Enzymatic synthesis of reactive RNA probes containing squaramate-linked cytidine or adenosine for bioconjugations and cross-linking with lysine-containing peptides and proteins

Check for updates

Ivana Ivancová [1,3], Tania Sánchez Quirante[1,2,3], Marek Ondruš [1], Radek Pohl [1], Marta Vlková[1], Eva Žilecká[1], Evžen Bouřa[1] & Michal Hocek [1,2]✉

Protein-RNA interactions play important biological roles and hence reactive RNA probes for cross-linking with proteins are important tools in their identification and study. To this end, we designed and synthesized 5′-O-triphosphates bearing a reactive squaramate group attached to position 5 of cytidine or position 7 of 7-deazaadenosine and used them as substrates for polymerase synthesis of modified RNA. In vitro transcription with T7 RNA polymerase or primer extension using TGK polymerase was used for synthesis of squaramate-modified RNA probes which underwent covalent bioconjugations with amine-linked fluorophore and lysine-containing peptides and proteins including several viral RNA polymerases or HIV reverse transcriptase. Inhibition of RNA-depending RNA polymerases from Japanese Encephalitis virus was observed through formation of covalent cross-link which was partially identified by MS/MS analysis. Thus, the squaramate-linked NTP analogs are useful building blocks for the synthesis of reactive RNA probes for bioconjugations with primary amines and cross-linking with lysine residues.

RNA-protein interactions are a very important yet underexplored field of research[1]. One of the most efficient ways how to study them is the use of covalent cross-linking of RNA with interacting proteins[2–4]. So far, the common methods include photochemical cross-linking[5–8] or the usage of external chemical cross-linking agents[9–11]. These methods are often inefficient, non-specific, and toxic for use in live cells. Therefore, there is a need for the introduction of specific reactive groups into RNA for cross-linking with peptides and proteins. In related DNA, many established reactive groups have been linked to nucleobases in 2'-deoxyribonucleoside triphosphates (dNTPs) and incorporated enzymatically[12], e.g. Michael acceptors reacting with Cys[13], chloroacetamide reacting with Cys or His[14], and glyoxal[15] or 1,3-diketones[16] for Arg, as well as nitroso benzaldehyde[17] or squaramate[18] for Lys. On the other hand, there were only a few examples of protein-cross-linking modifications of RNA relying on 5-fluoro- or 5-azapyrimidine nucleotides that form covalent adducts with Cys-

containing proteins, e.g. methyltransferases[19–21]. Other related examples of reactive modifications of RNA included posttranscriptional labelling through bioorthogonal reactions[22–29] or alkylation of RNA with methyltransferases[30–33]. Recently, we reported chloroacetamide (CA)-linked ATP that can be incorporated into RNA by in vitro transcription (IVT) using T7 RNA polymerase (T7 RNAP) and cross-links to Cys- or His-containing peptides and proteins (Fig. 1A)[34], but there is an urgent need for reactive groups and RNA probes specific for Lys as one of the most abundant amino acid in nucleic acids binding proteins[35–37].

The well-established IVT with T7 RNAP using modified ribonucleoside triphosphates (NTPs) incorporates the modifications replacing natural counterparts at all sites[23–25,38–40]. Enzymatic site-specific single modification with RNA polymerases is very complicated[41] and was previously reported only through laborious methods of limited efficiency, i.e. position-selective labeling by T7 RNAP (PLOR)[42], splint ligation[43] or use of the extended

[1]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nam. 2, CZ-16000 Prague 6, Prague, Czech Republic. [2]Department of Organic Chemistry, Faculty of Science, Charles University, Hlavova 8, CZ-12843 Prague 2, Prague, Czech Republic. [3]These authors contributed equally: Ivana Ivancová, Tania Sánchez Quirante. ✉e-mail: hocek@uochb.cas.cz
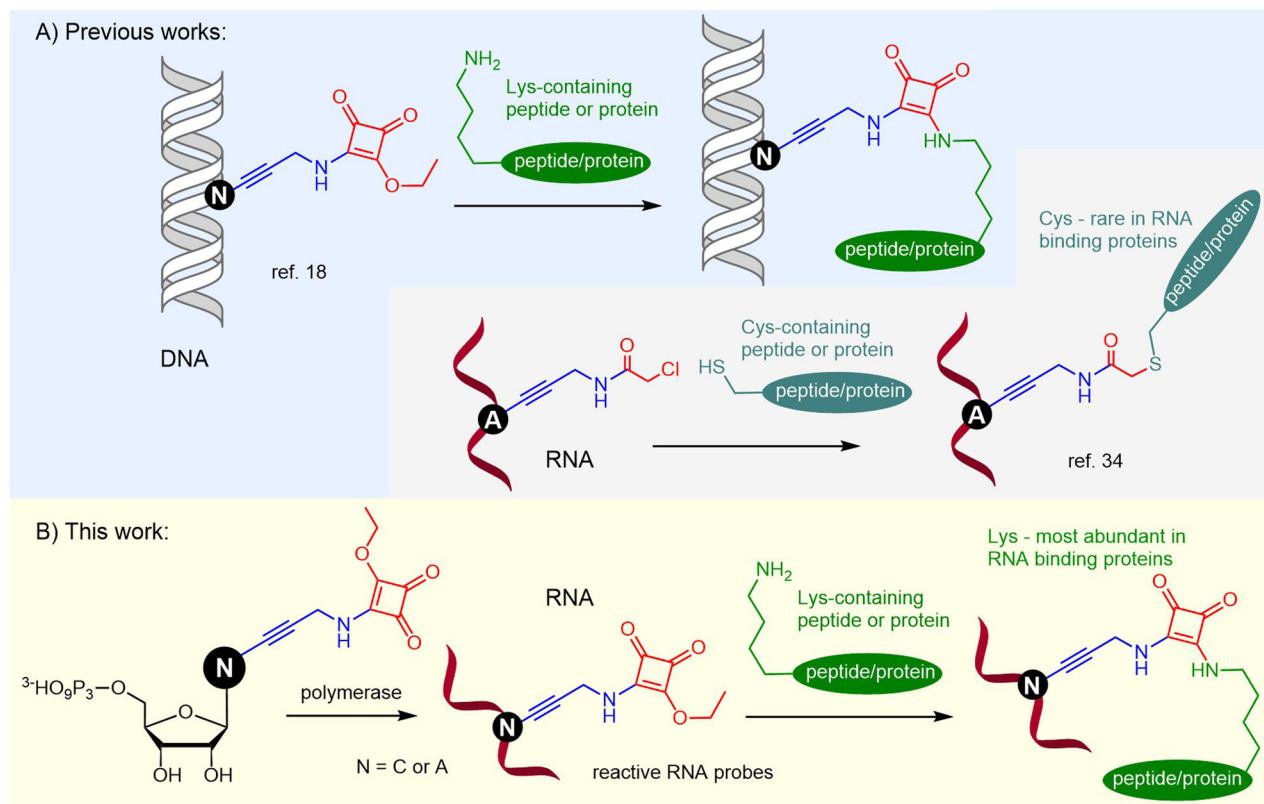
Fig. 1 | Overview of related reactive DNA and RNA probes. A Previous related works on reactive DNA and RNA probes and (**B**) conceptualization of this study: polymerase synthesis of squaramate-containing reactive RNA probes and their cross-linking with Lys-containing peptides or proteins.

genetic alphabet[26–29]. Only very recently we[44] and others[45] have independently reported enzymatic synthesis of base-modified RNA using engineered TGK DNA polymerase and DNA template for extension of an RNA primer (PEX) and an expedient site-specific or segmented incorporation of one or more modifications through (repeated) single nucleotide incorporation followed by primer extension. This approach takes advantage of the high substrate tolerance and efficiency of DNA polymerases for incorporation of even highly modified NTPs and of the possibility to exchange single-stranded DNA template to synthesize specifically modified RNA not accessible by classical IVT.

