta at the Institute of Biophysics AS CR in Brno, JEV and YFV NS5 RdRp and SARS-CoV-2 RdRp was kind donation from Mgr. Eva Konkoľová, Ph.D. from the group of E. Bouřa at the IOCB AS in Prague. Peptides were purchased at IOCB Prague.

Buffer composition (10×):

Buffer A: 400 mM Tris-HCl (pH 7.9), 60 mM $MgCl_2$, 100 mM DTT, 100 mM NaCl, 20mM spermidine.

Buffer B: 400 mM Tris-HCl (pH 7.9), 60 mM $MgCl_2$, 100 mM DTT, 100 mM NaCl, 20mM betain.

Buffer C: 400 mM Tris-HCl (pH 7.9), 60 mM $MgCl_2$, 100 mM DTT, 100 mM NaCl.

Buffer D: 100 mM Tris (pH. 8), 20 mM $MgCl_2$, 100 mM KCl, 12 mM βME.

Buffer E: 50 mM Tris-HCl (pH. 7.4), 100 mM DTT, 5% Triton X-100, 10% glycerol.

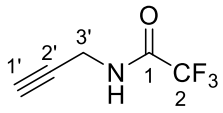
PAGE stop solution used after PEX reactions contains: 80% [v/v] formamide, 20 mM EDTA, 0.0250% [w/v] bromophenol blue and 0.0250% [w/v] xylene cyanol.

VPS loading buffer used for protein SDS gels contains: 0.0500 M TRIS (pH 6.8), 17% glycerol, 16 mM mercaptoethanol (βME), 3.5% SDS, bromophenol blue.

5.2 Enzymatic synthesis of reactive DNA probes for cross-linking with proteins

5.2.1 Synthesis and characterization of ESQ modified 2'-deoxycytidine and 2'-deoxycytidine mono- and triphosphate and 2'-deoxycytidine monophosphate adduct with Ac-Lys-OH and lysine containing tripeptide (Ac-Ala-Lys-Ala-NH_2)

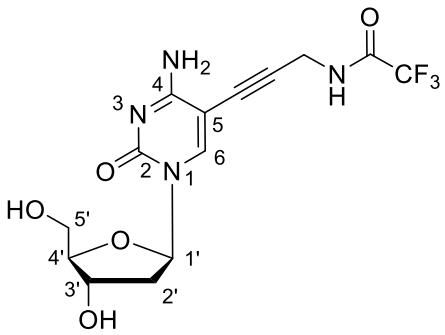
2, 2, 2-Trifluoro-*N*-(prop-2-ynyl)acetamide (14)[192]



Propargylamine (1.50 g, 27.2 mmol) and ethyl trifluoroacetate (5.03 g, 35.4 mmol) were dissolved in methanol (50 mL) and stirred for 24 h in the dark at room temperature. After removal of the solvent under reduced pressure saturated $NaHCO_3$ solution (56 mL) was added to the residue and the aqueous phase was extracted with DCM (4 × 70 mL). The combined organic layers were dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure. Desired product was obtained as yellow oil in 95% yield (3.89 g) and used without further purification.

All spectral data are consistent with the published literature.

5-[3-(Trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine[193,194]

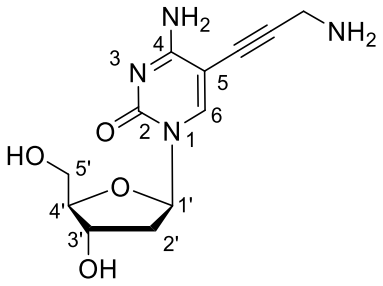


5-Iodo-2'-deoxycytidine (**15**; 1.11 g, 3.16 mmol), *N*-propargyl trifluoroacetamide (**14**, 1.44 g, 9.50 mmol), CuI (0.120 g, 0.630 mmol), $Pd(PPh_3)_4$ (0.365 g, 0.316 mmol) and Amberlite Ira-67 resin (2,81 g) were placed in a 50 mL round

bottom flask and dissolved in anhydrous DMF (17 mL) under argon atmosphere. The reaction mixture was stirred in the dark at r.t. for 24 hours then filtered over bed of SiO_2 topped with celite. The solvents were removed under reduced pressure and the title product was obtained by column chromatography using DCM/methanol (7:1) as a mobile phase in 94% yield (1.12 g) as an orange oil.

All spectral data are consistent with the published literature.

5-(3-Amino-1-propynyl)-2'-deoxycytidine (16)



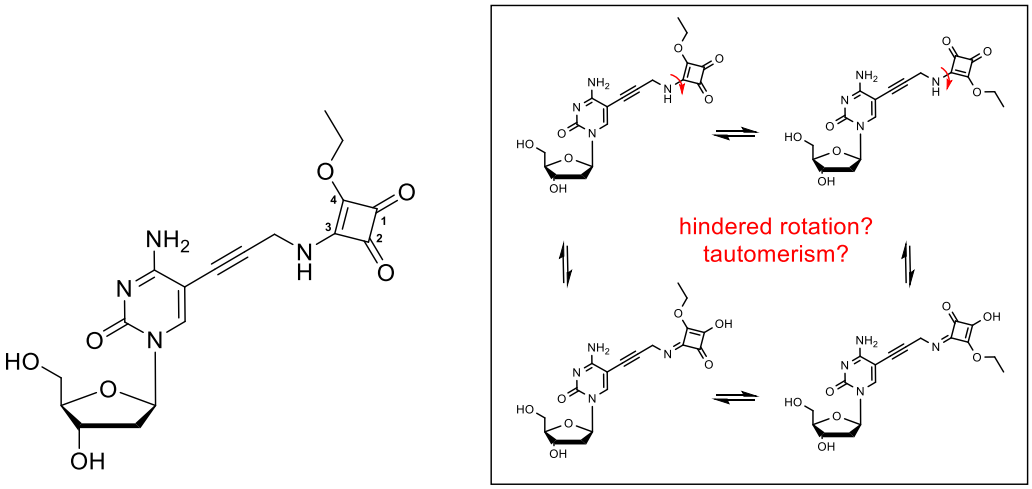
5-[3-(Trifluoroacetamido)-prop-1-ynyl]-2'-deoxycytidine (1.28 g, 3.40 mmol) was dissolved in 17 mL of HPLC water in the pressure tube. After complete dissolution, aqueous ammonium hydroxide (92 mL) was added, and the tube was closed properly. The reaction was stirred 24 hours at r.t. and then concentrated down. The crude product was redissolved in HPLC H_2O (26 mL) and DOWEX 50 × 8 resin (7 g) was added. The mixture was stirred for 30 min and filtered over a bed of DOWEX 50 × 8 resin (7 g) which was then washed with HPLC H_2O, and the product was eluted off the resin with HPLC H_2O/conc. NH_4OH (4:1, 800 mL). Final product was obtained as orange-brown oil (0.627 g, 66%) after removal of the solvents under reduced pressure.

^{1}H NMR (500.0 MHz, CD_3OD): 2.12 (ddd, 1H, J_{gem} = 13.4, $J_{2'b,1'}$ = 6.7, $J_{2',3'}$ = 6.2, H-2′b); 2.38 (ddd, 1H, J_{gem} = 13.4, $J_{2'a,1'}$ = 6.1, $J_{2'a,3'}$ = 4.0, H-2′a); 3.61 (s, 2H, CH_2N); 3.73 (dd, 1H, J_{gem} = 12.1, $J_{5'b,4'}$ = 3.6, H-5′b); 3.81 (dd, 1H, J_{gem} = 12.1, $J_{5'a,4'}$ = 3.1, H-5′a); 3.95 (ddd, 1H, $J_{4',3'}$ = 4.0, $J_{4',5'}$ = 3.6, 3.1, H-4′); 4.36 (dt, 1H, $J_{3',2'}$ = 6.2, 4.0, $J_{3',4'}$ = 4.0, H-3′); 6.21 (dd, 1H, $J_{1',2'}$ = 6.7, 6.1, H-1′); 8.30 (s, 1H, H-6).

^{13}C NMR (125.7 MHz, CD_3OD): 32.06 (CH_2N); 42.38 (CH_2-2′); 62.49 (CH_2-5′); 71.78 (CH-3′); 74.48 (cyt-**C**≡C-CH_2); 87.90 (CH-1′); 89.05 (CH-4′); 92.70 (C-5); 95.95 (cyt-C≡**C**-CH_2); 145.46 (CH-6); 156.77 (C-2); 166.49 (C-4).

HR/MS (ESI^-) for $C_{12}H_{15}O_4N_4$: $[M - H]^-$ calculated 279.11, found 279.11.

3-[5- (3-Amino-1-propynyl) -2'-deoxycytidine]-4- ethoxycyclobut-3-ene-1,2-dione (dC^{ESQ})



16 (0.200 g, 0.713 mmol) was well suspended in 4 ml of ethanol. Diethyl squarate (0.211 mL, 1.43 mmol) was added dropwise during 30 min to the stirred solution and the reaction mixture was stirred another 1 hour at room temperature. Column chromatography using DCM/methanol (8:2) as a mobile phase gave the desired product as white powder in 80% yield (0.231g).

^{1}H NMR (500.0 MHz, DMSO-d_6, T = 100 °C): 1.40 (t, 3H, J_{vic} = 7.0, C**H**$_3$$CH_2O$); 2.00 (dt, 1H, J_{gem} = 13.4, $J_{2'b,1'}$ = 7.0, $J_{2',3'}$ = 6.3, H-2′b); 2.21 (ddd, 1H, J_{gem} = 13.4, $J_{2'a,1'}$ = 6.1, $J_{2'a,3'}$ = 3.6, H-2′a); 3.57 (dd, 1H, J_{gem} = 11.8, $J_{5'b,4'}$ = 4.0, H-5′b); 3.63 (dd, 1H, J_{gem} = 11.8, $J_{5'a,4'}$ = 3.7, H-5′a); 3.82 (ddd, 1H, $J_{4',5'}$ = 4.0, 3.7, $J_{4',3'}$ = 3.6, H-4′); 4.23 (dt, 1H, $J_{3',2'}$ = 6.3, 3.6, $J_{3',4'}$ = 3.6, H-3′); 4.48 (bs, 2H, CH_2N); 4.67 (q, 2H, J_{vic} = 7.0, CH_3C**H**$_2$O); 6.10 (dd, 1H, $J_{1',2'}$ = 7.0, 6.1, H-1′); 6.96 (bs, 2H, NH_2); 8.10 (s, 1H, H-6); 8.74 (bs, 1H, NH).

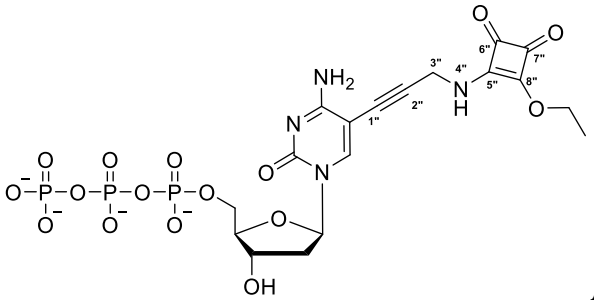
^{13}C NMR (125.7 MHz, DMSO-d_6, T = 100 °C): 15.17 (**C**H_3CH_2O); 33.84 (CH_2N); 40.66 (CH_2-2′); 61.00 (CH_2-5′); 68.75 (CH_3**C**H_2O); 70.00 (CH-3′); 76.28 (cyt-**C**≡C-CH_2); 85.50 (CH-1′); 87.42 (CH-4′); 98.66 (C-5); 90.13 (cyt-C≡**C**-CH_2);

144.22 (CH-6); 153.07 (C-2); 164.02 (C-4); 172.04 (C-4-cyclobut); 177.25 (C-3-cyclobut); 182.42, 188.75 (C-1,2-cyclobut).

HR/MS (ESI$^+$) for $C_{18}H_{21}O_7N_4$: $[M + H]^+$ calculated 405.14, found 405.14.

Synthesis of modified 2'-deoxycytidine triphosphate (dC^{ESQ}TP)

3-{[5- (3-Amino-1-propynyl) -2'-deoxycytidine]-5'-*O*-triphosphate}-4-ethoxycyclobut-3-ene-1,2-dione (dC^{ESQ}TP)



A mixture of two isomers A:B ~ 3:2

Ethoxy squarate modified nucleoside (**dC^{ESQ}**, 0.0500 g, 0.124 mmol) was dried at 80 °C for 2 hours in vacuo. After cooling on ice, $PO(OMe)_3$ (0.400 mL) and $POCl_3$ (14 µL) were added under argon atmosphere. The reaction mixture was stirred for 1 hour at 0 °C. In a separate flask, the mixture of $(NHBu_3)H_2P_2O_7$ (0,370 g) and Bu_3N (90 µL) in dry DMF (1 mL) was prepared under argon atmosphere cooled to 0 °C and then added by the syringe to the reaction mixture. The mixture was stirred at 0 °C 45 min. The reaction was stopped by addition of TEAB (2 M, 0.400 mL) and water (0.600 mL). The product was purified by C18 reversed-phase HPLC using water/methanol (5 to 50%) containing 0.100 M TEAB buffer as eluent. Several co-distillations with water and conversion to sodium salt (Dowex 50W × 8 in Na^+ cycle) followed by freeze-drying from water gave the product **dC^{ESQ}TP** as white powder in 7% yield (0.00600 g).

^{1}H NMR (500.0 MHz, D_2O, ref (dioxane) = 3.75 ppm): 1.40 – 1.46 (m, 6H, C**H$_3$**CH_2O-A,B); 2.30 (dt, 2H, J_{gem} = 13.9, $J_{2'b,1'} = J_{2'b,3'}$ = 6.4, H-2′b-A,B); 2.43 (ddd, 2H, J_{gem} = 13.9, $J_{2'a,1'}$ = 6.4, $J_{2'a,3'}$ = 4.5, H-2′a-A,B); 4.17 – 4.27 (m, 6H, H-4′,5′-A,B); 4.55 (bs, 2H, H-3″-A); 4.61 (dt, 1H, $J_{3',2'}$ = 6.4, 4.5, $J_{3',4'}$ = 4.5, H-3′);

4.65 (bs, 2H, H-3″-B); 4.69 – 4.80 (m, 4H, $CH_3C\mathbf{H_2}O$-A,B); 6.23 (t, 2H, $J_{1',2'}$ = 6.4, H-1′-A,B); 8.22 (s, 2H, H-6-A,B).

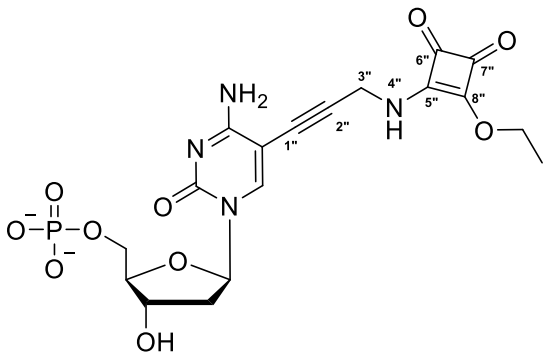
^{13}C NMR (125.7 MHz, D_2O, ref(dioxane) = 69.3 ppm): 17.78, 17.80 ($\mathbf{C}H_3CH_2O$-A,B); 37.09 (CH_2-3″-B); 37.36 (CH_2-3″-A); 42.17 (CH_2-2′-A,B); 67.66 (d, $J_{C,P}$ = 4.7, CH_2-5′-A,B); 72.62 (CH-3′-A,B); 73.55 ($CH_3\mathbf{C}H_2O$-A,B); 77.96 (cyt-**C**≡C-CH_2-B); 78.03 (cyt-**C**≡C-CH_2-A); 88.34 (d, $J_{C,P}$ = 9.0, CH-4′-A,B); 89.05 (CH-1′-A,B); 93.19 (cyt-C≡**C**-CH_2-A); 93.62 (cyt-C≡**C**-CH_2-B); 94.55 (C-5-A,B); 148.09 (CH-6-A,B); 158.81 (C-2-A,B); 167.79 (C-4-A,B); 175.75 (C-5″-B); 175.89 (C-5″-A); 180.11 (C-8″-A); 180.76 (C-8″-B); 186.47, 186.56, 191.41, 191.80 (C-6″,7″-A,B).

$^{31}P\{^1H\}$ NMR (202.4 MHz, D_2O): -21.62 (br, 2P, P_β); -10.58 (d, 2P, J = 19.4, P_α); –6.70 (br, 2P, P_γ).

HR/MS (ESI^-) for $C_{18}H_{22}O_{16}N_4P_3$: $[M - H]^-$ calculated 643.02, found 643.02.

Synthesis of modified 2'-deoxycytidine monophosphate (dC^{ESQ}MP) and its conjugates with Ac-Lys-OH and lysine containing tripeptide (Ac-Ala-Lys-Ala-NH$_2$) (dC^{ESQLys}MP, dC^{ESQ3pept}MP)

3-{[5- (3-Amino-1-propynyl) -2'-deoxycytidine]-5'-*O*-monophosphate}-4-ethoxycyclobut-3-ene-1,2-dione (dC^{ESQ}MP)



A mixture of two isomers A:B ~ 3:2

Ethoxy squarate modified nucleoside (**dC^{ESQ}**, 0.0500 g, 0.124 mmol) was dried at 80 °C for 2 hours in vacuo. After cooling on ice, $PO(OMe)_3$ (0.400 mL) and $POCl_3$ (14 μL) were added under argon atmosphere. The reaction mixture was stirred 2 hours at 0 °C. The reaction was stopped by addition of TEAB (2 M, 0.400 mL) and water (0.600 mL). The product was purified by C18 reversed phase HPLC using water/methanol (5 to 100%) containing 0.100 M TEAB buffer as eluent. Several co-distillations with water and conversion to sodium salt (Dowex 50W × 8 in Na^+ cycle) followed by freeze-drying from water gave the product **dC^{ESQ}MP** as white powder in 37% yield (0.0460 g).