In this paper, we present the design and development of Lys-reactive RNA probes through enzymatic synthesis by incorporation of the corresponding squaramate-modified ribonucleoside triphosphates ($N^{ESQ}TPs$) with T7 RNAP or TGK DNA polymerase and applications in bioconjugations and cross-linking with Lys-containing peptides and proteins and studying inhibition of viral RNA polymerases (Fig. 1B).

Results and discussion
Synthesis of modified nucleosides and corresponding triphosphates
We designed the squaramate-modified ribonucleoside triphosphates in a similar way to the previously reported dNTP. The squaramate was linked to position 5 of cytosine through a propargyl amide linker to ensure good compatibility with polymerase incorporation. The synthesis (Fig. S1 in SI and Fig. 2A) started by the Sonogashira reaction of 5-iodocytidine with *N*-propargyl trifluoroacetamide followed by deacylation to give 5-(3-amino-propyn-1-yl)-cytidine **3** in very good yield (87%). The reaction of nucleoside **3** with 2 equiv. of diethyl squarate gave squaramate-linked ribonucleoside (C^{ESQ}) in acceptable yield (22%). The triphosphorylation[46] of C^{ESQ} with $POCl_3$ in trimethyl phosphate followed by treatment with tributyl ammonium pyrophosphate gave the corresponding $C^{ESQ}TP$ (38%). Additionally, we synthesized $A^{ESQ}TP$ bearing squaramate group at position 7 of

7-deazaadenine base tethered through aminopropagyl linker. The two-step synthesis (Fig. S2 in SI and Fig. 2A) started by the Sonogashira cross-coupling reaction of 5-iodo-adenosine (A^I) with *N*-(propargyl)-squaramate (**PAS**) in the presence of Pd(OAc)$_2$, CuI, TPPTS, and Et$_3$N in a mixture of H$_2$O/acetonitrile (2:1). Ethoxy squarate modified adenosine (A^{ESQ}) was isolated after column chromatography in plausible yield (26%). Subsequent triphosphorylation reaction[43] of A^{ESQ} gave the corresponding $A^{ESQ}TP$ (2%). The yield of $A^{ESQ}TP$ was significantly affected by the extensive purification procedures.

Synthesis of ESQ-modified RNA by in vitro transcription using T7 RNAP
First, we tested the ability of $C^{ESQ}TP$ to serve as a CTP analog for T7 RNA polymerase in IVT. To describe the base generality, we also tested $A^{ESQ}TP$ as an ATP analog (Fig. 2B). We performed the IVT reactions in the presence of various DNA templates (for sequences, see Supplementary Table S1). The templates contained one, three, seven, or eight dG/dT residues in the coding region that guided the incorporation of one to eight C^{ESQ} or A^{ESQ} nucleotides into RNA transcript. We used two 5'-terminal 2'-MeO ribonucleotides of antisense strands of DNA templates to minimize non-templated nucleotide addition[47]. All transcription reactions were analyzed on a polyacrylamide gel and phosphor-imaged (Supplementary Fig. S3) or post-stained using Sybr-Gold (Fig. 2C, Supplementary Figs. S5–6). Since commercial transcription buffer supplied with the T7 RNAP contains spermidine and the squaramate group is known to readily react with primary amines[48], we had to optimize the IVT conditions and composition of the buffers to prevent an undesired reaction of the ESQ-linked nucleotide or RNA with spermidine. Hence the spermidine was either replaced with quaternary amine betaine or we used a buffer with no amines (Supplementary Figs. S4, S7, S21, and S35). The IVT synthesis of RNA containing one or three C^{ESQ} or A^{ESQ} modifications (i.e. **35_RNA_1C^{ESQ}** or **35_RNA_3C^{ESQ}**) proceeded smoothly with good conversions (Fig. 2C), whereas the formation of RNA with higher number of
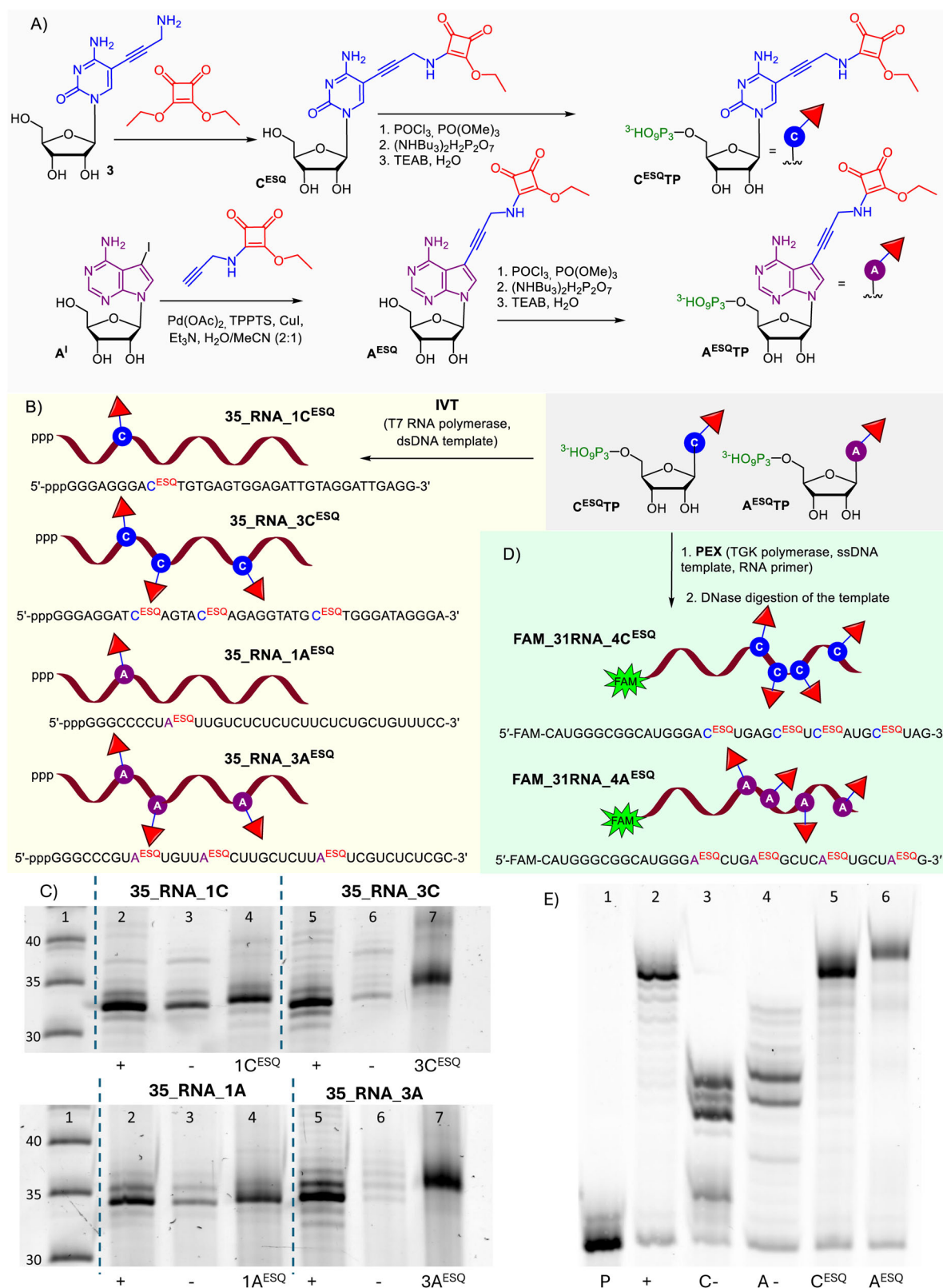
Fig. 2 | Synthesis of modified NTPs and RNA probes. A Synthesis of **C^ESQTP** and **A^ESQTP**. **B** IVT reactions with **C^ESQTP** or **A^ESQTP** and T7 RNAP. **C** PAGE analysis of IVT reactions. (+) positive controls with natural NTPs, (−) negative controls in the absence of CTP or ATP, (1C^ESQ, 3C^ESQ) IVT reactions with **C^ESQTP**, UTP, GTP, and ATP, (1A^ESQ, 3A^ESQ) IVT reactions with **A^ESQTP**, UTP, GTP, and CTP. **D** PEX reactions with **C^ESQTP** or **A^ESQTP** and TGK polymerase. **E** PAGE analysis of PEX reactions. (P) primer, (+) positive control with natural NTPs, (C−) negative control in the absence of CTP, (A−) negative control in the absence of ATP, (C^ESQ) PEX reaction with **C^ESQTP**, UTP, GTP and ATP, (A^ESQ) PEX reaction with **A^ESQTP**, UTP, GTP and CTP. For the uncropped gels see Supplementary Fig. 1.
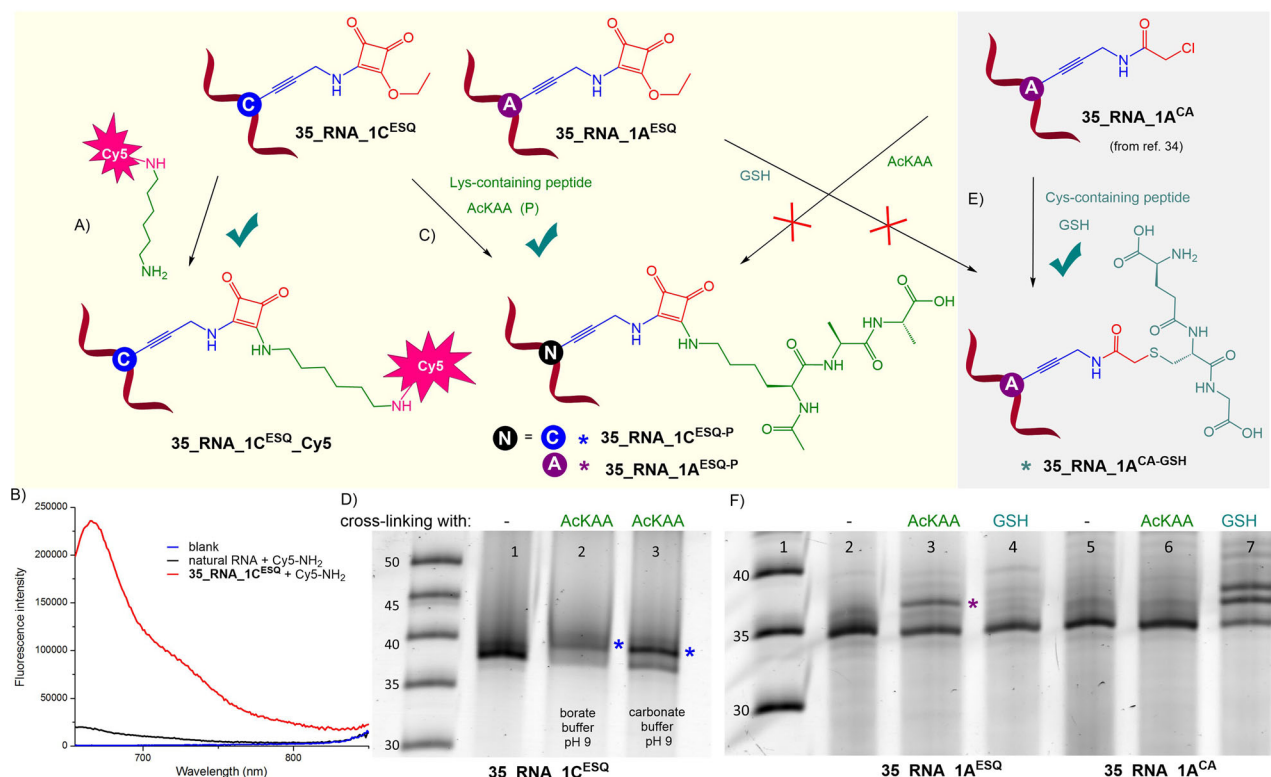
Fig. 3 | Reactivity and specificity of modified RNA probes. Cross-linking reaction scheme of ESQ-modified RNA with (A) amine-linked Cy5 fluorophore, (C) Lys-containing peptide, and (E) crossreactivity in comparison to CA-modified RNA with Cys-containing GSH. B) Fluorescence measurement of cross-linked **35_RNA_1C^{ESQ}_Cy5** (red line) and non-cross-linked natural RNA (black line). D) PAGE analysis of cross-linking reactions of **35_RNA_1C^{ESQ}** with AcKAA peptide.

(lane 1) **35_RNA_1C^{ESQ}**; (lane 2) cross-linking reaction in borate or (lane 3) carbonate buffer. F) PAGE analysis of crosslinking reactions of **35_RNA_1A^{ESQ}** and **35_RNA_1A^{CA}** with AcKAA and GSH. (lane 2) **35_RNA_1A^{ESQ}**, (lane 3) **35_RNA_1A^{ESQ}** with AcKAA, (lane 4) **35_RNA_1A^{ESQ}** with GSH, (lane 5) **35_RNA_1A^{CA}**; (lane 6) **35_RNA_1A^{CA}** with AcKAA, (lane 7) **35_RNA_1A^{CA}** with GSH. For the uncropped gels see Supplementary Fig. 1.

reactive modifications (i.e. **35_RNA_7C^{ESQ}**) was less efficient (Supplementary Figs. S3, 4). In some cases, we have observed formation of minor amounts of n + 1 products of non-templated extension of the RNA products (Fig. 2C and Supplementary Figs. S3–6). In case of IVT with templates encoding for incorporation of just one modified nucleotide, the negative control experiments in the absence of the corresponding non-modified NTP also showed some formation of full-length product due to misincorporation of uridine, which is quite common in such cases[49]. In all cases, the identity of the full-length modified RNA products was also confirmed by MALDI-TOF (Supplementary Fig. S22-27) or LC-MS (Supplementary Figs. S33, 34, S37, 38). Additional control analysis of the reaction mixture on both native and denaturing protein gels excluded crosslinking of squaramate-modified RNA with the T7 RNA polymerase during the transcription reaction (Supplementary Fig. S13).