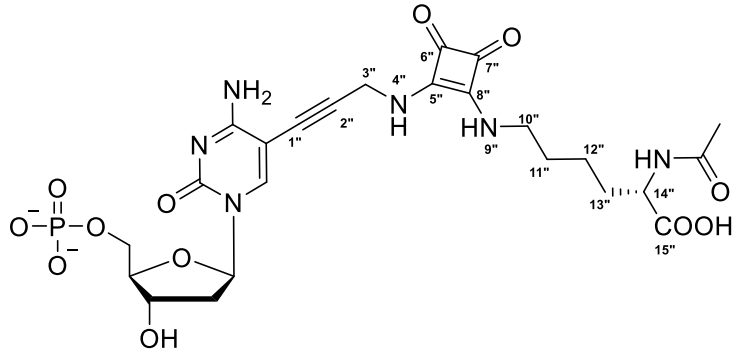
^{1}H NMR (500.0 MHz, D_2O, ref (dioxane) = 3.75 ppm): 1.43 (t, 3H, J_{vic} = 7.0, C**H$_3$**CH_2O-B); 1.45 (t, 3H, J_{vic} = 7.0, C**H$_3$**CH_2O-A); 2.29 (dt, 2H, J_{gem} = 13.9, $J_{2'b,1'} = J_{2'b,3'}$ = 6.5, H-2′b-A,B); 2.45 (ddd, 2H, J_{gem} = 13.9, $J_{2'a,1'}$ = 6.5, $J_{2'a,3'}$ = 3.9, H-2′a-A,B); 3.98 – 4.06 (m, 4H, H-5′-A,B); 4.18 (qd, 2H, $J_{4',3'} = J_{4',5'}$ = 3.9, $J_{H,P}$ = 1.0, H-4′-A,B); 4.52 (dt, 1H, $J_{3',2'}$ = 6.5, 3.9, $J_{3',4'}$ = 3.9, H-3′); 4.55 (bs, 2H, H-3″-A); 4.65 (bs, 2H, H-3″-B); 4.69 – 4.79 (m, 4H, CH_3C**H$_2$**O-A,B); 6.23 (t, 2H, $J_{1',2'}$ = 6.5, H-1′-A,B); 8.20 (s, 2H, H-6-A,B).

^{13}C NMR (125.7 MHz, D_2O, ref(dioxane) = 69.3 ppm): 17.72 (**C**H_3CH_2O-B); 17.79 (**C**H_3CH_2O-B); 37.06 (CH_2-3″-B); 37.31 (CH_2-3″-A); 42.30 (CH_2-2′-A,B); 66.77 (d, $J_{C,P}$ = 4.7, CH_2-5′-A,B); 73.38 (CH-3′-A,B); 73.56 (CH_3**C**H_2O-A,B); 77.95

(cyt-**C**≡C-CH_2-B); 78.05 (cyt-**C**≡C-CH_2-A); 88.61 (d, $J_{C,P}$ = 8.6, CH-4′-A,B); 89.22 (CH-1′-A,B); 93.11 (cyt-C≡**C**-CH_2-A); 93.55 (cyt-C≡**C**-CH_2-B); 94.40 (C-5-A,B); 147.95 (CH-6-A,B); 158.67 (C-2-A,B); 167.66 (C-4-A,B); 175.66 (C-5″-B); 175.80 (C-5″-A); 180.14 (C-8″-A); 180.74 (C-8″-B); 186.41, 186.51, 191.34, 191.67 (C-6″,7″-A,B). $^{31}P\{^1H\}$ NMR (202.4 MHz, D_2O): 2.29 (B); 2.33 (A).

HR/MS (ESI^-) for $C_{18}H_{20}O_{10}N_4P$: $[M - H]^-$ calculated 483.09, found 483.09.

N^2-acetyl-N^6-(2-{[5- (3-Amino-1-propynyl) -2'-deoxycytidine]-5'-O-monophosphate}-3,4-dioxocyclobut-1-en-1-yl)lysine ($dC^{ESQLys}MP$)



$dC^{ESQ}MP$ (0.0200 g, 0.0410 mmol) and N_α-acetyl-L-lysine (0.0150 g, 0.0820 mmol) were dissolved in borate buffer (0.500 M, pH 9, 1.30 mL) and the mixture was stirred overnight at room temperature. The product was purified by C18 reversed phase HPLC using water/methanol (5 to 100%) containing 0.100 M TEAB buffer as eluent. Several co-distillations with water and conversion to sodium salt (Dowex 50W × 8 in Na^+ cycle) followed by freeze-drying from water gave the product **$dC^{ESQLys}MP$** in 54% yield (0.0140 g).

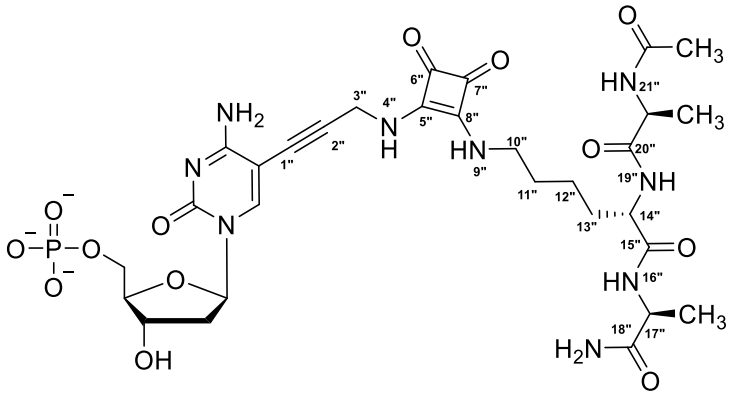
^1H NMR (500.0 MHz, D_2O, ref (dioxane) = 3.75 ppm): 1.33 – 1.44 (m, 2H, H-12″); 1.58 – 1.71 (m, 3H, H-11″,13″b); 1.80 (m, 1H, H-13″a); 1.99 (s, 3H, CH_3CO); 2.32 (dt, 1H, J_{gem} = 13.9, $J_{2'b,1'}$ = $J_{2'b,3'}$ = 6.3, H-2′b); 2.44 (ddd, 1H, J_{gem} = 13.9, $J_{2'a,1'}$ = 6.3, $J_{2'a,3'}$ = 4.7, H-2′a); 3.55 – 3.64 (m, 2H, H-10″); 3.94 – 4.03 (m, 2H, H-5′); 4.10 (dd, 1H, $J_{14'',13''}$ = 9.1, 4.5, H-14″); 4.15 (qd, 1H, $J_{4',3'}$ = $J_{4',5'}$ = 3.9, $J_{H,P}$ = 1.2, H-4′); 4.55 (ddd, 1H, $J_{3',2'}$ = 6.3, 4.7, $J_{3',4'}$ = 3.9, H-3′); 4.59, 4.63 (2 × d, 2 × 1H, J_{gem} = 17.7, H-3″); 6.23 (t, 1H, $J_{1',2'}$ = 6,3, H-1′); 8.38 (s, 1H, H-6).

^{13}C NMR (125.7 MHz, D_2O, ref(dioxane) = 69.3 ppm): 24.55 (**C**H_3CO); 24.87 (CH_2-12″); 32.59 (CH_2-11″); 33.81 (CH_2-13″); 37.14 (CH_2-3″); 42.32 (CH_2-2′); 46.82 (CH_2-10″); 57.85 (CH-14″); 65.85 (br, CH_2-5′); 72.82 (CH-3′); 78.56 (cyt-**C**≡C-CH_2); 88.93 (d, $J_{C,P}$ = 8.6, CH-4′); 88.96 (CH-1′); 93.78 (cyt-C≡**C**-CH_2); 94.50 (C-5); 148.62 (CH-6); 158.73 (C-2); 167.40 (C-4); 170.22 (C-5″); 171.39 (C-8″); 176.23 (CH_3**C**O); 182.23 (C-15″); 183.78, 184.54 (C-6″,7″).

^{31}P{^{1}H} NMR (202.4 MHz, D_2O): 4.44.

HR/MS (ESI^-) for $C_{24}H_{29}O_{12}N_6NaP$: $[M - 2H + Na]^-$ calculated 647.15, found 647.15.

[(2R,3S,5R)-5-(5-{3-({2-({5-(2-acetamidopropanamido)-6-({1-amino-1-oxopropan-2-yl}amino)-6-oxohexyl}amino)-3,4-dioxocyclobut-1-en-1-yl}amino)prop-1-yn-1-yl}-4-amino-2-oxopyrimidin-1(2H)-yl)-3-hydroxytetrahydrofuran-2-yl]methyl phosphate (dC^{ESQ3pept}MP)



dC^{ESQ}MP (0.0100 g, 0.0210 mmol) and a tripeptide (0.0130 g, 0.0410 mmol; for sequence see **Table 3**) were dissolved in borate buffer (0.500 M, pH 9; 0,650 mL) and the mixture was stirred overnight at room temperature. The product was purified by C18 reversed phase HPLC using water/methanol (0 to 100%) containing 0.100 M TEAB buffer as eluent. Several co-distillations with water and conversion to sodium salt (Dowex 50W × 8 in Na^+ cycle) followed by freeze-drying from water gave the product **dC^{ESQ3pept}MP** as white powder in 63% yield (0.0100 g).

^{1}H NMR (500.0 MHz, D_2O, ref (dioxane) = 3.75 ppm): 1.33 (d, 3H, $J_{CH3,21″}$ = 7.2, CH_3-21″); 1.37 (d, 3H, $J_{CH3,17″}$ = 7.2, CH_3-17″); 1.38 – 1.46 (bm, 2H, H-12″); 1.61 – 1.70 (bm, 2H, H-11″); 1.72 – 1.88 (bm, 2H, H-13″); 1.99 (s, 3H, CH_3CO);

2.33 (dt, 1H, J_{gem} = 13.6, $J_{2'b,1'} = J_{2'b,3'}$ = 6.3, H-2′b); 2.44 (ddd, 1H, J_{gem} = 13.6, $J_{2'a,1'}$ = 6.3, $J_{2'a,3'}$ = 5.2, H-2′a); 3.59 – 3.66 (bm, 2H, H-10″); 3.94 – 4.06 (bm, 2H, H-5′); 4.15 (q, 1H, $J_{4',3'} = J_{4',5'}$ = 3.5, H-4′); 4.21 (q, 1H, $J_{21'',CH3}$ = 7.2, H-21″); 4.24 (q, 1H, $J_{17'',CH3}$ = 7.2, H-17″); 4.25 (dd, 1H, $J_{14'',13''}$ = 8.9, 5.6, H-14″); 4.54 – 4.60 (m, 2H, H-3′,3″b); 4.64 (d, 1H, J_{gem} = 17.5, H-3″a); 6.24 (t, 1H, $J_{1',2'}$ = 6.3, H-1′); 8.47 (bs, 1H, H-6).

^{13}C NMR (125.7 MHz, D_2O, ref(dioxane) = 69.3 ppm): 19.22 (CH_3-21″); 19.33 (CH_3-17″); 24.29 (**C**H_3CO); 24.49 (CH_2-12″); 32.34 (CH_2-11″); 32.86 (CH_2-13″); 37.18 (CH_2-3″); 42.48 (CH_2-2′); 46.68 (CH_2-10″); 52.22 (CH-17″); 52.63 (CH-21″); 56.27 (CH-14″); 65.63 (d, $J_{C,P}$ = 4.5, CH_2-5′); 72.51 (CH-3′); 74.45 (cyt-**C**≡C-CH_2); 88.91 (CH-1′); 88.98 (d, $J_{C,P}$ = 9.1, CH-4′); 93.61 (cyt-C≡**C**-CH_2); 94.49 (C-5); 148.98 (CH-6); 158.74 (C-2); 167.29 (C-4); 170.16 (C-5″); 171.46 (C-8″); 176.53 (C-15″); 176.88 (CH_3**C**O); 178.33 (C-20″); 180.30 (C-18″); 183.72, 184.39 (C-6″,7″). ^{31}P{^{1}H} NMR (202.4 MHz, D_2O): 4.39.

HR/MS (ESI$^-$) for $C_{30}H_{40}O_{13}N_9NaP$: $[M + Na]^-$ calculated 788.24, found 788.24.

5.2.2 Synthesis of ESQ modified DNA and its conjugation with peptides

Incorporation of $dC^{ESQ}TP$ into DNA by PEX

19 mer, 1 modification

The reaction mixture (20 µL) contained $primer^{B}$ (3 µM, 1 µL), $template^{19\_1C}$ (3 µM, 1.50 µL), KOD XL DNA polymerase (0.250 U/µL, 0.250 µL), natural dGTP (4 mM, 0.500 µL), either natural or modified dCTP (4 mM, 1 µL) in enzyme reaction buffer (10×, 2 µL). Primer was labelled on its 5′-end by 6-carboxyfluorescein (6-FAM). The reaction mixture was incubated for 30 min at 60 °C in a thermal cycler. Finished PEX reaction was stopped by addition of PAGE stop solution and heated for 5 min at 95 °C prior to loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging (**Figure 11-a**).

20 mer, 1 modification

The reaction mixture (20 µL) contained $primer^{A}$ (3 µM, 1 µL), $template^{20\_1C}$ (3 µM, 1.50 µL), KOD XL DNA polymerase (0.250 U/µL, 0.250 µL), natural dNTPs (dGTP, dATP, dTTP; 4 mM, 0.500 µL), either natural or modified dCTP (4 mM, 1 µL) in enzyme reaction buffer (10×, 2 µL). Primer was labelled on its 5′-end by 6-carboxyfluorescein (6-FAM). The reaction mixture was incubated for 30 min at 60 °C in a thermal cycler. Finished PEX reaction was stopped by addition of PAGE stop solution and heated for 5 min at 95 °C prior to loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging (**Figure 11-b**).

31 mer, 4 modifications

The reaction mixture (20 µL) contained primer^{B} (3 µM, 1 µL), template^{31_4C} (3 µM, 1.5 µL), KOD XL DNA polymerase (0.250 U/µL, 0.300 µL), natural dNTPs (dGTP, dATP, dTTP; 4 mM, 0.700 µL), either natural or modified dCTP (4 mM, 0.700 µL) in enzyme reaction buffer (10×, 2 µL). Primer was labelled on its 5´-end by 6-carboxyfluorescein (6-FAM). The reaction mixture was incubated for 30 min at 60 °C in a thermal cycler. Finished PEX reaction was stopped by addition of PAGE stop solution and heated for 5 min at 95 °C prior to loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging (**Figure 11-b**).

98 mer, 18 modifications

The reaction mixture (150 µL) contained prim$^{REV_LT25\text{-}TH}$ (100 µM, 18 µL), temp$^{FVL\text{-}A}$ (100 µM, 15 µL), KOD XL DNA polymerase (2.50 U/µL, 7.50 µL), natural dNTPs (dGTP, dATP, dTTP; 4 mM, 18.8 µL), either natural or modified dCTP (4 mM, 18.8 µL) in enzyme reaction buffer (10×, 15 µL). Primer was labelled on its 5´-end by 6-carboxyfluorescein (6-FAM). The reaction mixture was incubated for 5 min at 95 °C, 1.5 min at 50 °C and then 2 h at 60 °C in a thermal cycler. Finished PEX reaction was stopped by addition of PAGE stop solution and heated for 5 min at 95 °C prior to loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging (**Figure 11-c**).

Preparation of ethoxy squarate-modified oligonucleotide (ON_C^{ESQ_19}, ON_C^{ESQ}) by PEX followed by magnetoseparation

19 mer, 1 modification (ON_C^{ESQ_19})

The reaction mixture (50 µL) contained KOD XL DNA polymerase (0.250 U/µL, 0.850 µL), primer^{B} (100 µM, 1.70 µL), 5′-biotinylated template^{19_1C} (100 µM, 1.70 µL), dNTPs (either natural or modified; 4 mM, 2.38 µL each) in KOD XL reaction buffer (10×, 5 µL). The reaction mixture was incubated in a thermal cycler for 30 min at 60 °C and stopped by cooling to 4 °C.

Streptavidine magnetic particles (Roche; 50 µL) were washed with binding buffer (3 × 200 µL; 10 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 7.5). The PEX solution and binding buffer (200 µL) were added. The mixture was incubated for 30 min at 15 °C and 1400 rpm. The magnetic beads were collected on a magnet (DynaMagTM-2, Invitrogen), and washed with wash buffer (3 × 300 µL; 10 mM Tris, 1 mM EDTA, 500 mM NaCl, pH 7.5) and water (4 × 300 µL). Then water (50 µL) was added and the sample was denatured for 2 min at 900 rpm and 55 °C. The beads were collected on a magnet and the solution was transferred into a clean vial. The product was evaporated to dryness, then dissolved in the water and analysed by MALDI-TOF mass spectrometry (the results are in **Table 4**).

20 mer, 1 modification (ON_C^{ESQ})

The reaction mixture (100 µL) contained KOD XL DNA polymerase (2.50 U/ µL, 0.480 µL), primer^{A} (100 µM, 4 µL), 5′-biotinylated template^{20_1C} (100 µM, 4 µL), dNTPs (either natural or modified; 4 mM, 0.375 µL each) in KOD XL reaction buffer (10×, 10 µL). The reaction mixture was incubated in a thermal cycler for 30 min at 60 °C and stopped by cooling to 4 °C.

Streptavidine magnetic particles (Roche; 100 µL) were washed with binding buffer (3 × 200 µL; 10 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 7.5). The PEX solution and binding buffer (200 µL) were added. The mixture was incubated for 30 min at 15 °C and 1400 rpm. The magnetic beads were collected on a magnet (DynaMagTM-2, Invitrogen), and washed with wash buffer (3 × 300 µL; 10 mM

Tris, 1 mM EDTA, 500 mM NaCl, pH 7.5) and water (4 × 300 µL). Then water (50 µL) was added and the sample was denatured for 2 min at 900 rpm and 55 °C. The beads were collected on a magnet and the solution was transferred into a clean vial. The product was evaporated to dryness, then dissolved in the water and analysed by MALDI-TOF mass spectrometry (the results are in **Table 4**).

PCR with squaramate-modified dCTP

Agarose gel electrophoresis

PCR product containing 6× DNA loading dye (60 mM EDTA, 10 mM Tris-HCl (pH 7.6), 60% glycerol, 0.0300% bromphenol blue, 0.0300% xylene cyanol FF; Thermo Scientific) was subjected to horizontal electrophoresis (Owl EasyCastB, Thermo Scientific) and analysed on agarose gel (containing 0.5× TBE buffer, pH 8). The gel was run at 118 V for ca. 90–120 min. PCR product was visualized with GelRed (Biotium, 10 000X in H_2O) under UV in the GBox.