Synthesis of ESQ-modified RNA by PEX using TGK polymerase

Following our recent paper on the enzymatic synthesis of base-modified RNA using engineered DNA polymerases[44], we tested **C^{ESQ}TP** and **A^{ESQ}TP** as substrates for RNA synthesis by PEX using TGK polymerase (Fig. 2D)[50]. We used a 31-nt single-stranded DNA template encoding for four cytosines and four adenosines respectively in the complementary strand and a 5'-FAM-labelled 15-nt RNA primer (for sequences see Supplementary Table S1). The PEX reaction was followed by the digestion of the template by Turbo DNase, and we analyzed the products by polyacrylamide gel electrophoresis (PAGE) (Fig. 2E and Supplementary Fig. S8). The formation of natural **FAM_31RNA** and the desired modified **FAM_31RNA_4C^{ESQ}** and **FAM_31RNA_4A^{ESQ}** containing four modified **C^{ESQ}** or **A^{ESQ}** nucleotides was also confirmed by LC-MS (Supplementary Fig. S30-32, Supplementary Table S3). MS analysis revealed a non-templated addition of one additional GMP resulting in n + 1 longer RNAs.

Conjugation reactions of ESQ-modified RNA with amine and Lys-containing peptide

First, we tested the reactivity of **C^{ESQ}**-modified RNA with a primary amine. We used amine-linked Cy5 fluorophore (Fig. 3A, Supplementary Fig. S9) for site-specific RNA labelling. Non-labelled natural RNA (for control experiment) and squaramate-modified **35_RNA_1C^{ESQ}** were prepared by in vitro transcription followed by DNase I treatment and purified by spin columns. Model conjugation reaction was performed with amine-linked Cy5 fluorophore (100 - 400 equiv.) in buffer (pH 7.4-7.9) at 37 °C. After overnight incubation, the Cy5-labelled RNA conjugate was purified from the excess of the free dye and analyzed by fluorescence spectroscopy and fluorescence imaging of the denaturing PAGE. The strong fluorescence signal was observed only for cross-linked **35_RNA_1C^{ESQ}_Cy5** while the non-modified RNA was not labelled under the same conditions (Fig. 3B, Supplementary Fig. S9a). The fluorescence imaging of the denaturing PAGE gel, shown in Supplementary Fig. S9b, supports the conclusion of fluorescence measurements that a stable covalent conjugate of Cy5 with **35_RNA_1C^{ESQ}** has been formed.

Then we tested the reactions of ESQ-modified RNAs with a Lys-containing peptide (AcKAA, P). We used **35_RNA_1C^{ESQ}** and **35_RNA_3C^{ESQ}** prepared by in vitro transcription using T7 RNA polymerase followed by DNase I treatment, purified and characterized by denaturing PAGE (Fig. 3C, D and Supplementary Fig. S5). Using **35_RNA_3C^{ESQ}** containing three reactive modifications, we observed formation of mixtures of conjugate products containing one, two or three peptides (Supplementary Fig. S11). We initially tested the influence of peptide excess (500 or 5000 equiv.) in borate buffer of different pH (9 or 10) for RNA-peptide conjugation efficacy. The buffer pH did not show any significant effect on the conjugation performance. On the other hand, increased excess of peptide pushed the conversion towards triply modified

Fig. 4 | Reactivity of modified RNA probe with recombinant proteins. A Cross-linking of ESQ-modified RNA (**36_vRNA_1gC^{ESQ}**) with Lys-containing protein. **B** 10% SDS-PAGE analysis of cross-linking experiments with natural or ESQ-modified RNA and various recombinant proteins (2–20 equiv. of protein; for the uncropped gels see Supplementary Fig. 1).
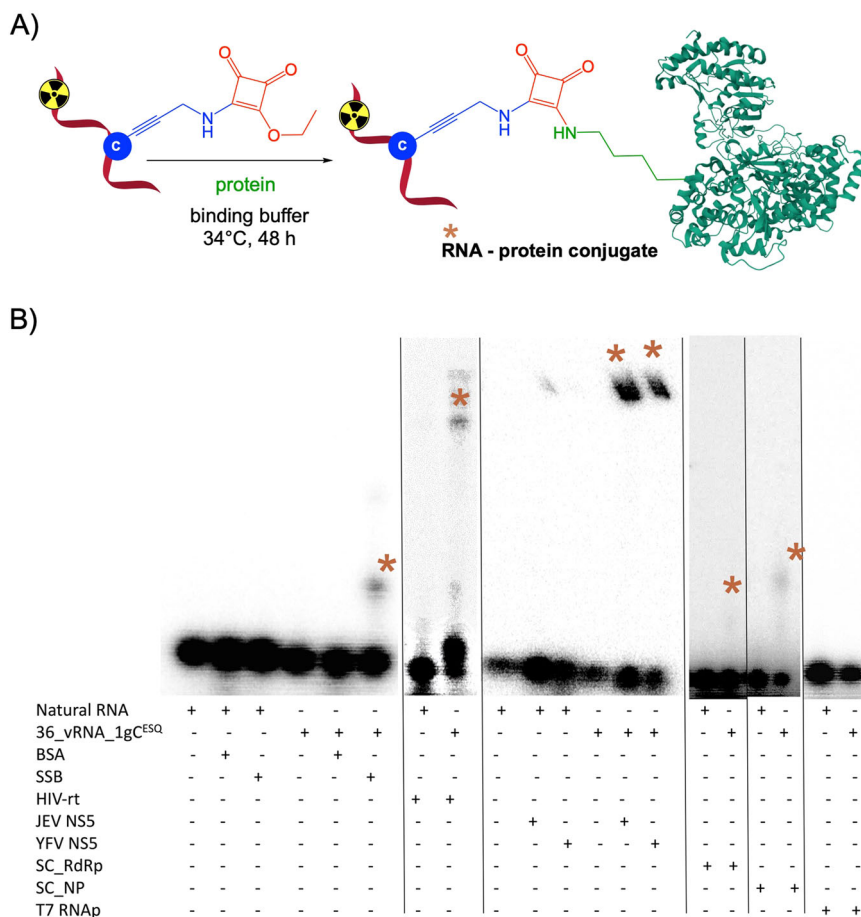
RNA-peptide conjugate but we never reached full conversion at all positions and we still observed a mixture of RNA-peptide conjugates with two and three peptide molecules attached. Having different electrophoretic mobilities, we could identify them on denaturing PAGE (Supplementary Fig. S10, S11b) and confirm by LC-MS (Supplementary Fig. S40-41, Supplementary Table S4).

We then used the higher peptide excess (9000 equiv.) in two different buffer systems (borate buffer, pH 9, and carbonate-bicarbonate buffer, pH 9.5) for the conjugation with **35_RNA_1C^{ESQ}** to yield the corresponding covalent conjugates **35_RNA_1C^{ESQ-P}** in 80 and 75% conversion, respectively (Fig. 3D, Supplementary Fig. S11a and S39). To compare the substrate specificity of squaramate-linked RNA probes, we performed a direct comparison experiment of cross-linking reactions of **A^{ESQ}**-modified **35_RNA_1A^{ESQ}** with previously published chloroacetamide (CA)-modified RNA (**35_RNA_1A^{CA}**)[34]. We used Cys-containing Glutathione (GSH) and the Lys-containing peptide in reactions with **35_RNA_1A^{ESQ}** and **35_RNA_1A^{CA}** (Fig. 3E, F, Supplementary Figs. S12a, S36). As expected, squaramate-modified **35_RNA_1A^{ESQ}** reacted only with Lys-containing peptide and not with GSH. On the other hand, the chloroacetamide-linked **35_RNA_1A^{CA}** reacted exclusively only with GSH showing the selectivity of both RNA probes towards specific amino acid residues. These results are supported by LC-MS measurements of both RNA-peptide and RNA-GSH conjugates (Supplementary Figs. S42, S43).

Cross-linking of ESQ-modified RNA with proteins
For cross-linking reactions of the ESQ-modified RNA with RNA binding proteins (RBP) we selected three RNA-dependent RNA polymerases (RdRp; Japanese encephalitis virus (JEV) NS5 protein, Yellow fever virus (YFV) NS5[51] and SARS-CoV-2 RdRp (SC_RdRp))[52], as well as SARS-CoV-2 nucleoprotein (SC_NP)[53] and HIV reverse transcriptase (HIV-rt)[54] as biologically relevant targets each of them containing several lysine amino acid

residues. Bovine serum albumin (BSA) as an RNA non-binding protein was used as a negative control together with single-strand binding protein (SSB) having a 10-fold lower affinity for ssRNA compared to DNA[55]. For all RNA-protein crosslinking studies, radioactively labelled natural or ESQ-modified RNA was used (prepared by IVT reaction described in the method section). RNA-protein cross-linking reactions were performed at pH 7.4-8, with 2 equiv. of HIV-rt, 4 equiv. of JEV and YFV NS5 protein, 8 equiv. of SC_RdRp and SC_NP, and higher excess of RNA non-binding or RNA weakly binding proteins (20 equiv. of BSA and SSB) (Fig. 4A). The ability of proteins to bind **36_vRNA_1gC^{ESQ}** bearing one modification (for sequence see Supplementary Table S1) was monitored by electrophoretic mobility shift assay (EMSA).

Reaction analysis by 10% native PAGE confirmed that the squaramate modification did not prevent protein binding to RNA (Supplementary Figs. S14a, S15a, S17a, S18a). In all shown cases, except for incubation of RNA with T7 RNA polymerase, a band with slower mobility was observed, although the complex formation between RNA and SC_RdRp was very weak (Fig. 4B). As expected, no RNA-protein complexes were formed in the case of T7 RNAP and BSA (Fig. 4A, Supplementary Fig. S13 and S14). The denaturing SDS-PAGE revealed new bands with slower electrophoretic mobility indicating the formation of corresponding RNA-protein covalent cross-linked conjugates with moderate to good conversions (**RNA$^{ESQ\_SSB}$** 9%, **RNA$^{ESQ\_HIV\_rt}$** 15%, **RNA$^{ESQ\_JEV\_NS5}$** 43%, **RNA$^{ESQ\_YFV\_NS5}$** 40%; Fig. 4B, Supplementary Figs. S14b, S18b, S15b, Table 1). The identity of the covalent **36_vRNA_1gC$^{ESQ\_SSB}$** conjugate was confirmed by MS (Supplementary Fig. S44). Only traces of covalent cross-link were detected with SC_RdRp and SC_NP proteins. More squaramate modifications within the RNA strand in the case of **36_vRNA_3gC^{ESQ}** did not increase protein binding, nor the conversions (Supplementary Figs. S15, S17, Table S5). It should be noted that proximity effect is needed for an efficient cross-linking reaction and hence it is not the number of lysines but the accessibility and optimum

Table 1 | Summary of cross-linking reactions between 36_vRNA_1gC^ESQ and proteins

Protein	Molecular weight [kDa]	RNA binding	No. of Lys residues	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.

positioning of one or more lysines in the vicinity to the reactive group is the driving force and decisive factor for the outcome of these reactions.

Gel-based RdRp polymerase assays in vitro

Following the cross-linking studies that have shown that the RdRps from *Flaviviridae* family readily cross-linked with squaramate-modified RNAs, we decided to investigate how the modified **C^ESQTP** influences the polymerase activity of the JEV recombinant RdRp (JEV_NS5) and SARS-CoV-2 RdRp (SC_RdRp) on an RNA template-product scaffold in vitro (Fig. 5A and Supplementary Figs. S19, S20). We used gel-based RdRp polymerase assay conditions similar to those published for various flaviviruses[56] and non-modified templates that represent the 3'-end of the JEV genome of the c5'-UTR, that folds into a hairpin with 5'-FAM-labelled overhang as a template for primer extension (for sequences see Supplementary Table 2)[57]. RdRp polymerases are known to go through repeated cycles of abortive initiation before starting elongation mode[58] and consecutive cytosines in the template may cause polymerase slippage[59] which leads to the presence of short abortive products and ladder-like patterns on the gel. SARS-CoV-2 RdRp produced full-length ESQ-modified RNA, which was confirmed by PAGE (Supplementary Fig. S20) and MALDI (Supplementary Figs. S28, 29). On the other hand, in the case of the JEV recombinant RdRp, we observed the full-length product of natural RNA and only traces of squaramate-modified RNA (Supplementary Fig. S19a-b). After subsequent treatment of the reaction mixture with Proteinase K (pK) to release all RNA from the JEV NS5 protein-RNA complex, both natural and modified full-length RNAs were observed (Fig. 5C, lanes 3 and 4, respectively; Supplementary Fig. S19a, b). In the case of ESQ-modified RNA, an additional band of slower mobility was formed which intensity increased after prolonged reaction time (Fig. 5C, lane 6; Supplementary Fig. S19b). We assumed that this band might be the squaramate-modified RNA cross-linked to short peptide resulting from JEV NS5 protein digestion. Additionally, the presence of the cross-linked product of newly modified RNA and full-length JEV NS5 RdRp was independently proved by the comparison of 10% SDS protein gel visualized both by FAM for RNA and immunodetection of JEV NS5 (Fig. 5B; Supplementary Fig. S19d, e). These results suggest that the JEV RdRp incorporates **C^ESQTP** and further cross-links with produced ESQ-modified RNA during the extension reaction.

Proteomic analysis of ESQ-modified RNA-JEV conjugate

To characterize cross-linked **46_RNA_HP^C4-JEV** conjugate and to determine the RNA-protein binding sites, two independent samples of the reaction mixture after polymerase assay were submitted to LC-MS/MS proteomic analysis (Fig. 5D; and Supplementary Fig. S45, 46) which revealed three lysine residues participating in protein cross-linking to **C^ESQ** in newly synthesized RNA (K269 in sample 1; K462 and/or K463 in sample 2). We identified fragments corresponding to cross-linking to neighbouring lysine residues in the spectrum of sample 2, and it can not be distinguished whether it is on K462 or K463. For sequence coverage of analyzed protein samples see

Supplementary Fig. S47, 48 and Supplementary Table S6 in supporting material. The comparison with the published crystal structure of JEV RdRp (PDB 4K6M) reveals the location of the K269 in the 10-residue linker region (residues 266-275) which connects the MTase on its C-terminus (Supplementary Fig. S49). The sequence of this linker varies highly among genus flavivirus. The lysines K462 and K463 are in the ring finger domain of RdRp (residues 453-479) that forms the roof of the NTP entry channel at the -5 position from the NTP binding motif F (Phe467). These amino acids are highly conserved in flaviviruses[60].