98-mer

The PCR mixture (20 µL) contained KOD XL DNA polymerase (2.50 U/µL, 1 µL), $primer^{FOR\_L20}$ (10 µM, 4 µL), $primer^{REV\_LT25\text{-}TH}$ (10 µM, 4 µL), $temp^{FVL\text{-}A}$ (1 µM, 0.500 µL), natural dNTPs (4 mM, 0.150 µL of each), **$dC^{ESQ}TP$** (4 mM, 1 µL) and KOD XL reaction buffer (10×, 2 µL) supplied by the manufacturer.

PCR cycler was preheated to 80 °C and 30 PCR cycles were run under following conditions: preheating for 3 min at 94 °C, denaturation for 1 min at 95 °C, annealing for 1 min at 53 °C, extension for 1 min at 72 °C followed by final extension step of 2 min at 75 °C. The reaction was stopped by cooling to 4 °C. The PCR product was analysed on a 2% agarose gel in 0.5× TBE running buffer (**Figure 12**).

235-mer

The PCR mixture (20 µL) contained KOD XL DNA polymerase (2.50 U/µL, 1 µL), primer^{FOR_235} (10 µM, 4 µL), primer^{REV_235} (10 µM, 4 µL), temp^{PCR_235} (1.90 µM, 0.145 µL), natural dNTPs (0.400 mM, 1.50 µL of each), **dC^{ESQ}TP** (0.400 mM, 1.50 µL) and KOD XL reaction buffer (10×, 2 µL) supplied by the manufacturer.

PCR cycler was preheated to 80 °C and 40 PCR cycles were run under following conditions: preheating for 3 min at 94 °C, denaturation for 1 min at 94 °C, annealing for 1 min at 55 °C, extension for 1 min at 72 °C followed by final extension step of 5 min at 75 °C. The reaction was stopped by cooling to 4 °C. The PCR product was analysed on a 1.30% agarose gel in 0.5× TBE running buffer (**Figure 12**).

Reaction of DNA_C^{ESQ} with sulfo-Cy-5-amine (DNA_C^{ESQCy5})

Fluorescence meassurements

Natural DNA and **DNA_C^{ESQ}** were prepared by PEX in semi-preparative scale as described above. The products were purified on QIAquick Nucleotide Removal Kit (QIAGEN) eluted with water, evaporated to dryness, and reconstituted in borate buffer (0.500 M, pH 9). The reaction mixtures for sulfo-Cy-5-amine addition (10 µL) contained purified PEX product (0.200 nmol of natural or modified DNA) and sulfo-Cy-5-amine (13.5 mM, 1.48 µL). All samples were incubated on 37 °C 36 h. The products were purified on QIAquick Nucleotide Removal Kit (QIAGEN) eluted with water and the fluorescence of the products was measured on spectrofluorimeter (water; λ_{ex} = 646 nm, λ_{em} = 662nm, **Figure 13**).

PAGE analysis

Natural DNA and **DNA_C^{ESQ}** were prepared by PEX in semi-preparative scale as described above. Primer was labelled on its 5´-end by 6-carboxyfluorescein (6-FAM). The products were purified on QIAquick Nucleotide Removal Kit (QIAGEN) and eluted with water. The reaction mixtures for sulfo-Cy-5-amine addition (10 µL) contained purified PEX product (0.200 nmol of natural

or modified DNA) and sulfo-Cy-5-amine (13.5 mM, 1.48 µL) in borate buffer (0.500 M, pH 9). All samples were incubated on 37 °C 36 h. The products were purified on QIAquick Nucleotide Removal Kit (QIAGEN) eluted with water evaporated to dryness and diluted with mixture water/50% glycerol (2:1). Samples were heated for 5 min at 95 °C and separated by 20% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging (**Figure 13**).

Reaction of DNA_C^{ESQ} with lysine and lysine containing peptides (DNA_C^{ESQLys}, DNA_C^{ESQ3pept}, DNA_C^{ESQ10pept})

PAGE analysis

DNA_C^{ESQ} was prepared by PEX in semi-preparative scale as described above. Primer was labelled on its 5´-end by 6-carboxyfluorescein (6-FAM). The product was purified on QIAquick Nucleotide Removal Kit (QIAGEN) eluted with water and evaporated to dryness. The reaction mixture (20 µL) contained **DNA_C^{ESQ}** (4.30 µM) and Ac-Lys-OH, tri- or decapeptide (11 mM) in borate buffer (0.500 M, pH 9). The reaction was incubated 36 h at 37 °C in a thermal cycler, evaporated to dryness, reconstituted in water (20 µL) and PAGE stop solution and heated for 5 min at 95 °C. Samples were separated by 20% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging (**Figure 14**).

MALDI-TOF analysis (*ON_C^{ESQLys}, ON_C^{ESQ3pept}, ON_C^{ESQ10pept}*)

DNA_C^{ESQ} was prepared by PEX in semi-preparative scale as described above. Primer was labelled on its 5´-end by biotin. The product was purified on QIAquick Nucleotide Removal Kit (QIAGEN) eluted with water and evaporated to dryness. The reaction mixture (20 µl) contained **DNA_C^{ESQ}** (0.200 nmol) and Ac-Lys-OH, tri- or decapeptide (22 mM) in borate buffer (0.500 M, pH 9). The reaction was incubated 36 h at 37 °C in a thermal cycler followed by magnetoseparation and

analysed by MALDI-TOF mass spectrometry (the results are in **Table 4**; for spectrum see **Figure A1-3**).

Cross-linking of DNA_C^{ESQ} and individual recombinant proteins (BSA, H2A, H2B, H3.1 and H4)

FAM labeled natural DNA and **DNA_C^{ESQ}** were prepared by PEX in semi-preparative scale as described above and purified on QIAquick Nucleotide Removal Kit (QIAGEN). 25 pmol of natural or modified DNA was incubated with 50 pmol of individual proteins (**Table 5**) in 10 µL of phosphate buffer (4.50 mM, ph 7.4) at 37 °C for 36 h. The reaction was diluted with 2× VPS loading buffer, denatured 2 min at 100 °C prior to loading and analysed by 17.5% SDS denaturing PAGE (acrylamide/methylenebisacrylamide 29:1; 1.92 M glycine, 0.250 M Tris, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by fluorescence imaging (**Figure 17-a**).

Alternatively, nonlabeled **DNA_C^{ESQ}** (483 pmol) was incubated with individual histone proteins (161 pmol) in 35 µL of phosphate buffer (4.50 mM, ph 7.4) at 37 °C for 36 h. The reaction was diluted with 2× VPS loading buffer, denatured 2 min at 100 °C prior to loading and analysed by 17.5% SDS denaturing PAGE (acrylamide/methylenebisacrylamide 29:1; 1.92 M glycine, 0.250 M Tris, 0.100% SDS) at room temperature (230 V, 70 min) and stained with PageBlue^{TM} protein staining solution (Thermofisher; **Figure 17-b**).

Cross-linking of 98-mer PEX product and H3.1 recombinant protein

FAM labeled natural and modified 98-mer PEX products were prepared as described above purified on QIAquick Nucleotide Removal Kit (QIAGEN) and eluted with water. 25 pmol of natural or modified DNA was incubated with 2 or 10 equiv of H3.1 protein in 10 µL of phosphate buffer (4.50 mM, ph 7.4) at 37 °C 36 h. The reaction mixtures were diluted with 2× VPS loading buffer, denatured 2 min at 100 °C prior to loading and analysed by 12% SDS-PAGE (acrylamide/methylenebisacrylamide 29:1; 1.92 M glycine, 0.250 M Tris,

0.100% SDS) at room temperature (42 mA, 80 min). Visualization was performed by fluorescence imaging (**Figure 17-c**).

Cross-linking of DNA_C^{ESQ_19}, DNA_4C^{ESQ} and GSTp53CD

Modified DNA_C^{ESQ_19} and DNA_4C^{ESQ} were prepared by PEX as described above. The products were purified on QIAquick Nucleotide Removal Kit (QIAGEN) and eluted with water. The reaction mixtures for GSTp53CD protein binding (20 µL) were prepared from purified PEX (6 ng/µL, 10 µL), KCl (500 mM, 2 µL), DTT (2 mM, 2 µL), VP buffer (50 mM Tris, 0.100% Triton-X100, pH 7.6; 2 µL) and GSTp53CD stock solution (400 ng/µL in 25 mM HEPES pH 7.6, 200 mM KCl, 10% glycerol, 1 mM DTT, 1 mM benzamidine; 2 µL). Control samples were prepared using corresponding natural DNA. All samples were incubated for 36 h at 37 °C. 2× VPS loading buffer was added, and the mixture was denatured 2 min at 100 °C prior to loading and analysed by 10% SDS-PAGE (0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (100 V/40 min then 150 V/60 min). Visualization was performed by fluorescence imaging (**Figure 17-d**).

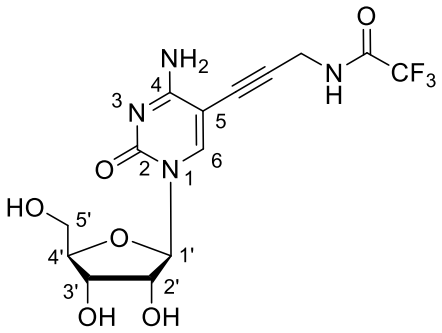
Characterization of crosslinks between DNA_C^{ESQ} and individual recombinant proteins (H2A, H2B, H3.1 and H4) by mass spectrometry

DNA_C^{ESQ} containing one modification (**Table 1**) was cross-linked to histones (H2A, H2B, H3.1, H4) as described above. 10 uL of reaction mixture was injected onto bioZen Intact C4 column (Phenomenex) and separated by gradient of acetonitrile in water (both mobile phases modified by 0.100% Formic acid). The separation was carried out by LC system (I-class, Waters) coupled to Mass Spectrometer (Synapt G2, Waters) to acquire *m/z* by positive electrospray ionization. Raw mass spectra of chromatographic peak containing conjugate was combined, subtracted and deconvoluted by MaxEnt1[220] algorithm (for spectrum see **Figure A4-6**).

5.3 Enzymatic synthesis of reactive RNA probes for cross-linking with proteins

5.3.1 Synthesis and characterization of ESQ modified cytidine and cytidine triphosphate

5-[3-(Trifluoroacetamido)-prop-1-ynyl]-cytidine (18)[193,221]



5-Iodo-cytidine (**17**; 0.443 g, 1.20 mmol), *N*-propargyl trifluoroacetamide (**14**; 0.544 g, 3.60 mmol), CuI (0.0450 g, 0.240 mmol), $Pd(PPh_3)_4$ (0.139 g, 0.120 mmol) and Amberlite Ira-67 resin (1.07 g) were placed in a 50 mL round bottom flask and dissolved in anhydrous DMF (6 mL) under argon atmosphere. The reaction mixture was stirred in the dark at r.t. for 24 hours then filtered over bed of SiO_2 topped with celite and eluted with mixture of DCM/methanol 7:1. The solvents were removed under reduced pressure and the title product was obtained by column chromatography using DCM/methanol (8:2) as a mobile phase in 71% yield (0.336 g) as an orange oil.

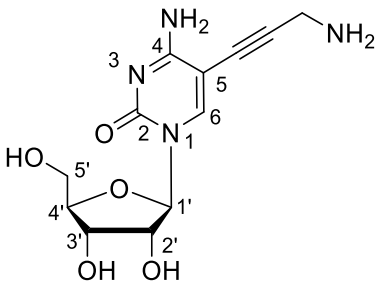
^{1}H NMR (400.1 MHz, CD_3OD): 3.76 (dd, 1H, J_{gem} = 12.4, $J_{5'b,4'}$ = 2.9, H-5′b); 3.91 (dd, 1H, J_{gem} = 12.4, $J_{5'a,4'}$ = 2.6, H-5′a); 4.03 (ddd, 1H, $J_{4',3'}$ = 5.6, $J_{4',5'}$ = 2.9, 2.6, H-4′); 4.09 – 4.18 (m, 2H, H-2′,3′); 4.31 (s, 2H, CH_2N); 5.84 (d, 1H, $J_{1',2'}$ = 2.9, H-1′); 8.41 (s, 1H, H-6).

^{13}C NMR (125,7 MHz, CD_3OD): 30.96 (CH_2N); 61.51 (CH_2-5′); 70.36 (CH-3′); 75.39 (**C**≡CCH_2); 76.37 (CH-2′); 85.83 (CH-4′); 90.92 (C≡**C**CH_2); 92.04 (C-5); 92.22 (CH-1′); 117.26 (q, $J_{C,F}$ = 286.5, **C**F_3CO); 146.43 (CH-6); 156.86 (C-2); 158.73 (q, $J_{C,F}$ = 37.5, CF_3**C**O); 166.30 (C-4).

^{19}F NMR (376.5 MHz, CD_3OD): -77.27.

HR/MS (ESI$^+$) for $C_{14}H_{16}O_6N_4F_3$: $[M + H]^+$ calcd 393.09, found 393.10.

5-(3-Amino-prop-1-ynyl)- cytidine (19)

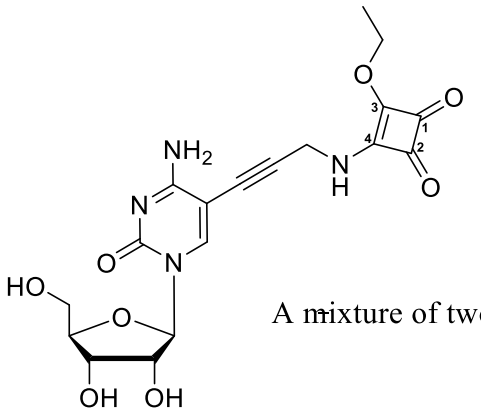


18 (0.336 g, 0.855 mmol) was dissolved in the mixture of 6 mL of HPLC water and 24 mL of aqueous ammonium hydroxide in the pressure tube and the tube was closed properly. The reaction was stirred overnight at r.t. and then concentrated down. The crude product was redissolved in HPLC H_2O (10 mL) and DOWEX 50 × 8 resin (2.50 g) was added. The mixture was stirred for 50 min and filtered over a bed of Dowex 50 × 8 resin (2.50 g) which was then washed with HPLC H_2O, and the product was eluted off the resin with HPLC H_2O/conc. NH_4OH (4:1, 400 mL). Final product was obtained as orange-brown oil (0.219 g, 87%) after removal of the solvents under reduced pressure.

^{1}H NMR (400,1 MHz, CD_3OD): 3.62 (s, 2H, CH_2N); 3.76 (dd, 1H, J_{gem} = 12.4, $J_{5'b,4'}$ = 2.8, H-5′b); 3.91 (dd, 1H, J_{gem} = 12.4, $J_{5'a,4'}$ = 2.5, H-5′a); 4.03 (ddd, 1H, $J_{4',3'}$ = 5.6, $J_{4',5'}$ = 2.8, 2.5, H-4′); 4.08 – 4.17 (m, 2H, H-2′,3′); 5.84 (d, 1H, $J_{1',2'}$ = 2.9, H-1′); 8.37 (s, 1H, H-6).

HR/MS (ESI$^+$) for $C_{12}H_{16}O_5N_4Na$: $[M + Na]^+$ calcd 319.10, found 319.10.

4-[5- (3-Amino-prop-1-ynyl) -cytidine]-3- ethoxycyclobut-3-ene-1,2-dione (C^{ESQ})



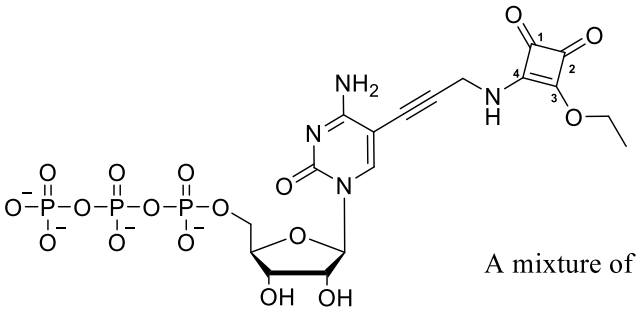
A mixture of two isomers 1:1

19 (0.120 g, 0.405 mmol) was well suspended in 2 ml of ethanol. Diethyl squarate (0.120 mL, 0.810 mmol) was added dropwise during 30 min to the stirred solution and the reaction mixture was stirred another 1 hour at room temperature. Column chromatography using DCM/methanol (8:2) as a mobile phase gave the desired product as white powder in 22% yield (0.0370g).

^{1}H NMR (500.0 MHz, DMSO-d_6): 1.37 (t, 6H, J_{vic} = 7.1, C**H$_3$**CH$_2$O); 3.55, 3.68 (2 × ddd, 2 × 2H, J_{gem} = 12.1, $J_{5',OH}$ = 4.9, $J_{5',4'}$ = 2.9, H-5′); 3.83 (dt, 2H, $J_{4',3'}$ = 5.2, $J_{4',5'}$ = 2.9, H-4′); 3.91 – 3.96 (m, 4H, H-2′,3′); 4.37, 4.56 (2 × bs, 2 × 2H, CH$_2$N); 4.66 (q, 4H, J_{vic} = 7.1, CH$_3$C**H$_2$**O); 4.49 (m, 2H, OH-3′); 5.15 (t, 2H, $J_{OH,5'}$ = 4.9, OH-5′); 5.36 (m, 2H, OH-2′); 5.75 (d, 2H, $J_{1',2'}$ = 3.5, H-1′); 6.82, 6.84 (2 × bs, 2 × 1H, NH$_a$**H$_b$**); 7.84 (s, 2H, N**H$_a$**H$_b$); 8.26 (s, 2H, H-6); 8.95, 9.11 (2 × bs, 2 × 1H, NH). ^{13}C NMR (125.7 MHz, DMSO-d_6): 15.17 (**C**H$_3$CH$_2$O); 34.06, 34.51 (CH$_2$N); 60.29 (CH$_2$-5′); 69.19 (CH-3′); 69.28 (CH$_3$**C**H$_2$O); 74.48 (CH-2′); 76.42, 76.55 (cyt-**C**≡C-CH$_2$); 84.45 (CH-4′); 89.12 (C-5); 89.95 (CH-1′); 90.22, 90.82 (cyt-C≡**C**-CH$_2$); 145.08 (CH-6); 153.83 (C-2); 164.43 (C-4); 171.80, 172.47 (C-4-cyclobut); 177.33, 177.76 (C-3-cyclobut); 182.47, 182.79, 188.92, 189.48 (C-1,2-cyclobut).