Conclusions

Herein we report the synthesis of squaramate-modified ribonucleotides **C^ESQTP** and **A^ESQTP** and their incorporation into the RNA by in vitro transcription. We have found that the spermidine (primary amine) present in the commercial reaction buffer interfered with the ESQ moiety during the transcription reaction. After optimization of reaction conditions by exchanging the spermidine for quaternary betaine or using a buffer with no amine, full-length RNA transcripts containing one to eight **C^ESQ** and **A^ESQ** modifications were synthesized. We have also successfully used **C^ESQTP** and **A^ESQTP** in PEX for the synthesis of modified RNAs using engineered TGK DNA polymerase in analogy to our recent paper[44]. The ESQ-modified RNA was used in post-synthetic labelling with amine-linked Cy5 fluorophore and conjugation with short Lys-containing peptide. The reactions proceeded with high conversion only in the presence of a large excess of the amine or peptide. Squaramate moiety in RNA probes showed its specificity for cross-linking reactions with lysine-containing peptides and no reactivity with cysteine-containing peptides and thus is fully complementary to the previously reported[34] chloroacetamide (CA) modification that reacts with Cys (and His) but not Lys. Most importantly, the ESQ-modified RNA probes reacted with various lysine-containing RNA-binding proteins (i.e., JEV and YFV NS5, and HIV-rt) and gave the covalently cross-linked conjugates in 9-43% conversions. On the other hand, negative control experiments with RNA non-binding or weakly binding proteins did not show any significant formation of covalent conjugates confirming that the proximity effect is crucial for good conversion. Finally, we investigated how **C^ESQTP** affects the polymerase activity of RNA-dependent RNA polymerases in vitro. Squaramate modification did not interfere with SARS-CoV-2 RdRp and gave full-length RNA. On the other hand, experimental results showed that newly synthesized RNA with incorporated **C^ESQTP** cross-linked with JEV RdRp during polymerase reaction. Proteomic analysis of the reaction mixtures of JEV RdRp RNA extension identified K269, K462, and K463 lysine residues modified by **C^ESQ** from modified RNA. Hence there is certain potential in using reactive modified nucleotides for inhibition of viral RNA polymerases, though probably not in the therapeutically relevant mode.

The squaramate moiety specifically reacts with primary amines including lysine to form stable amide bonds under physiological conditions without any external reagents and hence it has a good potential in post-synthetic labelling of RNA, bioconjugation of RNA with various biomolecules, cross-linking with proteins with abundant lysine or in developing irreversible inhibitors of RNA-binding enzymes containing free lysine residues (e.g., viral polymerases). It is complementary to the previously reported chloroacetamide reactive group[34] and together or even in combination they can be used for RNA proteomics to identify new RNA-binding proteins.

Methods

Complete experimental part including the chemical syntheses of modified nucleotides and their characterization, all enzymatic IVT and PEX reactions together with cross-linking reactions, and others are given in Supplementary Methods section in Supplementary Information.

In vitro transcription with A^ESQTP and C^ESQTP

In vitro transcription reaction (10 µL) contained dsDNA template (1 µM, ssDNA strands annealed in 10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.8), natural NTPs (2 mM), **N^ESQTP** (1 mM for **C^ESQTP** or 2 mM for **A^ESQTP**), T7 RNA polymerase (2 U/µL) and buffer (40 mM Tris-HCl, 6 mM

$MgCl_2$, 10 mM DTT, 10 mM NaCl, pH 7.9). Water was used instead of $C^{ESQ}TP$ or $A^{ESQ}TP$ in the negative control experiments. The mixture was incubated at 37 °C for 5 h. When the reaction was completed, the dsDNA

template was digested with DNase I (0.1 U/µL) for 1 h at 37 °C followed by the addition of EDTA (50 mM) and heated for 15 min at 75 °C. The reaction mixture was purified using Monarch Kit 50 µg for further use.
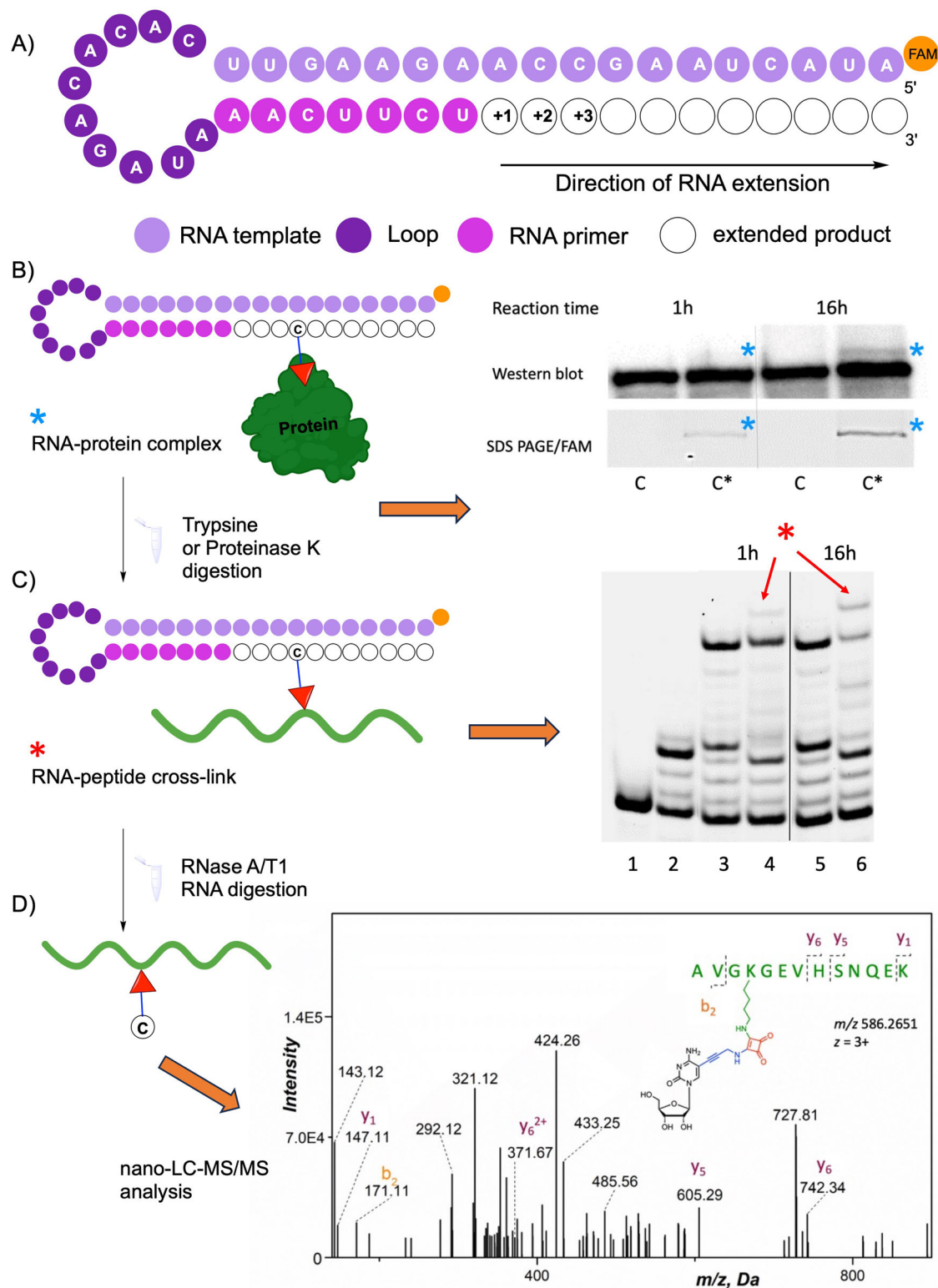
Fig. 5 | Summary of RdRp assay. A General scheme of RdRp polymerase assay. Hairpin RNA substrate with template-product scaffold. **B** Gel-based JEV RdRp assay analyzed by western blot and SDS-PAGE (lane C) positive control – reaction with natural NTPs; (lane C*) reaction with $C^{ESQ}TP$, GTP, ATP, UTP. For the uncropped gels see Supplementary Fig. 1. **C** Proteinase K or Trypsin digestion of the RNA-protein conjugate; (lane 1) template; (lane 2) negative control – reaction without CTP; (lanes 3 and 5) positive control – reaction with natural NTPs; (lanes 4 and 6) reaction with $C^{ESQ}TP$, GTP, ATP, UTP. **D** RNA digestion for nano LC-MS/MS analysis and spectrum of **46_RNA_HP$^{C4\_JEVNS5}$** with identification of K269 (m/z acquired: 586.26 Da; Supplementary Fig. S45).

PEX using TGK polymerase

PEX reaction (10 μL) contained DNA template Prb4basII (4.8 μM), FAM-labelled DNA primer Prim248-short (4 μM), natural NTPs (0.8 mM), $\textbf{N}^{\textbf{ESQ}}\textbf{TP}$ (0.8 mM), TGK polymerase (1.5 μM) and ThermoPol reaction buffer (1X). The mixture was heated at 95 °C for 30 sec and followed by 60 °C for 2 h. Positive control experiment was performed using natural CTP (0.8 mM) and negative control in the absence of $\textbf{N}^{\textbf{ESQ}}\textbf{TP}$. Reactions were then treated with TurboDNase (2U) and heated at 37 °C for 30 min. The crude reaction mixture was purified using a QIAquick kit.

Cross-linking reaction of ESQ-modified RNA (35_RNA_1A$^{\textbf{ESQ}}$) and CA-modified RNA (35_RNA_1A$^{\textbf{CA}}$) with AcKAA peptide and Glutathione

Cross-linking reaction (25 μL) contained **35_RNA_1A$^{\textbf{ESQ}}$** or **35_RNA_1A$^{\textbf{CA}}$** (2 μM), peptide AcKAA (10000 equiv.) or GSH (10000 equiv) and borate buffer (0.1 M, pH 9). The mixture was incubated at 37 °C for 48 h. Crude reaction mixtures were purified using a Monarch RNA cleanup kit.

Cross-linking of ESQ-modified RNA with proteins NS5 of Japanese Encephalitis Virus (JEV) and Yellow Fever Virus (YFV)

Radioactively-labelled natural RNA and **36_vRNA_1gC$^{\textbf{ESQ}}$, 36_vRNA_3fC$^{\textbf{ESQ}}$** (with one and three modifications respectively) were prepared by in vitro transcription as described above. Natural or modified RNA (0.5 μM) was incubated with 2 μM of NS5 protein (JEV resp. YFV) in 10X binding buffer (50 mM Tris-HCl of pH 7.4, 100 mM DTT, 5% Triton X-100, 10% glycerol; 2 μL), 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 for 48 h at 34 °C, then diluted with 2×VPS loading buffer, denatured for 10 min at 95 °C before loading, and analyzed 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).

SARS-CoV-2 RdRp mediated RNA extension assay[61]

The polymerase activity of SARS-CoV-2 RdRp was determined in a PEX reaction using fluorescently-labelled RNA templates. The reaction mixture (20 μL) contained a reaction buffer (10 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$, 10 mM KCl, 1 mM β ME), 10 μM NTPs, 0.5 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 $\textbf{C}^{\textbf{ESQ}}\textbf{TP}$ 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 before loading. Samples were separated by 12.5% PAGE (acrylamide/bisacrylamide 19:1, 25% urea) under denaturing conditions (TBE 1×, 42 mA, 1 hour).

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

LC-MS characterization of ESQ-modified RNA and their corresponding peptide conjugates

LC-MS analyses were carried out using bioZen 1.7 μm Oligo, 2.1 × 150 mm column (Kinetex) using mobile phase A (15 mM Et_3N, 300 mM HFIP in H_2O) and mobile phase B (15 mM Et_3N, 300 mM HFIP in MeOH) by 30 min gradient from 5% B to 50% B. Deconvolutions of LC-MS spectra were carried out using the UniDec program[62].

Data availability

Detailed procedures and data are given in Supplementary Information. NMR data are provided in Supplementary Data 1. Uncropped gels from Figs. 1–5 are provided in Supplementary Data 2. Primary data are available from repository: https://doi.org/10.48700/datst.6pjaf-g8g07.

References

1. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018).
2. Nechay, M. & Kleiner, R. E. High-throughput approaches to profile RNA-protein interactions. *Curr. Opin. Chem. Biol.* **54**, 37–44 (2020).
3. Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA-protein interactions. *Nat. Methods* **16**, 225–234 (2019).
4. Bae, J. W., Kwon, S. C., Na, Y., Kim, V. N. & Kim, J.-S. Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nat. Struct. Mol. Biol.* **27**, 678–682 (2020).
5. Willis, M., Hicke, B., Uhlenbeck, O., Cech, T. & Koch, T. Photocrosslinking of 5-Iodouracil-Substituted RNA and DNA to Proteins. *Science* **262**, 1255–1257 (1993).
6. Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070 (2014).
7. Panhale, A. et al. CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nat. Commun.* **10**, 2682 (2019).
8. Luo, H. et al. Photocatalytic Chemical Crosslinking for Profiling RNA-Protein Interactions in Living. *Cells Angew. Chem. Int. Ed.* **61**, e202202008 (2022).
9. 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).
10. Chiaruttini, C. et al. Protein-RNA crosslinking in Escherichia coli 30S ribosomal subunits. Identification of a 16S rRNA fragment crosslinked to protein S12 by the use of the chemical crosslinking reagent 1-ethyl-3-dimethyl-aminopropylcarbodiimide. *Nucleic Acids Res.* **23**, 7657–7676 (1982).
11. Zaman, U. et al. Dithiothreitol (DTT) acts as a specific, UV-inducible cross-linker in elucidation of protein-RNA interactions. *Mol. Cell. Proteom.* **14**, 3196–3210 (2015).
12. Ivancová, I., Leone, D.-L. & Hocek, M. Reactive modifications of DNA nucleobases for labelling, bioconjugations, and cross-linking. *Curr. Opin. Chem. Biol.* **52**, 136–144 (2019).
13. Dadová, J. et al. Vinylsulfonamide and acrylamide modification of DNA for cross-linking with proteins. *Angew. Chem. Int. Ed.* **52**, 10515–10518 (2013).
14. Olszewska, A., Pohl, R., Brázdová, M., Fojta, M. & Hocek, M. Chloroacetamide-linked nucleotides and DNA for cross-linking with peptides and proteins. *Bioconjugate Chem.* **27**, 2089–2094 (2016).
15. Leone, D.-L. et al. Glyoxal-linked nucleotides and DNA for bioconjugations and crosslinking with arginine-containing peptides and proteins. *Chem. Eur. J.* **28**, e202104208 (2022).
16. Leone, D., Hubálek, M., Pohl, R., Sýkorová, V. & Hocek, M. 1,3-diketone-modified nucleotides and DNA for cross-linking with arginine-containing peptides and proteins. *Angew. Chem. Int. Ed.* **133**, 17523–17527 (2021).
17. Guo, A.-D. et al. Spatiotemporal and global profiling of DNA-protein interactions enables discovery of low-affinity transcription factors. *Nat. Chem.* **15**, 803–814 (2023).
18. 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.* **58**, 13345–13348 (2019).
19. Liu, Y. & Santi, D. V. m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. *Proc. Natl Acad. Sci. USA* **97**, 8263–8265 (2000).
20. Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* **31**, 458–464 (2013).
21. Dai, W. et al. Activity-based RNA-modifying enzyme probing reveals DUS3L-mediated dihydrouridylation. *Nat. Chem. Biol.* **17**, 1178–1187 (2021).
22. Fantoni, N. Z., El-Sagheer, A. H. & Brown, T. A Hitchhiker's guide to click-chemistry with nucleic acids. *Chem. Rev.* **121**, 7122–7154 (2021).