HR/MS (ESI$^+$) for $C_{18}H_{21}O_8N_4$: $[M + H]^+$ calcd 421.13, found 421.13.

4-{[5- (3-Amino-propynyl) -cytidine]-5'-*O*-triphosphate}-3- ethoxycyclobut-3-ene-1,2-dione ($C^{ESQ}TP$)



A mixture of two isomers 1:1

Ethoxy squarate modified nucleoside (**C^{ESQ}**, 0.0310 g, 0.0740 mmol) was dried at 80 °C for 2 hours in vacuo. After cooling, $PO(OMe)_3$ (0.240 mL) and $POCl_3$ (8.40 µL) were added on ice under argon atmosphere. The reaction mixture was stirred for 90 min at 0 °C. In a separate flask, the mixture of $(NHBu_3)H_2P_2O_7$ (0,220 g) and Bu_3N (54 µL) in dry acetonitrile (0.600 mL) was prepared under argon atmosphere, cooled to 0 °C and then added by the syringe to the reaction mixture. The mixture was stirred for 1 hour at 0 °C. The reaction was stopped by addition of TEAB (2 M, 0.400 mL) and water (1.19 mL). The product was purified by C18 reversed phase HPLC using water/methanol (5 to 50%) containing 0.100 M TEAB buffer as eluent. Several co-distillations with water and conversion to sodium salt (Dowex 50W × 8 in Na^+ cycle) followed by freeze-drying from water gave the product **$C^{ESQ}TP$** as white powder in 38% yield (0.202 g).

^{1}H NMR (500.0 MHz, D_2O, ref(*t*BuOH) = 1.24 ppm): 1.45 (bt, 6H, J_{vic} = 7.0, C**H₃**CH_2O-A,B); 4.25 – 4.33 (m, 8H, H-2′,4′,6′); 4.40 (t, 2H, $J_{3',2'} = J_{3',4'}$ = 5.4, H-3′); 4.57, 4.67 (2 × bs, 2 × 2H, CH_2N); 4.70 – 4.85 (m, 4H, CH_3C**H₂**O, overlapped with water signal); 5.94 (td, 2H, $J_{1',2'}$ = 3.8, H-1′); 8.25 (s, 2H, H-6).

^{13}C NMR (125.7 MHz, D_2O, ref(*t*BuOH) = 32.43 ppm): 17.87 (**C**H_3CH_2O); 37.20, 37.46 (CH_2N); 67.26 (d, $J_{C,P}$ = 4.4, CH_2-5′); 71.51 (CH-3′); 73.64 (CH_3**C**H_2O); 77.26 (CH-2′); 78.04 (cyt-**C**≡C-CH_2); 85.51 (d, $J_{C,P}$ = 9.0, CH-4′); 92.54 (CH-1′); 93.37, 93.79 (cyt-C≡**C**-CH_2); 94.79 (C-5); 148.09 (CH-6); 159.03 (C-2); 167.86 (C-4); 175.84, 176.01 (C-4-cyclobut); 180.21, 180.88 (C-3-cyclobut); 186.61, 186.67, 191.50, 191.88 (C-1,2-cyclobut).

^{31}P{^{1}H} NMR (202.4 MHz, D_2O): -21.94, -21.89 (2 × bt, *J* = 19.7, 2P, P$_\beta$); -10.64 (d, 2P, *J* = 19.7, P$_\alpha$); -8.10 (br, 2P, P$_\gamma$).

HR/MS (ESI$^-$) for $C_{18}H_{22}O_{17}N_4P_3$: $[M - H]^-$ calcd 659.02, found 659.02.

5.3.2 Synthesis of ESQ modified RNA and its conjugation with proteins

In vitro transcription with $C^{ESQ}TP$

A solution of template DNA oligonucleotides (100 µM each) in 1:1 ratio in annealing buffer (Tris (10 mM, NaCl (50 mM), EDTA (1 mM), pH 7.8) was heated to 95 °C for 5 minutes and slowly cooled to 25 °C in 45 minutes. The resulting DNA (50 uM) was used as a template for transcription reactions.

In vitro transcription reactions were performed in the total volume of 20 µL containing: dsDNA template (2 µM), ATP (1 mM), UTP (1 mM), GTP (0.800 mM), CTP (1 mM) for positive control, $C^{ESQ}TP$ (1.60 mM) for modified RNA, [α-^{32}P]-GTP (111 TBq/mmol, 370 MBq/mL, 0.200 µL), T7 RNA polymerase (2.50 U/µL, Thermoscientific), $MgCl_2$ (4.80 mM), dithiothreitol (DTT; 12 mM), Triton X-100 (0.120%), dimethyl sulfoxide (DMSO; 5%) and 1× transcription buffer B (40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl, 2mM betain). Water was used instead of the solution of **$C^{ESQ}TP$** in the negative control experiment. The mixture was incubated at 37 °C for 2 h. The aliquots (3 µL) were mixed with 2× PAGE stop solution, denatured (75 °C, 10 min) and analysed by 12.5% PAGE (7M urea, AA/bisAA 19:1, 40 mA/1h, 1× TBE). Visualization was performed by phosphoimaging (**Figure 19**).

DNAse I treatment

The DNA template was digested by DNAse I to attain pure RNA. Transcription mixture (25 µL), 4 µL of DNAse I buffer (supplied with the enzyme), 4 µL of DNAse I (4 U; NEB), and 7 µL DEPC water were incubated 60 min at 37 °C. The enzyme was heat deactivated at 75 °C for 10 min followed with cooling on ice. All samples were purified on NucAway Spin Columns (50 µL) for further use.

MALDI-TOF analysis of ethoxy squarate-modified RNAs

A solution of template DNA oligonucleotides (100 μM each) in 1:1 ratio in annealing buffer (Tris (10 mM, NaCl (50 mM), EDTA (1 mM), pH 7.8) was heated to 95 °C for 5 minutes and slowly cooled to 25 °C in 45 minutes. The resulting DNA (50 uM) was used as a template for transcription reactions.

In vitro transcription reactions were performed in the total volume of 20 μL containing: dsDNA template (2 μM), ATP (1 mM), UTP (1 mM), GTP (1 mM), CTP (1 mM) for positive control, **C^{ESQ}TP** (1.60 mM) for modified RNA, T7 RNA polymerase (2.50 U/μL, Thermoscientific), $MgCl_2$ (4.80 mM), dithiothreitol (DTT; 12 mM), Triton X-100 (0.120%), dimethyl sulfoxide (DMSO; 5%) and 1× transcription buffer B (40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl, 2mM betain). The mixture was incubated at 37 °C for 2 h.

DNAse I treatment

The DNA template was digested by DNAse I to attain pure RNA. Transcription mixture (25 μL), 4 μL of DNAse I buffer (supplied with the enzyme), 4 μL of DNAse I (4 U; NEB), and 7 μL DEPC water were incubated 60 min at 37 °C. The enzyme was heat deactivated at 75 °C for 10 min followed with cooling on ice. All samples were purified on NucAway Spin Columns. The product was freeze-dried, then redissolved in the water and analysed by MALDI-TOF mass spectrometry (**Table 8-9**).

Reaction of ethoxy squarate-modified RNA (RNA_1C^{ESQ}) with sulfo-Cy-5-amine (RNA_C^{ESQ_Cy5})

Fluorescence measurements

Unlabelled natural RNA and **RNA_1C^{ESQ}** were prepared by *in vitro* transcription reaction described above. The reaction mixtures for sulfo-Cy-5-amine addition (10 µL) were prepared from purified transcribed product (5 µM) reconstituted in PBS buffer (pH 7.4) and sulfo-Cy-5-amine was added (2 mM). All samples were incubated on 37 °C overnight. The products were purified on NucAway Spin Columns and the fluorescence of the products was measured on a Fluoromax 4 spectrofluorimeter (HORIBA Scientific, **Figure 20-a**).

PAGE analysis

Unlabelled natural RNA and **RNA_1C^{ESQ}** were prepared by *in vitro* transcription reaction described above. The reaction mixtures for sulfo-Cy-5-amine addition (15 µL) were prepared from purified transcription product (9 µM) in TRIS-HCl buffer (40 mM, pH 7.9) (19 µL) and sulfo-Cy-5-amine (0.900 mM). All samples were incubated on 37 °C overnight. The products were purified on NucAway Spin Columns mixed with 2× PAGE stop solution without xylene cyanol and heated for 5 min at 95 °C. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour), and the gel was stained with Sybr Gold. Visualization was performed by fluorescence imaging using Typhoon FLA 9500, GE Healthcare (**Figure 20-b**).

Cross-linking of RNA_C^{ESQ} with proteins

Cross-linking of RNA_C1g^{ESQ}, RNA_C3f ^{ESQ} with T7 RNA polymerase

Radioactively labelled natural RNA and **RNA_C1g^{ESQ}, RNA_C3f ^{ESQ}** (with one and three modifications) were prepared by *in vitro* transcription as described above. Natural or modified RNA (1µM) was incubated with 2U/µL of T7 RNA polymerase at 37 °C. Reaction mixture contained $MgCl_2$ (4.80 mM), dithiothreitol (DTT; 12 mM), Triton X-100 (0.120%), dimethyl sulfoxide (DMSO; 5%) and 1× transcription buffer B (40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl, 2mM betain) and DEPC water (total reaction volume 25 µL). After 30 min 5 µL of the reaction mixture was separated by 5-10% native PAGE (acrylamide/bisacrylamide 37.5:1; 4 °C, 200 V, 0.250 M Tris, 0.192 M glycine,). The rest of the reaction was incubated up to 2 hours at 37 °C, then diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and analysed by 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by phosphoimaging (**Figure 21**).

Cross-linking of RNA_C1g^{ESQ} with Bovine Serum Albumin (BSA) and Single Strand Binding Protein (SSB)

Radioactively labelled natural RNA and **RNA_C1g^{ESQ}** were prepared by *in vitro* transcription as described above. Natural or modified RNA (1µM) was incubated with 20 µM protein (BSA resp. SSB) in 10× binding buffer D (100 mM Tris pH. 8.0, 20 mM $MgCl_2$, 100 mM KCl, 12 mM βME; 2µL), 50% glycerol (2 µL) and DEPC water (total reaction volume 20 µL) at 37 °C. After 1 hour 5 µL of the reaction mixture was separated by 5-10% native PAGE (acrylamide/bisacrylamide 37.5:1; 4 °C, 200V, 0.250 M Tris, 0.192 M glycine,). The rest of the reaction was incubated overnight at 37 °C, then diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and analysed by 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by phosphoimaging (**Figure 22**).

Cross-linking of RNA_C1g^{ESQ}, RNA_C3f^{ESQ} with Nonstructural Protein NS5 of Japanese Encephalitis Virus (JEV) and Yellow Fever Virus (YFV)

Radioactively labelled natural RNA and **RNA_C1g^{ESQ}, RNA_C3f^{ESQ}** (with one and three modifications) were prepared by *in vitro* transcription as described above. Natural or modified RNA (0.500 µM) was incubated with 2 µM of NS5 protein (JEV resp. YFV) in 10× binding buffer E (50 mM Tris-HCl pH. 7.4, 100 mM DTT, 5% Triton X-100, 10% glycerol; 2µL), 10 mM $MnCl_2$ (2 µL), 10 mM $MgCl_2$ (2 µL), 50% glycerol (2 µL) and DEPC water (total reaction volume 20 µL) at 34 °C. After 1 hour 3 µL of the reaction mixture was separated by 5-7% native PAGE (acrylamide/bisacrylamide 37.5:1; 4 °C, 200 V, 1× Tris-glycine). The rest of the reaction was incubated 48 h at 34 °C, then diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and analysed by 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by phosphoimaging (**Figure 24**).

Cross-linking of RNA_C1g^{ESQ}, RNA_C3f^{ESQ} with SARS-CoV-2 RdRp

Radioactively labelled natural RNA and **RNA_C1g^{ESQ}, RNA_C3f^{ESQ}** (with one and three modifications) were prepared by *in vitro* transcription as described above. Natural or modified RNA (1 µM) was incubated with 8 µM SARS-CoV-2 RdRp protein (complex of 2 µM SC nsp12 and 6 µM nsp7/8 proteins) in 10× binding buffer D (100 mM Tris-HCl pH. 8, 20 mM $MgCl_2$, 100 mM KCl 12 mM βME; 2µL), 50% glycerol (2 µL) and DEPC water (total reaction volume 20 µL) at 30 °C. After 1 hour 3 µL of the reaction mixture was separated by 10% native PAGE (acrylamide/bisacrylamide 37.5:1; 4 °C, 200 V, 1× Tris-glycine). The rest of the reaction was incubated overnight at 30 °C, then diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and analysed by 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by phosphoimaging (**Figure 25**).

Cross-linking of RNA_C1g^{ESQ}, RNA_C3f ^{ESQ} with SARS-CoV-2 nucleoprotein

Radioactively labelled natural RNA and **RNA_C1g^{ESQ}, RNA_C3f ^{ESQ}** (with one and three modifications) were prepared by *in vitro* transcription as described above. Natural or modified RNA (1 μM) was incubated with 8 μM SARS-CoV-2 nucleoprotein in 10× binding buffer D (100 mM Tris-HCl pH. 8, 20 mM $MgCl_2$, 100 mM KCl 12 mM βME; 2μL), 50% glycerol (2 μL) and DEPC water (total reaction volume 20 μL) at 30 °C. After 1 hour 3 μL of the reaction mixture was separated by 10% native PAGE (acrylamide/bisacrylamide 37.5:1; 4 °C, 200 V, 1× Tris-glycine). The rest of the reaction was incubated overnight at 30 °C, then diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and analysed by 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by phosphoimaging (**Figure 25**).

Cross-linking of RNA_C1g^{ESQ} with HIV-rt

Radioactively labelled natural RNA and **RNA_C1g^{ESQ}** (with one modification) was prepared by *in vitro* transcription as described above.

Natural or modified RNA (0.500 μM) was incubated with 1 μM HIV-rt in buffer B (40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl, 2mM betain), 50% glycerol (2 μL) and DEPC water (total reaction volume 20 μL) at 37 °C. After 1 hour 3 μL of the reaction mixture was separated by 10% native PAGE (acrylamide/bisacrylamide 37.5:1; 4 °C, 200 V, 1× Tris-glycine). The rest of the reaction was incubated overnight at 37 °C, then diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and analysed by 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization was performed by phosphoimaging (**Figure 26**).

Gel-based polymerase assays in vitro

JEV and YFV RdRp mediated RNA extension assay[218]

The polymerase activity of NS5 proteins was determined in a PEX reaction using fluorescently labelled RNA templates (**Table 13**). The reaction mixture (20 µl) contained reaction buffer (5 mM Tris-HCl pH. 7.4, 10 mM DTT, 0.100% Triton X-100, 1% glycerol, 1 mM $MnCl_2$, 1 mM $MgCl_2$), 10 µM NTPs, 200 nM template and 800 nM NS5 proteins. In the positive control, all natural NTPs were used and in the modified version 10 µM **C^{ESQ}TP** instead of CTP was used. The reactions were incubated for 1 h and overnight respectively, at 34 °C. Then the reactions were incubated with Proteinase K (1U) 30 min at room temperature. Reactions were stopped by adding of 2× PAGE stop solution and denatured at 95 °C for 10 min prior to loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging using Typhoon FLA 9500, GE Healthcare (**Figure 27-b/c**). Next, the gel was stained with PageBlue™ protein staining solution (Thermofisher; **Figure 27-e**).

Immunodetection of RNA_C^{ESQ_JEVNS5} conjugate

Natural RNA and **RNA_C^{ESQ}** were prepared by *in vitro* polymerase assay as described above. The reaction mixtures were diluted with 2× VPS loading buffer, denatured 10 min at 95 °C prior to loading and electrophoresed on 5-10% SDS denaturing PAGE (acrylamide/bisacrylamide 37.5:1; 0.250 M Tris, 0.192 M glycine, 0.100% SDS) at room temperature (230 V, 70 min). Visualization of FAM labelled cross-linked product was performed by fluorescence imaging using Typhoon FLA 9500, GE Healthcare (**Figure 27-d**). The gel was blotted (dry; 0.250 M Tris, 0.192 M glycine, 10% methanol, 0.100% SDS; 12 V/2 hours at r.t.) to polyvinylidene fluoride Immobilon-P transfer membrane (Millipore). The membrane was washed with TBS (2.50 mM Tris, 15 mM NaCl, 0.200 mM KCl, pH 7.4), blocked with 5% non-fat milk (in T-TBS; 2.50 mM Tris, 15 mM NaCl, 0.200 mM KCl, 0.500% Tween-20, pH 7.4) for 30 min and afterwards washed with T-TBS (20 ml, 10 min). The membrane was incubated with JEV NS5 polyclonal

antibody (rabbit/IgG; Invitrogen) at dilution 1:1000 in 5% non-fat milk in T-TBS at 4 °C overnight. The unbound antibodies were washed out with T-TBS (20 ml, 3 × 5 min). Next, the membrane was incubated with anti-rabbit IgG HRP linked secondary antibody (Invitrogen) at 1:1000 dilution in in 5% non-fat milk in T-TBS at 25 °C for 1 hour. The access of antibodies was removed by washing with T-TBS (20 ml, 3 × 5 min) and TBS (20 ml, 10 min). The membrane was incubated with SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) for 1 min. Chemiluminescence was measured on an ImageQuant LAS 4000 Mini luminescence analyser (GE Healthcare; **Figure 27-d**).

SARS-CoV-2 RdRp mediated RNA extension assay[222]

The polymerase activity of SARS-CoV-2 RdRp was determined in a PEX reaction using fluorescently labelled RNA templates (**Table 13**). The reaction mixture (20 µl) contained reaction buffer (10 mM Tris-HCl (pH. 8.0), 2 mM $MgCl_2$, 10 mM KCl, 1 mM βME), 10 µM NTPs, 0.500 µM template, 1 µM nsp12 and 3 µM nsp7/8 proteins. In the positive control, all natural NTPs were used and in the modified version 10 µM **C^{ESQ}TP** instead of CTP was used. The reactions were incubated for 1 h at 30 °C. Reactions were stopped by adding of 2× PAGE stop solution and denatured at 95 °C for 10 min prior to loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour). Visualization was performed by fluorescence imaging using Typhoon FLA 9500, GE Healthcare (**Figure 28**).

For MALDI TOF analysis, the reaction mixture was desalted with Bio-Spin6/Biorad columns (buffer was exchanged for water according to supplied protocol; the results are in **Table 13**).

Characterization of cross-linking products between RNA_C^{ESQ} and individual proteins by mass spectrometry

Intact mass measurement of RNA_C1g^{ESQ_SSB}

RNA_C1g^{ESQ} containing one modification (**Table 7**) was cross-linked to SSB protein as described above.

The volume of 18 µl of the sample was injected onto a MassPREP Micro Desalting column (20 µm, 5-mm by 2.1-mm ID, Waters) and desalted and eluted by fast gradient (4 min). Mobile phase A (10 mM ammonium acetate in H_2O) and B (acetonitrile) were used for elution. The separation was carried out by AQUITY UPLC I-Class system (Waters) and was on-line coupled to Mass Spectrometer Synapt G2 (Waters) to acquire mass spectra using electrospray ionization in positive mode. The TOF mass range was set from *m/z* 500 to 4000. The raw spectrum was subtracted and deconvoluted (MaxEnt 1, Waters) to produce the final spectrum (**Figure A7**).

LC-MS analysis of digested RNA_C1g^{ESQ_JEV} samples after polymerase assay

The JEV NS5 mediated RNA extension assay was performed as described above (20 × 20 µl reaction mixture), except the reaction was incubated overnight. The reaction was followed by trypsine digestion of the protein to peptides and RNA digestion by RNase A/T1 mix.

Samples were dissolved in 15 µl of 0.1 TFA in H_2O and 3 µl of the sample was injected on an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a Mass Spectrometer Orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific). The peptides were trapped on a PepMap100 column (5 µm, 5 mm by 300-µm internal diameter (ID); Thermo Fisher Scientific) and desalted with 2% acetonitrile in 0.100% formic acid at a flow rate of 5 µL/min. Eluted peptides were separated using an EASY-Spray PepMap100 C18 analytical column (2 µm, 50-cm by 75-µm ID; Thermo Fisher Scientific). The 30-min elution gradient at a constant flow rate of 300 nL/min was set to start at 5% phase B (0.100% formic acid in 99.9% acetonitrile) and 95% phase A (0.100% formic acid). Then,

the content of acetonitrile was increased gradually up to 50 % of phase B. The orbitrap mass range was set from *m/z* 350 to 2000 in the MS mode, and for ions with a charge state 2-6 the fragmentation spectra were acquired. A Proteome Discoverer 2.5 (Thermo Fisher Scientific) was used for identification of peptide and protein using Sequest HS and MS Amanda as search engines and databases of protein sequence and common contaminants (**Table 15; Figure 29-30; Figure A10-11**).

6. Appendices

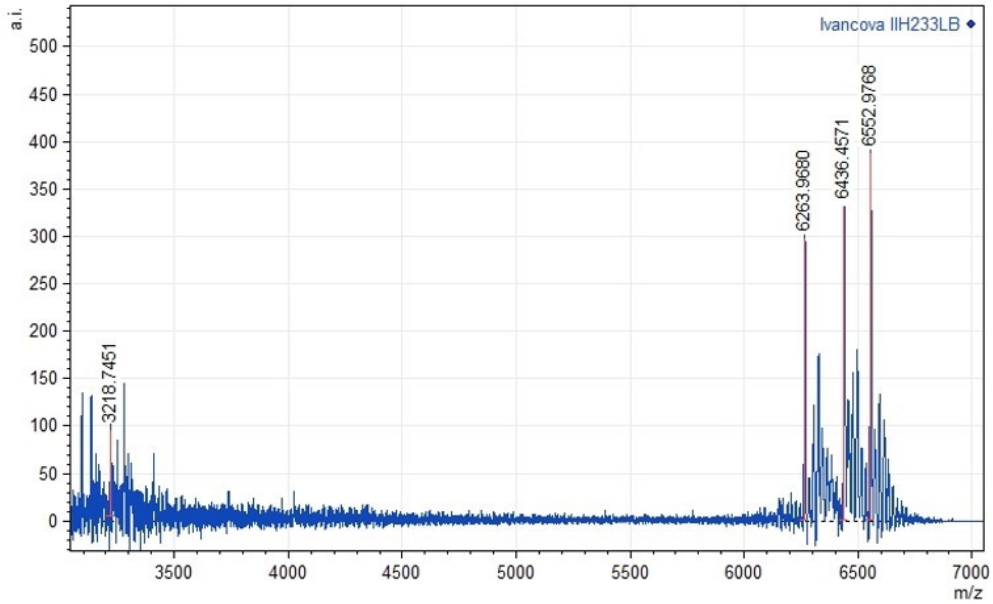


Figure A1. MALDI-TOF MS spectrum of **ON_C^{ESQLys}**, calculated for [M + H]: 6435.5 Da; found 6436.4 Da; the peak at *m/z* = 6552.9 can be assigned to the biotinylated template and the peak at *m/z* = 6263.9 can be assigned to the hydrolysed **ON_C^{ESQ}** [- CH_2CH_3 - H^+] (**Table 4**).

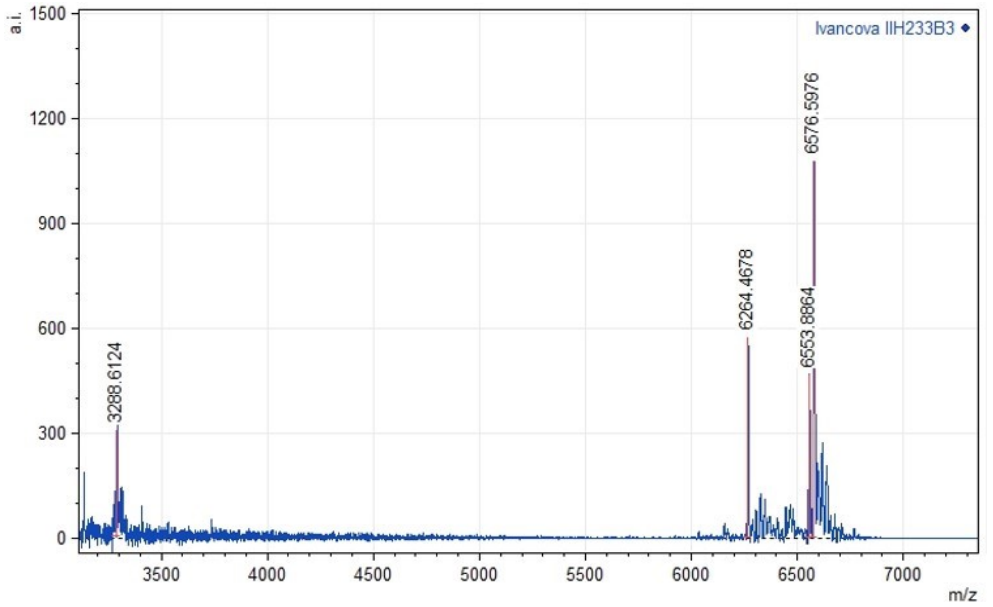


Figure A2. MALDI-TOF MS spectrum of **ON_C^{ESQ3pept}**, calculated for [M + H]: 6576.5 Da; found 6576.6 Da; the peak at *m/z* = 6553.9 can be assigned to the biotinylated template and the peak at *m/z* = 6264.5 can be assigned to the hydrolysed **ON_C^{ESQ}** [- CH_2CH_3 - H^+] (**Table 4**).

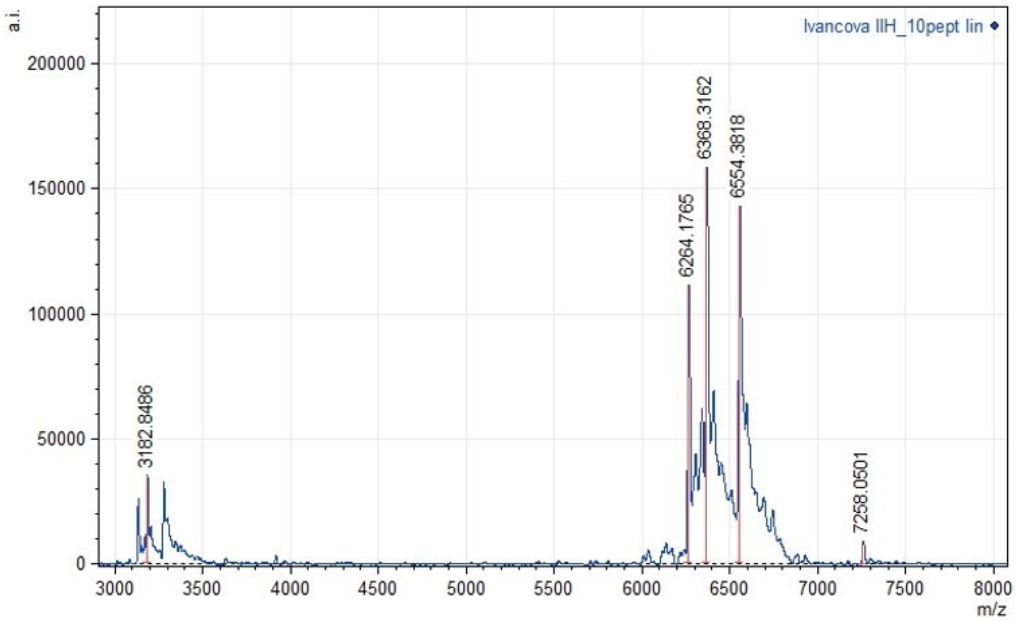


Figure A3. MALDI-TOF MS spectrum of **ON_C^{ESQ10pept}**, calculated for $[M + H]^+$: 7257.8 Da; found 7258.0 Da; the peak at *m/z* = 6554.4 can be assigned to the biotinylated template, the peak at *m/z* = 6264.2 can be assigned to the hydrolysed **ON_C^{ESQ}** [- CH_2CH_3 - H^+] and the peak at 6368.3 Da can be assigned to the product of decomposition of peptide part of the conjugate **ON_C^{ESQ10pept}** in MS (**Table 4**).

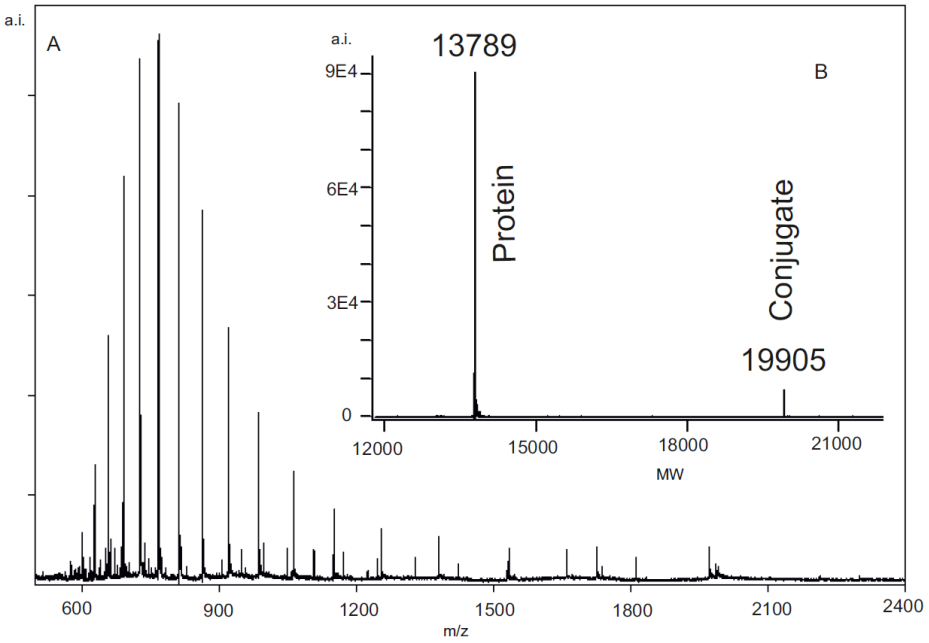


Figure A4. MS spectrum of chromatographic peak of histone H2B (M = 13789) cross-linked to **DNA_C^{ESQ}** (M = 6293.3 for single strand DNA with one modification (**ON_C^{ESQ}**); see **Table 4**). A) combined raw spectrum B) deconvoluted spectrum by MaxEnt1. The signal at 19905 corresponds to **ON_C^{ESQH2B}** conjugate (see **Table 6**).

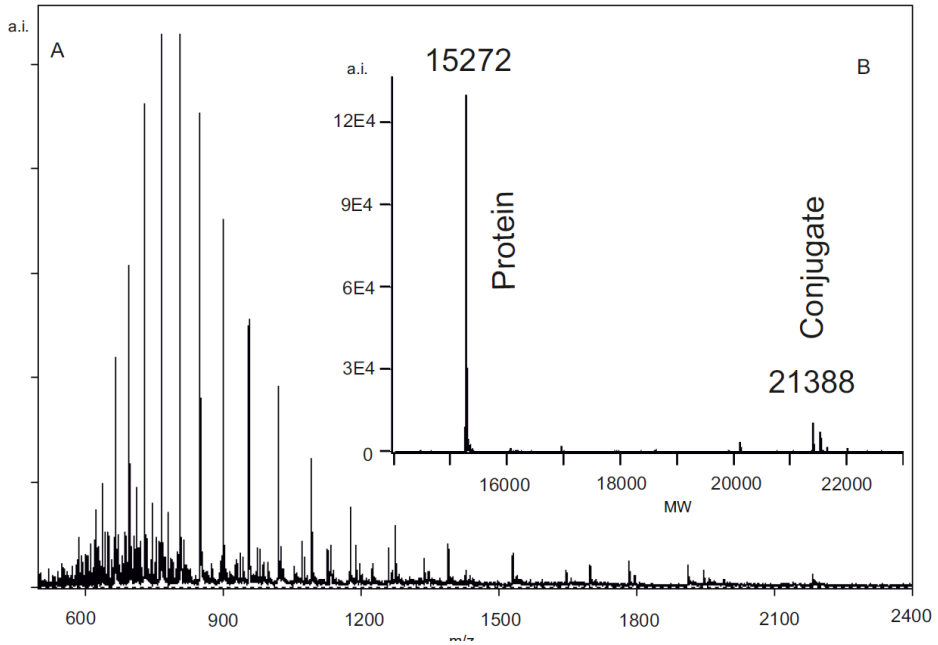


Figure A5. MS spectrum of chromatographic peak of histone H3.1 (M = 15273) cross-linked to **DNA_C^{ESQ}** (M = 6293.3 for single strand DNA with one modification (**ON_C^{ESQ}**); see **Table 4**). A) combined raw spectrum B) deconvoluted spectrum by MaxEnt1. The signal at 21388 corresponds to **ON_C$^{ESQH3.1}$** conjugate (see **Table 6**).

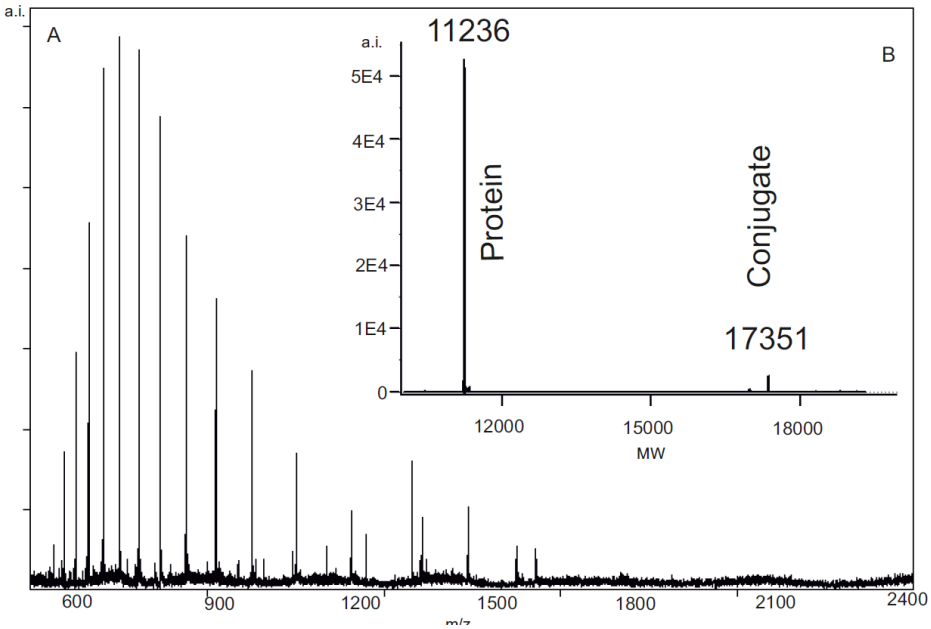


Figure A6. MS spectrum of chromatographic peak of histone H4 (M = 11236) cross-linked to **DNA_C^{ESQ}** (M = 6293.3 for single strand with one modification (**ON_C^{ESQ}**); see **Table 4**). A) combined raw spectrum B) deconvoluted spectrum by MaxEnt1. The signal at 17351 corresponds to **ON_C^{ESQH4}** conjugate (see **Table 6**).

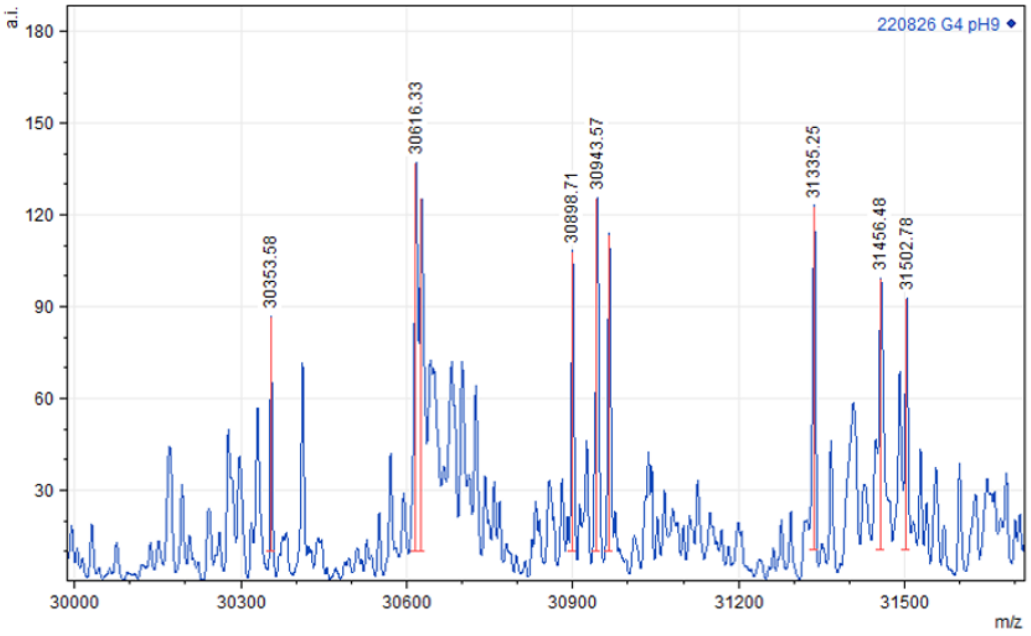


Figure A7. Spectrum of intact mass measurement of **RNA_C1g^{ESQ_SSB}**. M (calc.) = 30920.52 Da, M (found) = 30943.57 $[M + Na]^+$. Peak at 30616.33 Da is $[M - UMP]^+$ (**Table 10**).