23. George, J. T. & Srivatsan, S. G. Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. *Methods* **120**, 28–38 (2017).

24. Sawant, A. A. et al. A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **44**, e16 (2016).

25. Walunj, M. B., Tanpure, A. A. & Srivatsan, S. G. Post-transcriptional labeling by using Suzuki-Miyaura cross-coupling generates functional RNA probes. *Nucleic Acids Res.* **46**, e65 (2018).

26. Someya, T., Ando, A., Kimoto, M. & Hirao, I. Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry. *Nucleic Acids Res* **43**, 6665–6676 (2015).

27. Eggert, F. & Kath-Schorr, S. A cyclopropene-modified nucleotide for site-specific RNA labeling using genetic alphabet expansion transcription. *Chem. Commun.* **52**, 7284–7287 (2016).

28. Eggert, F., Kulikov, K., Domnick, C., Leifels, P. & Kath-Schorr, S. Illuminated by foreign letters—strategies for site-specific cyclopropene modification of large functional RNAs via in vitro transcription. *Methods* **120**, 17–27 (2017).

29. Bornewasser, L., Domnick, C. & Kath-Schorr, S. Stronger together for in-cell translation: natural and unnatural base modified mRNA. *Chem. Sci.* **13**, 4753–4761 (2022).

30. Mattay, J., Dittmar, M. & Rentmeister, A. Chemoenzymatic strategies for RNA modification and labeling. *Curr. Opin. Chem. Biol.* **63**, 46–56 (2021).

31. Anhäuser, L., Hüwel, S., Zobel, T. & Rentmeister, A. Multiple covalent fluorescence labeling of eukaryotic mRNA at the poly(A) tail enhances translation and can be performed in living cells. *Nucleic Acids Res.* **47**, e42 (2019).

32. Hartstock, K. et al. MePMe-seq: antibody-free simultaneous m6A and m5C mapping in mRNA by metabolic propargyl labeling and sequencing. *Nat. Commun.* **14**, 7154 (2023).

33. van Dülmen, M., Muthmann, N. & Rentmeister, A. Chemo-enzymatic modification of the 5' cap maintains translation and increases immunogenic properties of mRNA. *Angew. Chem. Int. Ed.* **60**, 13280–13286 (2021).

34. Brunderová, M., Krömer, M., Vlková, M. & Hocek, M. Chloroacetamide-modified nucleotide and RNA for bioconjugations and cross-linking with RNA-binding. *Proteins Angew. Chem. Int. Ed.* **62**, e202213764 (2023).

35. Bartas, M., Červeň, J., Guziurová, S., Slychko, K. & Pečinka, P. Amino acid composition in various types of nucleic acid-binding proteins. *Int. J. Mol. Sci.* **22**, 922 (2021).

36. Hacker, S. et al. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).

37. Ukmar-Godec, T. et al. Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat. Commun.* **10**, 2909 (2019).

38. 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).

39. Pawar, M. G., Nuthanakanti, A. & Srivatsan, S. G. Heavy atom containing fluorescent ribonucleoside analog probe for the fluorescence detection of RNA-ligand binding. *Bioconjug. Chem.* **24**, 1367–1377 (2013).

40. 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).

41. Flamme, M., McKenzie, L. K., Sarac, I. & Hollenstein, M. Chemical methods for the modification of RNA. *Methods* **161**, 64–82 (2019).

42. Liu, Y. et al. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368–372 (2015).

43. Hertler, J. et al. Synthesis of point-modified mRNA. *Nucleic Acids Res.* **50**, e115 (2022).

44. Brunderová, M. et al. Expedient production of site specifically nucleobase-labelled or hypermodified RNA with engineered thermophilic DNA polymerases. *Nat. Commun.* **15**, 3054 (2024).

45. Haslecker, R. et al. Extending the toolbox for RNA biology with SegModTeX: a polymerase-driven method for site-specific and segmental labeling of RNA. *Nat. Commun.* **14**, 8422 (2023).

46. Yoshikawa, M., Kato, T. & Takenishi, T. A novel method for phosphorylation of nucleosides to 5'-nucleotides. *Tetrahedron Lett.* **8**, 5065–5068 (1967).

47. Kao, C., Zheng, M. & Rüdisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272 (1999).

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

49. Pomerantz, R. T. et al. Mechanism of nucleotide misincorporation during transcription due to template-strand misalignment. *Mol. Cell* **24**, 245–255 (2006).

50. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. *Proc. Natl Acad. Sci. USA* **109**, 8067–8072 (2012).

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

52. 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).

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

54. 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).

55. 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).

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

57. 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).

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

59. Lovett, S. T. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **52**, 1243–1253 (2004).

60. Lu, G. & Gong, P. Crystal structure of the full-length japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog.* **9** (2013).

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

62. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

Author contributions

I.I. and M.H. designed the study, analyzed the results and wrote the paper. I.I. and M.O. performed chemical syntheses, biochemistry, cross-linking experiments and RdRp assays. T.S.Q. performed IVT, PEX and LC-MS analyses. E.Z. and E. B. produced the JEV, YFV and SARS-CoV-2 RdRps. R.P. measured and characterized NMR spectra. M.V. performed proteomic analysis.

Competing interests

The authors declare no competing interests. Michal Hocek is a Guest Editor for Communications Chemistry's Nucleic Acid Chemistry Collection, but was not involved in the editorial review of, or the decision to publish this article.

Additional information

Supplementary information The online version contains supplementary material available at
https://doi.org/10.1038/s42004-024-01399-6.

Correspondence and requests for materials should be addressed to Michal Hocek.

Peer review information *Communications Chemistry* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at
http://www.nature.com/reprints