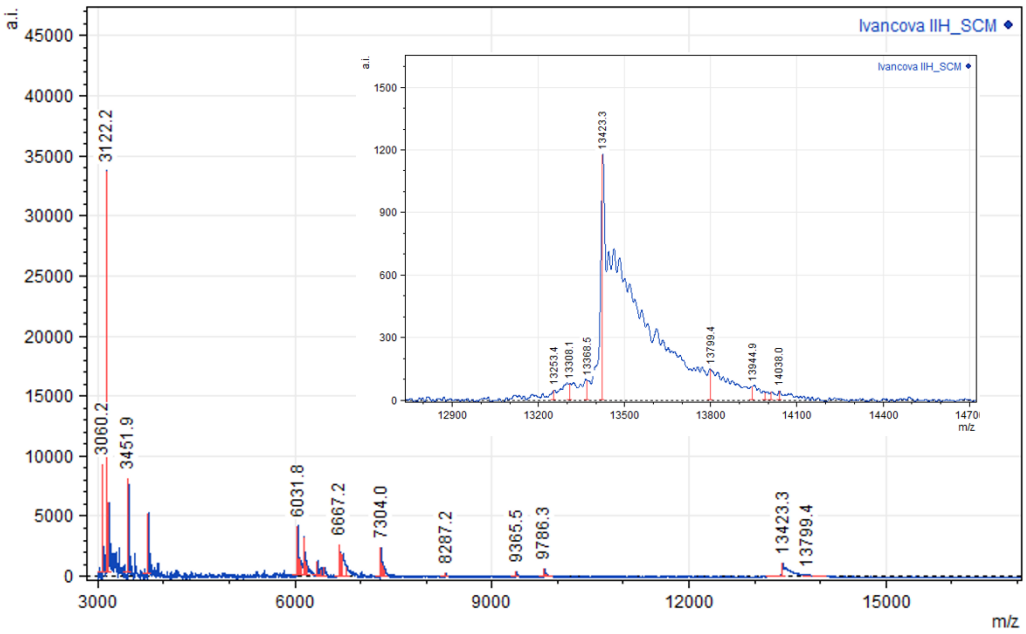


Figure A8. MALDI-TOF MS spectrum of **RNA_C^{ESQ_SC1}**, M (calc.) = 13451.1 Da; M (found) = 13423.3 Da of hydrolysed **RNA_C^{ESQ_SC1}** $[M - CH_2CH_3 - H]$ (**Table 14**).

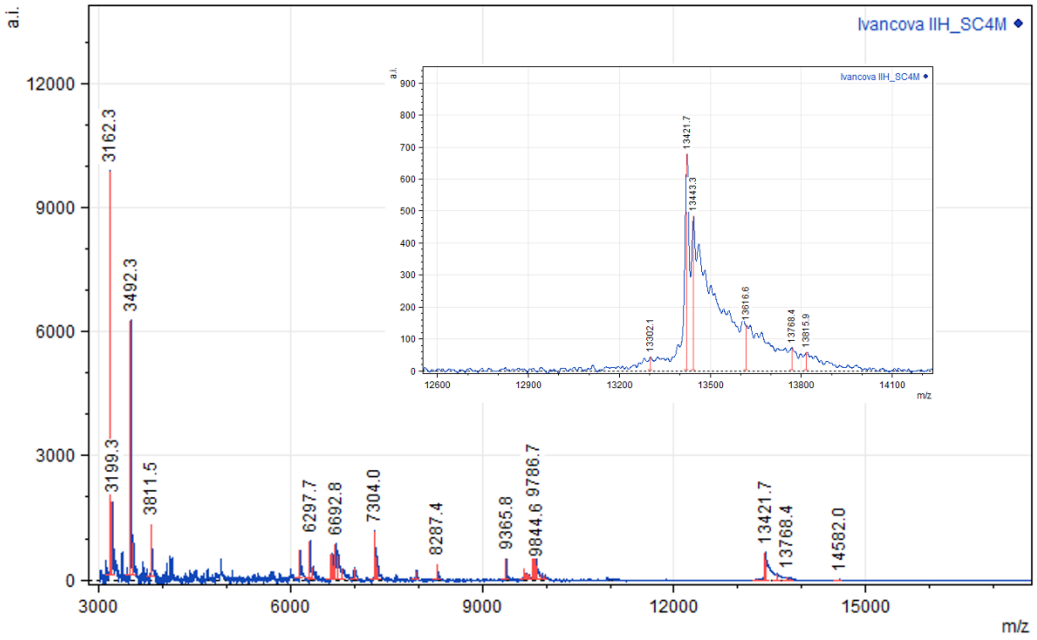


Figure A9. MALDI-TOF MS spectrum of **RNA_C^{ESQ_SC4}**, M (calc.) = 13451.1 Da; M (found) = 13421.7 Da of hydrolysed **RNA_C^{ESQ_SC4}** [M - CH_2CH_3] (**Table 14**).

JEV NS5 protein sequence

```
        10         20         30         40         50         60
GRPGGRTLGE QWKEKLNAMS REEFFKYRRE AIIEVDRTEA RRARRENNIV GGHPVSRGSA

        70         80         90        100        110        120
KLRWLVEKGF VSPIGKVIDL GCGRGGWSYY AATLKKVQEV RGYTKGGAGH EEPMLMQSYG

       130        140        150        160        170        180
RNLVSLKSGV DVFYKPSEPS DTLFCDIGES SPSPEVEEQR TLRVLEMTSD WLHRGPREFC

       190        200        210        220        230        240
IKVLCPYMPK VIEKMEVLQR RFGGGLVRLP LSRNSNHEMY WVSGAAGNVV HAVNMTSQVL

       250        260        269        280        290        300
LGRMDRTVWR GPKYEEDVNL GSGTRAVGK*G EVHSNQEKIK KRIQKLKEEF ATTWHKDPEH

       310        320        330        340        350        360
PYRTWTYHGS YEVKATGSAS SLVNGVVKLM SKPWDAIANV TTMAMTDTTP FGQQRVFKEK

       370        380        390        400        410        420
VDTKAPEPPA GAKEVLNETT NWLWAHLSRE KRPRLCTKEE FIKKVNSNAA LGAVFAEQNQ

       430        440        450        460        470        480
WSTAREAVDD PRFWEMVDEE RENHLRGECH TCIYNMMGKR EKKPGEFGKA KGSRAIWFMW

       490        500        510        520        530        540
LGARYLEFEA LGFLNEDHWL SRENSGGGVE GSGVQKLGYI LRDIAGKQGG KMYADDTAGW

       550        560        570        580        590        600
DTRITRTDLE NEAKVLELLD GEHRMLARAI IELTYRHKVV KVMRPAAEGK TVMDVISRED

       610        620        630        640        650        660
QRGSGQVVTY ALNTFTNIAV QLVRLMEAEG VIGPQHLEQL PRKTKIAVRT WLFENGEERV

       670        680        690        700        710        720
TRMAISGDDC VVKPLDDRFA TALHFLNAMS KVRKDIQEWK PSHGWHDWQQ VPFCSNHFQE

       730        740        750        760        770        780
IVMKDGRSIV VPCRGQDELI GRARISPGAG WNVKDTACLA KAYAQMWLLL YFHRRDLRLM

       790        800        810        820        830        840
ANAICSAVPV DWVPTGRTSW SIHSKGEWMT TEDMLQVWNR VWIEENEWMM DKTPITSWTD

       850        860        870        880        890        900
VPYVGKREDI WCGSLIGTRS RATWAENIYA AINQVRAVIG KENYVDYMTS LRRYEDVLIQ

EDRVI
```

Figure A10. Sequence coverage (91%) of the JEV NS5 proteomic analysis from sample 1. The identified parts of protein are highlighted in green, peptide with RNA modification is underlined and in red. K* = $\mathbf{C}^{\mathbf{ESQ}}$ modified lysine K269.

JEV NS5 protein sequence

```
        10         20         30         40         50         60
GRPGGRTLGE QWKEKLNAMS REEFFKYRRE AIIEVDRTEA RRARRENNIV GGHPVSRGSA

        70         80         90        100        110        120
KLRWLVEKGF VSPIGKVIDL GCGRGGWSYY AATLKKVQEV RGYTKGGAGH EEPMLMQSYG

       130        140        150        160        170        180
RNLVSLKSGV DVFYKPSEPS DTLFCDIGES SPSPEVEEQR TLRVLEMTSD WLHRGPREFC

       190        200        210        220        230        240
IKVLCPYMPK VIEKMEVLQR RFGGGLVRLP LSRNSNHEMY WVSGAAGNVV HAVNMTSQVL

       250        260        270        280        290        300
LGRMDRTVWR GPKYEEDVNL GSGTRAVGKG EVHSNQEKIK KRIQKLKEEF ATTWHKDPEH

       310        320        330        340        350        360
PYRTWTYHGS YEVKATGSAS SLVNGVVKLM SKPWDAIANV TTMAMTDTTP FGQQRVFKEK

       370        380        390        400        410        420
VDTKAPEPPA GAKEVLNETT NWLWAHLSRE KRPRLCTKEE FIKKVNSNAA LGAVFAEQNQ

       430        440        450        459 462/3                  480
WSTAREAVDD PRFWEMVDEE RENHLRGECH TCIYNMMGKR EK*K*PGEFGKA KGSRAIWFMW

       490        500        510        520        530        540
LGARYLEFEA LGFLNEDHWL SRENSGGGVE GSGVQKLGYI LRDIAGKQGG KMYADDTAGW

       550        560        570        580        590        600
DTRITRTDLE NEAKVLELLD GEHRMLARAI IELTYRHKVV KVMRPAAEGK TVMDVISRED

       610        620        630        640        650        660
QRGSGQVVTY ALNTFTNIAV QLVRLMEAEG VIGPQHLEQL PRKTKIAVRT WLFENGEERV

       670        680        690        700        710        720
TRMAISGDDC VVKPLDDRFA TALHFLNAMS KVRKDIQEWK PSHGWHDWQQ VPFCSNHFQE

       730        740        750        760        770        780
IVMKDGRSIV VPCRGQDELI GRARISPGAG WNVKDTACLA KAYAQMWLLL YFHRRDLRLM

       790        800        810        820        830        840
ANAICSAVPV DWVPTGRTSW SIHSKGEWMT TEDMLQVWNR VWIEENEWMM DKTPITSWTD

       850        860        870        880        890        900
VPYVGKREDI WCGSLIGTRS RATWAENIYA AINQVRAVIG KENYVDYMTS LRRYEDVLIQ


EDRVI
```

Figure A11. Sequence coverage (88%) of the JEV NS5 proteomic analysis from sample 2. The identified parts of protein are highlighted in green, peptide with RNA modification is underlined and in red. K* = $\mathbf{C^{ESQ}}$ modified lysine K462/463.

7. List of publications of the author

1) Krömer, M.; Brunderová, M.; Ivancová, I.; Poštová Slavetínská, L.; Hocek, M.: "2-Formyl-dATP as Substrate for Polymerase Synthesis of Reactive DNA bearing an Aldehyde Group in the Minor Groove" *ChemPlusChem* **2020**, *85*, 1164-1170.

2) Ivancová, I.; Leone, D.; Hocek, M.: "Reactive modifications of DNA nucleobases for labelling, bioconjugations, and cross-linking" *Curr. Opin. Chem. Biol.* **2019**, *52*, 136-144.

3) Ivancová, I.; Pohl, R.; Hubálek, M.; Hocek, M.: "Squaramate-Modified Nucleotides and DNA for Specific Cross-Linking with Lysine-Containing Peptides and Proteins" *Angew. Chem. Int. Ed.* **2019**, *58*, 13345-13348.

8. References

1. Belmont, P., Constant, J. F. & Demeunynck, M. Nucleic acid conformation diversity: From structure to function and regulation. *Chem Soc Rev* **30**, 70–81 (2001).

2. Crick, F. & Watson, J. © 1953 Nature Publishing Group. (1953).

3. Matta, C. F., Castillo, N. & Boyd, R. J. Extended weak bonding interactions in DNA: π-stacking (base-base), base-backbone, and backbone-backbone interactions. *Journal of Physical Chemistry B* **110**, 563–578 (2006).

4. Svoboda, P. & Di Cara, A. Hairpin RNA: A secondary structure of primary importance. *Cellular and Molecular Life Sciences* vol. 63 901–918 Preprint at https://doi.org/10.1007/s00018-005-5558-5 (2006).

5. Blackburn, G. M., Egli, M., Gait, M. J. (Michael J.) & Watts, J. K. *Nucleic Acids in Chemistry and Biology*. (2022).

6. Butcher, S. E. & Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* **44**, 1302–1311 (2011).

7. Tan, Z. J. & Chen, S. J. Salt contribution to RNA tertiary structure folding stability. *Biophys J* **101**, 176–187 (2011).

8. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence. *J Mol Biol* **403**, 1–10 (2010).

9. Noller, H. F. RNA structure: Reading the ribosome. *Science* vol. 309 1508–1514 Preprint at https://doi.org/10.1126/science.1111771 (2005).

10. Blythe, A. J., Fox, A. H. & Bond, C. S. The ins and outs of lncRNA structure: How, why and what comes next? *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**, 46–58 (2016).

11. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology* vol. 22 96–118 Preprint at https://doi.org/10.1038/s41580-020-00315-9 (2021).

12. Jones, C. P. & Ferré-D'Amaré, A. R. Long-Range Interactions in Riboswitch Control of Gene Expression. *Annual Review of Biophysics* vol. 46 455–481 Preprint at https://doi.org/10.1146/annurev-biophys-070816-034042 (2017).

13. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins. *Nature Reviews Genetics* vol. 8 776–790 Preprint at https://doi.org/10.1038/nrg2172 (2007).

14. Famulok, M. & Jenne, A. Catalysis Based on Nucleic Acid Structures. in 101–131 (Springer, Berlin, Heidelberg, 1999). doi:10.1007/3-540-48990-8_4.

15. Ma, L. & Liu, J. Catalytic Nucleic Acids: Biochemistry, Chemical Biology, Biosensors, and Nanotechnology. *iScience* vol. 23 100815 Preprint at https://doi.org/10.1016/j.isci.2019.100815 (2020).

16. Hollenstein, M. DNA catalysis: The chemical repertoire of DNAzymes. *Molecules* vol. 20 20777–20804 Preprint at https://doi.org/10.3390/molecules201119730 (2015).

17. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).

18. Zhang, Y. Z., Wu, W. C., Shi, M. & Holmes, E. C. The diversity, evolution and origins of vertebrate RNA viruses. *Current Opinion in Virology* vol. 31 9–16 Preprint at https://doi.org/10.1016/j.coviro.2018.07.017 (2018).

19. Esteller, M. Non-coding RNAs in human disease. *Nature Reviews Genetics* vol. 12 861–874 Preprint at https://doi.org/10.1038/nrg3074 (2011).

20. Sun, B. *et al.* DNA methylation perspectives in the pathogenesis of autoimmune diseases. *Clinical Immunology* vol. 164 21–27 Preprint at https://doi.org/10.1016/j.clim.2016.01.011 (2016).

21. Davegårdh, C., García-Calzón, S., Bacos, K. & Ling, C. DNA methylation in the pathogenesis of type 2 diabetes in humans. *Molecular Metabolism* vol. 14 12–25 Preprint at https://doi.org/10.1016/j.molmet.2018.01.022 (2018).

22. Teller, C. & Willner, I. Functional nucleic acid nanostructures and DNA machines. *Current Opinion in Biotechnology* vol. 21 376–391 Preprint at https://doi.org/10.1016/j.copbio.2010.06.001 (2010).

23. Guo, P. The emerging field of RNA nanotechnology. *Nature Nanotechnology* vol. 5 833–842 Preprint at https://doi.org/10.1038/nnano.2010.231 (2010).

24. Bennett, C. F., Krainer, A. R. & Cleveland, D. W. Antisense Oligonucleotide Therapies for Neurodegenerative Diseases. *Annu Rev Neurosci* **42**, 385–406 (2019).

25. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **17**, 547–558 (2018).

26. Patil, S. D., Rhodes, D. G. & Burgess, D. J. DNA-based therapeutics and DNA delivery systems: A comprehensive review. *AAPS Journal* vol. 7 E61 Preprint at https://doi.org/10.1208/aapsj070109 (2005).

27. Taslem Mourosi, J., Awe, A., Jain, S. & Batra, H. Nucleic Acid Vaccine Platform for DENGUE and ZIKA Flaviviruses. *Vaccines (Basel)* **10**, 834 (2022).

28. Khan, K. H. DNA vaccines: Roles against diseases. *GERMS* vol. 3 26–35 Preprint at https://doi.org/10.11599/germs.2013.1034 (2013).

29. Lundstrom, K. RNA Viruses as Tools in Gene Therapy and Vaccine Development. *Genes (Basel)* **10**, 189 (2019).

30. Huang, Q., Zeng, J. & Yan, J. COVID-19 mRNA vaccines. *Journal of Genetics and Genomics* vol. 48 107–114 Preprint at https://doi.org/10.1016/j.jgg.2021.02.006 (2021).

31. Dunn, J. & Grider, M. H. *Physiology, Adenosine Triphosphate (ATP)*. *StatPearls* (StatPearls Publishing, 2020).

32. Chang, Y. F. & Carman, G. M. CTP synthetase and its role in phospholipid synthesis in the yeast Saccharomyces cerevisiae. *Progress in Lipid Research* vol. 47 333–339 Preprint at https://doi.org/10.1016/j.plipres.2008.03.004 (2008).

33. Gitzelmann, R. Formation of galactose-1-phosphate from uridine diphosphate galactose in erythrocytes from patients with galactosemia. *Pediatr Res* **3**, 279–286 (1969).

34. Ives, H. E. GTP binding proteins and growth factor signal transduction. *Cellular Signalling* vol. 3 491–499 Preprint at https://doi.org/10.1016/0898-6568(91)90026-Q (1991).

35. Ichikawa, E. & Kato, K. Sugar-Modified Nucleosides in Past 10 Years, A Review. *Curr Med Chem* **8**, 385–423 (2012).

36. P. Burke, M., M. Borland, K. & A. Litosh, V. Base-Modified Nucleosides as Chemotherapeutic Agents: Past and Future.

37. *Modified Nucleic Acids*. vol. 31 (Springer International Publishing, 2016).

38. Hottin, A. & Marx, A. Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Accounts of Chemical Research* vol. 49 418–427 Preprint at https://doi.org/10.1021/acs.accounts.5b00544 (2016).

39. Kielkowski, P., Fanfrlík, J. & Hocek, M. 7-Aryl-7-deazaadenine 2′-Deoxyribonucleoside Triphosphates (dNTPs): Better Substrates for DNA Polymerases than dATP in Competitive Incorporations. *Angewandte Chemie International Edition* **53**, 7552–7555 (2014).

40. Hocek, M. & Fojta, M. Cross-coupling reactions of nucleoside triphosphates followed by polymerase incorporation. Construction and applications of base-functionalized nucleic acids. *Org Biomol Chem* **6**, 2233–2241 (2008).

41. Yoshikawa, M., Kato, T. & Takenishi, T. A novel method for phosphorylation of nucleosides to 5′-nucleotides. *Tetrahedron Lett* **8**, 5065–5068 (1967).

42. Yoshikawa, M., Kato, T. & Takenishi, T. Studies of Phosphorylation. III. Selective Phosphorylation of Unprotected Nucleosides. *Bull Chem Soc Jpn* **42**, 3505–3508 (1969).

43. Burgess, K. & Cook, D. Syntheses of nucleoside triphosphates. *Chem Rev* **100**, 2047–2059 (2000).

44. Ludwig, J. & Eckstein, F. Rapid and Efficient Synthesis of Nucleoside 5'-O-(1-thiotriphosphates), 5'-Triphosphates and 2',3'-Cyclophosphorothioates Using 2-Chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *Journal of Organic Chemistry* **54**, 631–635 (1989).

45. Wu, W., Meyers, C. L. F. & Borch, R. F. A novel method for the preparation of nucleoside triphosphates from activated nucleoside phosphoramidates. *Org Lett* **6**, 2257–2260 (2004).

46. Shaughnessy, K. H. Palladium-catalyzed modification of unprotected nucleosides, nucleotides, and oligonucleotides. *Molecules* **20**, 9419–9454 (2015).

47. Hocek, M. Synthesis of base-modified 2′-deoxyribonucleoside triphosphates and their use in enzymatic synthesis of modified DNA for applications in bioanalysis and chemical biology. *Journal of Organic Chemistry* **79**, 9914–9921 (2014).

48. Čapek, P. *et al.* An Efficient Method for the Construction of Functionalized DNA Bearing Amino Acid Groups through Cross-Coupling Reactions of Nucleoside Triphosphates Followed by Primer Extension or PCR. *Chemistry – A European Journal* **13**, 6196–6203 (2007).

49. Hervé, G., Sartori, G., Enderlin, G., MacKenzie, G. & Len, C. Palladium-catalyzed Suzuki reaction in aqueous solvents applied to unprotected nucleosides and nucleotides. *RSC Adv* **4**, 18558–18594 (2014).

50. Biteau, N., Hervin, V., Roy, V. & Agrofoglio, L. A. Suzuki-Miyaura Cross-Coupling as a Synthetic Tool for Nucleoside and Nucleotide Modification. *Palladium-Catalyzed Modification of Nucleosides, Nucleotides and Oligonucleotides* 37–74 (2018) doi:10.1016/B978-0-12-811292-2.00003-9.

51. Hervé, G. & Len, C. Heck and Sonogashira couplings in aqueous media – application to unprotected nucleosides and nucleotides. *Sustainable Chemical Processes* **3**, 3 (2015).

52. Dadová, J. *et al.* Aqueous heck cross-coupling preparation of acrylate-modified nucleotides and nucleoside triphosphates for polymerase synthesis

of acrylate-labeled DNA. *Journal of Organic Chemistry* **78**, 9627–9637 (2013).

53. Langer, P. R., Waldrop, A. A. & Ward, D. C. Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes. *Proc Natl Acad Sci U S A* **78**, 6633–6637 (1981).

54. Bartlett, J. M. S. & Stirling, D. A Short History of the Polymerase Chain Reaction. in *PCR Protocols* 3–6 (Humana Press, 2003). doi:10.1385/1-59259-384-4:3.

55. Little, J. W. An Exonuclease Induced by Bacteriophage λ. *Journal of Biological Chemistry* **242**, 679–686 (1967).

56. Hultman, T., Stahl, S., Homes, E. & Uhlén, M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Res* **17**, 4937–4946 (1989).

57. Ménová, P., Raindlová, V. & Hocek, M. Scope and limitations of the nicking enzyme amplification reaction for the synthesis of base-modified oligonucleotides and primers for PCR. *Bioconjug Chem* **24**, 1081–1093 (2013).

58. Hollenstein, M. Generation of long, fully modified, and serum-resistant oligonucleotides by rolling circle amplification. *Org Biomol Chem* **13**, 9820–9824 (2015).

59. Cahová, H. *et al.* Synthesis of 8-bromo-, 8-methyl- and 8-phenyl-dATP and their polymerase incorporation into DNA. *Org Biomol Chem* **6**, 3657–3660 (2008).

60. Jakubovska, J., Tauraitė, D. & Meškys, R. Transient N4-Acyl-DNA Protection against Cleavage by Restriction Endonucleases. *ChemBioChem* **20**, 2504–2512 (2019).

61. Cherkasov, D., Biet, T., Bäuml, E., Traut, W. & Lohoff, M. New nucleotide analogues with enhanced signal properties. *Bioconjug Chem* **21**, 122–129 (2010).

62. Matyašovský, J., Perlíková, P., Malnuit, V., Pohl, R. & Hocek, M. 2-Substituted dATP Derivatives as Building Blocks for Polymerase-

Catalyzed Synthesis of DNA Modified in the Minor Groove. *Angewandte Chemie International Edition* **55**, 15856–15859 (2016).

63. Lodish, H. *et al. Molecular Cell Biology (Lodish, Molecular Cell Biology)*. (2007).

64. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783–8798 (1987).

65. Yin, Y. W. & Steitz, T. A. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science (1979)* **298**, 1387–1395 (2002).

66. Kuzmine, I., Gottlieb, P. A. & Martin, C. T. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *Journal of Biological Chemistry* **278**, 2819–2823 (2003).

67. Kennedy, W. P., Momand, J. R. & Yin, Y. W. Mechanism for De Novo RNA Synthesis and Initiating Nucleotide Specificity by T7 RNA Polymerase. *J Mol Biol* **370**, 256–268 (2007).

68. Padmanabhan, R. & Miller, D. The Length of Promoter Sequence Affects The De Novo Initiation By T7 RNA Polymerase in Vitro: New Insights Into The Evolution of Promoters For Single Subunit RNA Polymerases. *bioRxiv* 619395 (2019) doi:10.1101/619395.

69. Beckert, B. & Masquida, B. Synthesis of RNA by In Vitro Transcription BT - RNA: Methods and Protocols. *Methods Mol Biol* **703**, 29–41 (2011).

70. Gößringer, M. *et al.* Enzymatic RNA Synthesis Using Bacteriophage T7 RNA Polymerase. *Handbook of RNA Biochemistry: Second, Completely Revised and Enlarged Edition* **1–2**, 1–28 (2014).

71. Vaught, J. D., Dewey, T. & Eaton, B. E. T7 RNA polymerase transcription with 5-position modified UTP derivatives. *J Am Chem Soc* **126**, 11231–11237 (2004).

72. Paredes, E. & Das, S. R. Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **12**, 125–131 (2011).

73. Rao, H., Tanpure, A. A., Sawant, A. A. & Srivatsan, S. G. Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nat Protoc* **7**, 1097–1112 (2012).

74. George, J. T. & Srivatsan, S. G. Vinyluridine as a Versatile Chemoselective Handle for the Post-transcriptional Chemical Functionalization of RNA. *Bioconjug Chem* **28**, 1529–1536 (2017).

75. Smith, C. C., Hollenstein, M. & Leumann, C. J. The synthesis and application of a diazirine-modified uridine analogue for investigating RNA-protein interactions. *RSC Adv* **4**, 48228–48235 (2014).

76. Tanpure, A. A. & Srivatsan, S. G. Synthesis, Photophysical Properties and Incorporation of a Highly Emissive and Environment-Sensitive Uridine Analogue Based on the Lucifer Chromophore. *ChemBioChem* **15**, 1309–1316 (2014).

77. Pawar, M. G. & Srivatsan, S. G. Synthesis, photophysical characterization, and enzymatic incorporation of a microenvironment-sensitive fluorescent uridine analog. *Org Lett* **13**, 1114–1117 (2011).

78. Milisavljevič, N., Perlíková, P., Pohl, R. & Hocek, M. Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org Biomol Chem* **16**, 5800–5807 (2018).

79. Wong, S. S. & Jameson, D. M. *Chemistry of protein and nucleic acid cross-linking and conjugation, second edition*. *Chemistry of Protein and Nucleic Acid Cross-Linking and Conjugation, Second Edition* (CRC Press, 2012). doi:10.1201/B11175.

80. Schneider, M., Belsom, A. & Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem Sci* **43**, 157 (2018).

81. Kobsa, S. & Saltzman, W. M. Bioengineering Approaches to Controlled Protein Delivery. *Pediatric Research 2008 63:5* **63**, 513–519 (2008).

82. Beck, A., Goetsch, L., Dumontet, C. & Corvaïa, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat Rev Drug Discov* **16**, 315–337 (2017).

83. Wang, R. E. *et al.* An immunosuppressive antibody-drug conjugate. *J Am Chem Soc* **137**, 3229–3232 (2015).

84. Greineder, C. F. *et al.* Site-Specific Modification of Single-Chain Antibody Fragments for Bioconjugation and Vascular Immunotargeting. *Bioconjug Chem* **29**, 56–66 (2018).

85. Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L. & Bernardes, G. J. L. Contemporary approaches to site-selective protein modification. *Nature Reviews Chemistry* vol. 3 147–171 Preprint at https://doi.org/10.1038/s41570-019-0079-1 (2019).

86. Koniev, O. & Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem Soc Rev* **44**, 5495–5551 (2015).

87. Maruyama, K. & Kanai, M. Synthetic methodology-driven chemical protein modifications. *Chem Lett* **48**, 1421–1432 (2019).

88. Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat Chem Biol* **13**, 697–705 (2017).

89. Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat Chem* **10**, 205–211 (2018).

90. Murza, A. *et al.* C-Terminal modifications of apelin-13 significantly change ligand binding, receptor signaling, and hypotensive action. *J Med Chem* **58**, 2431–2440 (2015).

91. Willwacher, J., Raj, R., Mohammed, S. & Davis, B. G. Selective Metal-Site-Guided Arylation of Proteins. *J Am Chem Soc* **138**, 8678–8681 (2016).

92. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nature Chemistry 2017 9:12* **9**, 1181–1190 (2017).

93. Gray, G. R. The direct coupling of oligosaccharides to proteins and derivatized gels. *Arch Biochem Biophys* **163**, 426–428 (1974).

94. McFarland, J. M. & Francis, M. B. Reductive alkylation of proteins using iridium catalyzed transfer hydrogenation. *J Am Chem Soc* **127**, 13490–13491 (2005).

95. Fujiki, K. & Tanaka, K. Biomolecular labeling based on lysine-clickable 6π-azaelectrocyclization toward innovative cancer theranostics. *Bioorg Med Chem* **42**, (2021).

96. Gilis, D., Massar, S., Cerf, N. J. & Rooman, M. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biology 2001 2:11* **2**, 1–12 (2001).

97. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem Sci* **2**, 1666–1676 (2011).

98. Gevaert, K., Van Damme, P., Martens, L. & Vandekerckhove, J. Diagonal reverse-phase chromatography applications in peptide-centric proteomics: Ahead of catalogue-omics? *Anal Biochem* **345**, 18–29 (2005).

99. Foettinger, A., Leitner, A. & Lindner, W. Selective enrichment of tryptophan-containing peptides from protein digests employing a reversible derivatization with malondialdehyde and solid-phase capture on hydrazide beads. *J Proteome Res* **6**, 3827–3834 (2007).

100. Antos, J. M. & Francis, M. B. Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J Am Chem Soc* **126**, 10256–10257 (2004).

101. Seki, Y. *et al.* Transition Metal-Free Tryptophan-Selective Bioconjugation of Proteins. *J Am Chem Soc* **138**, 10798–10801 (2016).

102. Imiołek, M. *et al.* Selective radical trifluoromethylation of native residues in proteins. *J Am Chem Soc* **140**, 1568–1571 (2018).

103. Joshi, N. S., Whitaker, L. R. & Francis, M. B. A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J Am Chem Soc* **126**, 15942–15943 (2004).

104. Hooker, J. M., Kovacs, E. W. & Francis, M. B. Interior Surface Modification of Bacteriophage MS2. *J Am Chem Soc* **126**, 3718–3719 (2004).

105. Tilley, S. D. & Francis, M. B. Tyrosine-selective protein alkylation using pi-allylpalladium complexes. *J Am Chem Soc* **128**, 1080–1081 (2006).

106. Ban, H., Gavrilyuk, J. & Barbas, C. F. Tyrosine bioconjugation through aqueous ene-type reactions: A click-like reaction for tyrosine. *J Am Chem Soc* **132**, 1523–1525 (2010).

107. Stipanuk, M. H. SULFUR AMINO ACID METABOLISM: Pathways for Production and Removal of Homocysteine and Cysteine. *http://dx.doi.org/10.1146/annurev.nutr.24.012003.132418* **24**, 539–577 (2004).

108. Zang, J., Chen, Y., Zhu, W. & Lin, S. Chemoselective Methionine Bioconjugation on a Polypeptide, Protein, and Proteome. *Biochemistry* **59**, 132–138 (2020).

109. Takahashi, K. The Reaction of Phenylglyoxal with Arginine Residues in Proteins. *Journal of Biological Chemistry* **243**, 6171–6179 (1968).

110. Yankeelov, J. A. [52] Modification of arginine by diketones. in *Methods in Enzymology* vol. 25 566–579 (Academic Press, 1972).

111. Toi, K., Bynum, E., Norris, E. & Itano, H. A. Studies on the Chemical Modification of Arginine. *Journal of Biological Chemistry* **242**, 1036–1043 (1967).

112. Gauthier, M. A. & Klok, H. A. Arginine-specific modification of proteins with polyethylene glycol. *Biomacromolecules* **12**, 482–493 (2011).

113. Bouhlel, M., Lambert, M. & David-Cordonnier, M.-H. Targeting Transcription Factor Binding to DNA by Competing with DNA Binders as an Approach for Controlling Gene Expression. *Curr Top Med Chem* **15**, 1323–1358 (2015).

114. Nevinsky, G. A. Structural, thermodynamic, and kinetic basis for the activities of some nucleic acid repair enzymes. *J Mol Recognit* **24**, 656–677 (2011).

115. Gujar, H., Weisenberger, D. J. & Liang, G. The Roles of Human DNA Methyltransferases and Their Isoforms in Shaping the Epigenome. *Genes 2019, Vol. 10, Page 172* **10**, 172 (2019).

116. Sternburg, E. L. & Karginov, F. V. Global Approaches in Studying RNA-Binding Protein Interaction Networks. *Trends Biochem Sci* **45**, 593–603 (2020).

117. Tretyakova, N. Y., Groehler, A. & Ji, S. DNA-Protein Cross-links: Formation, Structural Identities, and Biological Outcomes. *Acc Chem Res* **48**, 1631 (2015).

118. Zhang, H., Xiong, Y. & Chen, J. DNA–protein cross-link repair: what do we know now? *Cell & Bioscience 2020 10:1* **10**, 1–10 (2020).

119. Trakselis, M. A., Alley, S. C. & Ishmael, F. T. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug Chem* **16**, 741–750 (2005).

120. Wimmer, E. Genome-linked proteins of viruses. *Cell* **28**, 199–201 (1982).

121. Drygin, Y. F. Natural covalent complexes of nucleic acids and proteins: some comments on practice and theory on the path from well-known complexes to new ones. *Nucleic Acids Res* **26**, 4791–4796 (1998).

122. Baker, N. M., Rajan, R. & Mondragńn, A. Structural studies of type I topoisomerases. *Nucleic Acids Res* **37**, 693 (2009).

123. Flanegan, J. B., Petterson, R. F., Ambros, V., Hewlett, N. J. & Baltimore, D. Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus. *Proceedings of the National Academy of Sciences* **74**, 961–965 (1977).

124. Kealey, J. T., Gu, X. & Santi, D. v. Enzymatic mechanism of tRNA (m5U54)methyltransferase. *Biochimie* **76**, 1133–1142 (1994).

125. Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: biological functions and applications. *Nucleic Acids Res* **44**, 7511–7526 (2016).

126. Gladkova, E. D. *et al.* The First Berberine-Based Inhibitors of Tyrosyl-DNA Phosphodiesterase 1 (Tdp1), an Important DNA Repair Enzyme. *International Journal of Molecular Sciences 2020, Vol. 21, Page 7162* **21**, 7162 (2020).

127. Gu, X., Liu, Y. & Santi, D. v. The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc Natl Acad Sci U S A* **96**, 14270–14275 (1999).

128. Verdine, G. L. & Norman, D. P. G. Covalent Trapping of Protein-DNA Complexes. *https://doi.org/10.1146/annurev.biochem.72.121801.161447* **72**, 337–366 (2003).

129. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).

130. Brooks, P. J. & Zakhari, S. Acetaldehyde and the genome: beyond nuclear DNA adducts and carcinogenesis. *Environ Mol Mutagen* **55**, 77–91 (2014).

131. Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes *. *Journal of Biological Chemistry* **290**, 26404–26411 (2015).

132. Feldman, M. Y. Reactions of Nucleic Acids and NucleoDroteins with Formaldehyde. *Prog Nucleic Acid Res Mol Biol* **13**, 1–49 (1973).

133. Orlando, V., Strutt, H. & Paro, R. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* **11**, 205–214 (1997).

134. Niranjanakumari, S., Lasda, E., Brazas, R. & Garcia-Blanco, M. A. Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. *Methods* **26**, 182–190 (2002).

135. Sutherland, B. W., Toews, J. & Kast, J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions. *Journal of Mass Spectrometry* **43**, 699–715 (2008).

136. Klockenbusch, C., O'Hara, J. E. & Kast, J. Advancing formaldehyde cross-linking towards quantitative proteomic applications. *Anal Bioanal Chem* **404**, 1057–1067 (2012).

137. Tayri-Wilk, T. *et al.* Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins. *Nature Communications 2020 11:1* **11**, 1–9 (2020).

138. Kim, Y. C. *et al.* Functional analysis of RNA binding by the hepatitis C virus RNA-dependent RNA polymerase. *J Biol Chem* **280**, 38011–38019 (2005).

139. Patton, R. D. *et al.* Chemical crosslinking enhances RNA immunoprecipitation for efficient identification of binding sites of proteins that photo-crosslink poorly with RNA. *RNA* **26**, 1216–1233 (2020).

140. Alexander, P. & Moroson, H. Cross-linking of Deoxyribonucleic Acid to Protein following Ultra-Violet Irradiation of Different Cells. *Nature 1962 194:4831* **194**, 882–883 (1962).

141. Kantor, G. J. & Hull, D. R. An effect of ultraviolet light on RNA and protein synthesis in nondividing human diploid fibroblasts. *Biophys J* **27**, 359–370 (1979).

142. Willis, M. C., Lecuyer, K. A., Meisenheimer, K. M., Uhlenbeck, O. C. & Koch, T. H. An RNA—protein contact determined by 5-bromouridine substitution, photocrosslinking and sequencing. *Nucleic Acids Res* **22**, 4947–4952 (1994).

143. Sontheimer, E. J. Site-specific RNA crosslinking with 4-thiouridine. *Molecular Biology Reports 1994 20:1* **20**, 35–44 (1994).

144. Preston, G. W. & Wilson, A. J. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem Soc Rev* **42**, 3289–3301 (2013).

145. Lercher, L., McGouran, J. F., Kessler, B. M., Schofield, C. J. & Davis, B. G. DNA Modification under Mild Conditions by Suzuki–Miyaura Cross-Coupling for the Generation of Functional Probes. *Angew Chem Int Ed Engl* **52**, 10553 (2013).

146. Smith, C. C., Hollenstein, M. & Leumann, C. J. The synthesis and application of a diazirine-modified uridine analogue for investigating RNA-protein interactions. *RSC Adv* **4**, 48228–48235 (2014).

147. Luo, H. *et al.* Photocatalytic Chemical Crosslinking for Profiling RNA–Protein Interactions in Living Cells. *Angewandte Chemie International Edition* **61**, e202202008 (2022).

148. Wickramaratne, S., Mukherjee, S., Villalta, P. W., Schärer, O. D. & Tretyakova, N. Y. Synthesis of sequence-specific DNA-protein conjugates via a reductive amination strategy. *Bioconjug Chem* **24**, 1496–1506 (2013).

149. Carrette, L. L. G., Gyssels, E., de Laet, N. & Madder, A. Furan oxidation based cross-linking: a new approach for the study and targeting of nucleic acid and protein interactions. *Chemical Communications* **52**, 1539–1554 (2016).

150. Dadová, J. *et al.* Vinylsulfonamide and acrylamide modification of DNA for cross-linking with proteins. *Angewandte Chemie - International Edition* **52**, 10515–10518 (2013).

151. Olszewska, A., Pohl, R., Brázdová, M., Fojta, M. & Hocek, M. Chloroacetamide-Linked Nucleotides and DNA for Cross-Linking with Peptides and Proteins. *Bioconjug Chem* **27**, 2089–2094 (2016).

152. Leone, D. L., Hubálek, M., Pohl, R., Sýkorová, V. & Hocek, M. 1,3-Diketone-Modified Nucleotides and DNA for Cross-Linking with Arginine-Containing Peptides and Proteins. *Angewandte Chemie - International Edition* **60**, 17383–17387 (2021).

153. Leone, D. L. *et al.* Glyoxal-Linked Nucleotides and DNA for Bioconjugations and Crosslinking with Arginine-Containing Peptides and Proteins. *Chemistry - A European Journal* **28**, (2022).

154. Matyašovský, J. & Hocek, M. 2-Substituted 2′-deoxyinosine 5′-triphosphates as substrates for polymerase synthesis of minor-groove-modified DNA and effects on restriction endonuclease cleavage. *Org Biomol Chem* **18**, 255–262 (2020).

155. Raindlová, V., Pohl, R., Šanda, M. & Hocek, M. Direct polymerase synthesis of reactive aldehyde-functionalized DNA and Its conjugation and staining with hydrazines. *Angewandte Chemie - International Edition* **49**, 1064–1066 (2010).

156. Krömer, M., Bártová, K., Raindlová, V. & Hocek, M. Synthesis of Dihydroxyalkynyl and Dihydroxyalkyl Nucleotides as Building Blocks or

Precursors for Introduction of Diol or Aldehyde Groups to DNA for Bioconjugations. *Chemistry - A European Journal* **24**, 11890–11894 (2018).

157. Krömer, M., Brunderová, M., Ivancová, I., Poštová Slavětínská, L. & Hocek, M. 2-Formyl-dATP as Substrate for Polymerase Synthesis of Reactive DNA Bearing an Aldehyde Group in the Minor Groove. *Chempluschem* **85**, 1164–1170 (2020).

158. Tretyakova, N. Y., Groehler, A. & Ji, S. DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Acc Chem Res* **48**, 1631–1644 (2015).

159. Ji, S., Shao, H., Han, Q., Seiler, C. L. & Tretyakova, N. Y. Reversible DNA-Protein Cross-Linking at Epigenetic DNA Marks. *Angewandte Chemie International Edition* **56**, 14130–14134 (2017).

160. Li, F. *et al.* 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *J Am Chem Soc* **139**, 10617–10620 (2017).

161. Marchetti, L. A., Kumawat, L. K., Mao, N., Stephens, J. C. & Elmes, R. B. P. The Versatility of Squaramides: From Supramolecular Chemistry to Chemical Biology. *Chem* **5**, 1398–1485 (2019).

162. Ian Storer, R., Aciro, C. & Jones, L. H. Squaramides: physical properties, synthesis and applications. *Chem Soc Rev* **40**, 2330 (2011).

163. Davis, A. P., Draper, S. M., Dunne, G. & Ashton, P. The N-carbamoyl squaramide dimer: A compact, strongly associated H- bonding motif. *Chemical Communications* 2265–2266 (1999) doi:10.1039/A907179B.

164. Rotger, C. *et al.* Crystallographic and Theoretical Evidence of Anion–π and Hydrogen-Bonding Interactions in a Squaramide–Nitrate Salt. *European J Org Chem* **2008**, 1864–1868 (2008).

165. López, C., Ximenis, M., Orvay, F., Rotger, C. & Costa, A. Supramolecular Hydrogels Based on Minimalist Amphiphilic Squaramide–Squaramates for Controlled Release of Zwitterionic Biomolecules. *Chemistry – A European Journal* **23**, 7590–7594 (2017).

166. Popova, E. A. *et al.* Squaramide-Based Catalysts in Organic Synthesis (A Review). *Russ J Gen Chem* **92**, 287–347 (2022).

167. Gale, P. A., Howe, E. N. W. & Wu, X. Anion Receptor Chemistry. *Chem* **1**, 351–422 (2016).

168. Piña, M. N. *et al.* Selective sensing of competitive anions by non-selective hosts: the case of sulfate and phosphate in water. *New Journal of Chemistry* **32**, 1919–1923 (2008).

169. Elmes, R. B. P., Turner, P. & Jolliffe, K. A. Colorimetric and luminescent sensors for chloride: Hydrogen bonding vs deprotonation. *Org Lett* **15**, 5638–5641 (2013).

170. Marchetti, L. A., Mao, N., Krämer, T., Kitchen, J. A. & Elmes, R. B. P. A long wavelength colourimetric chemosensor for fluoride. *https://doi.org/10.1080/10610278.2018.1461873* **30**, 795–805 (2018).

171. Busschaert, N. *et al.* Squaramides as Potent Transmembrane Anion Transporters. *Angewandte Chemie International Edition* **51**, 4426–4430 (2012).

172. Bao, X. *et al.* Fluorescent squaramides as anion receptors and transmembrane anion transporters. *Chemical Communications* **54**, 1363–1366 (2018).

173. Tantry, S. J. *et al.* Discovery of Imidazo[1,2-a]pyridine Ethers and Squaramides as Selective and Potent Inhibitors of Mycobacterial Adenosine Triphosphate (ATP) Synthesis. *J Med Chem* **60**, 1379–1399 (2017).

174. Ribeiro, C. J. A. *et al.* Novel squaramides with in vitro liver stage antiplasmodial activity. *Bioorg Med Chem* **24**, 1786–1792 (2016).

175. Berney, M. *et al.* Synthesis and evaluation of squaramide and thiosquaramide inhibitors of the DNA repair enzyme SNM1A. *Bioorg Med Chem* **46**, 116369 (2021).

176. Gauthier, M. A. & Klok, H. A. Peptide/protein–polymer conjugates: synthetic strategies and design concepts. *Chemical Communications* 2591–2611 (2008) doi:10.1039/B719689J.

177. Wurm, F., Steinbach, T. & Klok, H.-A. One-pot squaric acid diester mediated aqueous protein conjugation. *Chem Commun (Camb)* **49**, 7815–7 (2013).

178. Sejwal, P., Han, Y., Shah, A. & Luk, Y. Y. Water-driven chemoselective reaction of squarate derivatives with amino acids and peptides. *Org Lett* **9**, 4897–4900 (2007).

179. Hou, S. jie, Saksena, R. & Kováč, P. Preparation of glycoconjugates by dialkyl squarate chemistry revisited. *Carbohydr Res* **343**, 196–210 (2008).

180. Dürr, E. M. *et al.* Squaramide-Based 5'-Phosphate Replacements Bind to the DNA Repair Exonuclease SNM1A. *ChemistrySelect* **3**, 12824–12829 (2018).

181. Sato, K., Seio, K. & Sekine, M. Synthesis and properties of a new oligonucleotide analogue containing an internucleotide squaryl amide linkage. 121–122 (2001) doi:10.1093.

182. Fonvielle, M. *et al.* Electrophilic RNA for Peptidyl-RNA Synthesis and Site-Specific Cross-Linking with tRNA-Binding Enzymes. *Angewandte Chemie - International Edition* **55**, 13553–13557 (2016).

183. Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* **79**, 233 (2010).

184. Nielsen, T. B. *et al.* Peptide-Directed DNA-Templated Protein Labelling for The Assembly of a Pseudo-IgM. *Angewandte Chemie International Edition* **58**, 9068–9072 (2019).

185. Vinkenborg, J. L., Mayer, G. & Famulok, M. Aptamer-Based Affinity Labeling of Proteins. *Angewandte Chemie International Edition* **51**, 9176–9180 (2012).

186. Niemeyer, C. M. Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. *Angew Chem Int Ed Engl* **49**, 1200–1216 (2010).

187. Norris, C. L., Meisenheimer, P. L. & Koch, T. H. Mechanistic studies of the 5-iodouracil chromophore relevant to its use in nucleoprotein photo-cross-linking. *J Am Chem Soc* **118**, 5796–5803 (1996).

188. Li, F. *et al.* 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *J Am Chem Soc* **139**, 10617–10620 (2017).

189. Ji, S., Shao, H., Han, Q., Seiler, C. L. & Tretyakova, N. Y. Reversible DNA–Protein Cross-Linking at Epigenetic DNA Marks. *Angewandte Chemie International Edition* **56**, 14130–14134 (2017).

190. Raindlová, V., Pohl, R. & Hocek, M. Synthesis of Aldehyde-Linked Nucleotides and DNA and Their Bioconjugations with Lysine and Peptides through Reductive Amination. *Chemistry – A European Journal* **18**, 4080–4087 (2012).

191. Hocek, M. Enzymatic Synthesis of Base-Functionalized Nucleic Acids for Sensing, Cross-linking, and Modulation of Protein-DNA Binding and Transcription. *Acc Chem Res* **52**, 1730–1737 (2019).

192. Hashmi, A. S. K. *et al.* Gold catalysis: Non-spirocyclic intermediates in the conversion of furanynes by the formal insertion of an alkyne into an aryl-alkyl C-C single bond. *Chemistry - A European Journal* **18**, 10480–10486 (2012).

193. Garg, N. K., Woodroofe, C. C., Lacenere, C. J., Quake, S. R. & Stoltz, B. M. A ligand-free solid-supported system for Sonogashira couplings: Applications in nucleoside chemistry. *Chemical Communications* 4551–4553 (2005) doi:10.1039/b505737j.

194. Jäger, S. *et al.* A versatile toolbox for variable DNA functionalization at high density. *J Am Chem Soc* **127**, 15071–15082 (2005).

195. Mishra, N. C. & Broom, A. D. A novel synthesis of nucleoside 5′-triphosphates. *J Chem Soc Chem Commun* 1276–1277 (1991) doi:10.1039/C39910001276.

196. Tietze, L. F. *et al.* Anticancer Agents, 15. Squaric Acid Diethyl Ester: A New Coupling Reagent for the Formation of Drug Biopolymer Conjugates. Synthesis of Squaric Acid Ester Amides and Diamides. *Chem Ber* **124**, 1215–1221 (1991).

197. Brázdilová, P. *et al.* Ferrocenylethynyl Derivatives of Nucleoside Triphosphates: Synthesis, Incorporation, Electrochemistry, and Bioanalytical Applications. *Chemistry – A European Journal* **13**, 9527–9533 (2007).

198. Moran, S., Ren, R. X. F., Sheils, C. J., Rumney IV, S. & Kool, E. T. Non-hydrogen bonding 'terminator' nucleosides increase the 3'-end homogeneity of enzymatic RNA and DNA synthesis. *Nucleic Acids Res* **24**, 2044 (1996).

199. Sampson, J. R. & Uhlenbeck, O. C. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proceedings of the National Academy of Sciences* **85**, 1033–1037 (1988).

200. Yoo, J. S., Kim, C. M., Kim, J. H., Kim, J. Y. & Oh, J. W. Inhibition of Japanese encephalitis virus replication by peptide nucleic acids targeting cis-acting elements on the plus- and minus-strands of viral RNA. *Antiviral Res* **82**, 122–133 (2009).

201. Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **180**, 51–62 (1989).

202. Andersen, T. E., Kirpekar, F. & Haselmann, K. F. RNA fragmentation in MALDI mass spectrometry studied by H/D-exchange: Mechanisms of general applicability to nucleic acids. *Journal of The American Society for Mass Spectrometry 2006 17:10* **17**, 1353–1368 (2006).

203. Malet, H. *et al.* The flavivirus polymerase as a target for drug discovery. *Antiviral Res* **80**, 23–35 (2008).

204. Mishra, A. & Rathore, A. S. RNA dependent RNA polymerase (RdRp) as a drug target for SARS-CoV2. *J Biomol Struct Dyn* **40**, 6039–6051 (2022).

205. Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications 2021 12:1* **12**, 1–17 (2021).

206. Singh, A. K. & Das, K. Insights into HIV-1 Reverse Transcriptase (RT) Inhibition and Drug Resistance from Thirty Years of Structural Studies. *Viruses* **14**, (2022).

207. Chase, J. W. & Williams, K. R. SINGLE-STRANDED DNA BINDING PROTEINS REQUIRED FOR DNA REPLICATION. *https://doi.org/10.1146/annurev.bi.55.070186.000535* **VOL. 55**, 103–136 (2003).

208. Boehmer, P. E. RNA binding and R-loop formation by the herpes simplex virus type-1 single-stranded DNA-binding protein (ICP8). *Nucleic Acids Res* **32**, 4576–4584 (2004).

209. Venkataraman, S., Prasad, B. V. L. S. & Selvarajan, R. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses* **10**, (2018).

210. Peng, Q. *et al.* Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell Rep* **31**, (2020).

211. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**, (2019).

212. Terrier, O. *et al.* Antiviral Properties of the NSAID Drug Naproxen Targeting the Nucleoprotein of SARS-CoV-2 Coronavirus. *Molecules* **26**, (2021).

213. Zeng, W. *et al.* Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun* **527**, 618 (2020).

214. Sarafianos, S. G. *et al.* Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* **385**, 693 (2009).

215. Selisko, B. *et al.* Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. *Virology* **351**, 145–158 (2006).

216. Wu, J., Liu, W., Gong, P. & Gong, P. A structural overview of RNA-dependent RNA polymerases from the Flaviviridae family. *Int J Mol Sci* **16**, 12943–12957 (2015).

217. Kim, Y. G., Yoo, J. S., Kim, J. H., Kim, C. M. & Oh, J. W. Biochemical characterization of a recombinant Japanese encephalitis virus RNA-dependent RNA polymerase. *BMC Mol Biol* **8**, 1–12 (2007).

218. Konkolova, E. *et al.* Remdesivir triphosphate can efficiently inhibit the RNA-dependent RNA polymerase from various flaviviruses. *Antiviral Res* **182**, 104899 (2020).

219. Lu, G. & Gong, P. Crystal Structure of the Full-Length Japanese Encephalitis Virus NS5 Reveals a Conserved Methyltransferase-Polymerase Interface. *PLoS Pathog* **9**, (2013).

220. Ferrige, A. G. *et al.* Disentangling electrospray spectra with maximum entropy. *Rapid Communications in Mass Spectrometry* **6**, 707–711 (1992).

221. Asare-Okai, P. N., Agustin, E., Fabris, D. & Royzen, M. Site-specific fluorescence labelling of RNA using bio-orthogonal reaction of trans-cyclooctene and tetrazine. *Chemical Communications* **50**, 7844–7847 (2014).

222. Dejmek, M. *et al.* Non-nucleotide RNA-dependent RNA polymerase inhibitor that blocks SARS-CoV-2 replication. *Viruses* **13**, 1585 (2021).