Reversible site-specific tagging of enzymatically synthesized RNAs using aldehyde–hydrazine chemistry and protease-cleavable linkers

Stephanie Pfander, Roberto Fiammengo, Srećko I. Kirin, Nils Metzler-Nolte and Andres Jäschke*

Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Im Neuenheimer Feld 364, 69120 Heidelberg, Germany

ABSTRACT

The investigation of RNA structure, dynamics and biological function often requires the site-specific incorporation of non-natural moieties. Here we describe the functionalization of RNA transcripts by aldehyde–hydrazine chemistry using a simple initiator nucleotide that carries an acetal-protected aldehyde function. This initiator nucleotide was efficiently incorporated into RNA, and the modified RNAs were quantitatively coupled to a peptide derivative displaying a hydrazine moiety at one end, a biotin tag at the other, and a trypsin-cleavable sequence in between. RNA conjugates could be easily isolated by affinity chromatography on streptavidin agarose and quantitatively cleaved off the support by trypsin treatment without detectable RNA degradation. The strategy described here may allow the incorporation of various new features into enzymatically synthesized RNA under mild conditions.

INTRODUCTION

The investigation of RNA structure, dynamics and biological function often requires the site-specific incorporation of non-natural moieties (labels, dyes, affinity tags, tethers, cross-linkers) (1–6). For short chemically synthesized RNAs this is a moderately complicated task and can be accomplished by using either the appropriate phosphoramidites during solid-phase synthesis or by post-synthetic derivatization (7). Longer RNAs are, however, typically synthesized by enzymatic transcription, and the site-specific modification of these molecules represents a considerable challenge. The only positions that are easily accessible are the two termini, while the specific modification of internal positions requires more complex approaches involving splinted ligations (8,9) or non-natural base pairs (4,10). At the 3′-end, modifications are commonly introduced by ligation of modified (oligo-)nucleotides (11,12) or by specific oxidation of the ribose, followed by derivatization of the resulting aldehyde (13). Chemically modified guanosine derivatives have been effectively used as initiator nucleotides during transcription to incorporate various functional groups at the 5′-end (14–19). However, the range of different modification strategies is still limited, and there is a particular need for new methods to introduce selectively addressable functionalities that are orthogonal to the commonly used amine–NHS ester and thiol–maleimide chemistries (20).

Aldehyde–amine and aldehyde–hydrazine coupling have successfully been used to modify oligonucleotides with many different functionalities, such as ligands for metal complexes, solid supports, amino-sugars or peptides (21,22). So far aldehyde groups have been selectively incorporated into long RNA transcripts only by post-transcriptional modification of a thiophosphate-modified RNA (23) or by periodate oxidation of the 3′-terminal ribose, which may result in unstable conjugates due to elimination reactions (24,25). In order to develop a simple aldehyde-based conjugation protocol we describe here the design and synthesis of an initiator nucleotide that carries a protected aldehyde function. We report the site-specific incorporation of this compound into RNA transcripts, including a 109 residue long RNA pool with 70 randomized positions, the efficient coupling of a biotinylated peptide using hydrazine chemistry, and the targeted release of the appended moieties by enzymatic cleavage (Figure 1).

MATERIALS AND METHODS

General procedures

All reagents used were purchased from Aldrich, Fluka, Novabiochem, Glen research (phosphoramidite **1**) or
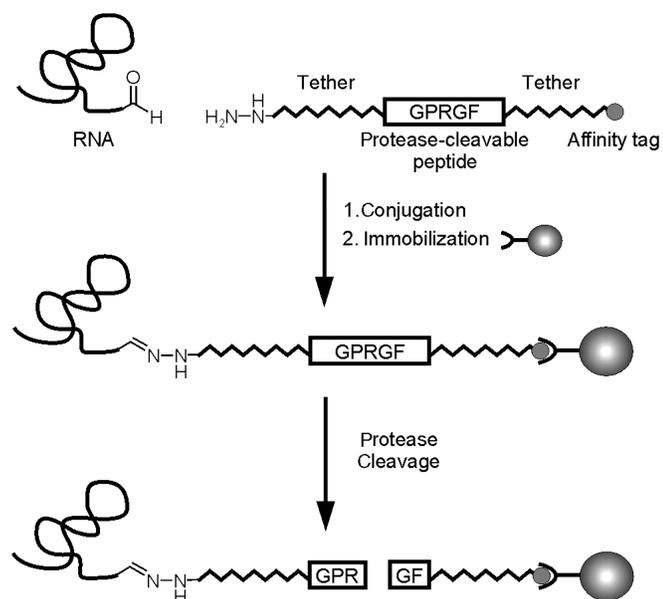
Figure 1. Site-specific tagging of RNA transcripts by hydrazone formation, immobilization of labelled RNA and subsequent proteolytic cleavage.

Acros Organics and used without further purification. Dry solvents (acetonitrile = ACN, tetrahydrofurane = THF) were purchased from Fluka (dry solvents over molecular sieves). Reversed phase thin layer chromatography (RP-TLC) analyses were carried out using RP-18 $F_{254 \ s}$ plates (20×20 cm aluminium sheets) from Merck. Ready-to-use Lobar® columns from Merck (stationary phase LiChroprep® RP-18) were used for preparative reversed phase chromatography. Cation exchange was performed using DOWEX® 50WX8-200 resin (sodium form). HPLC analyses were performed on an Agilent 1100 Series HPLC system equipped with an diode array detector using a Phenomenex® Luna 5U C18 column (4.6×250 mm) and eluting at 1 ml/min with a gradient of 100 mM triethyl-ammonium acetate (TEAA) pH 7.0 (buffer A) and 100 mM TEAA in 80% ACN (buffer B) for initiator nucleotide and RNA conjugates, and water/ACN both containing 0.1% TFA for peptides. NMR spectra were recorded on a Varian VNMR S 500 spectrometer. ^{1}H NMR spectra were calibrated to TMS on the basis of the relative chemical shift of the solvent as an internal standard. Abbreviations used are as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad signal. Matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra were recorded on a Bruker BIFLEX III spectrometer. Samples for MALDI-TOF measurements were dissolved in water to a final concentration of 10 µM. The samples for analysis were prepared using the dried droplet method with the following matrix solution: 3-hydroxy-picolinic acid/diammonium hydrogen citrate in 1.2 : 1 v/v water/acetonitrile (detection in positive mode). ESI mass spectra were measured on a Finnigan TSQ 700 instrument. Gel scans were acquired using a Typhoon 9400 Variable Mode

Imager (Amersham Biosiences). Radioactive samples were quantified using a Beckman Coulter LS 6500 Liquid Scintillation Counter.

Initiator nucleotide 3

$2',3'$-*Bis*-(t-butyldimethylsilyl)guanosine **2** (26,27) (0.102 g, 200 µmol, 2 equiv) was dried under high vacuum overnight and suspended in 1 ml dry ACN under argon. A solution of 2-[4-(5,5-diethyl-1,3-dioxan-2-yl)phenox-y]ethan-1-yl-[(2-cyanoethyl)-(N,N-diisopropyl)]-phos-phoramidite **1** (48 mg, 100 µmol) in 1 ml dry ACN was then added, followed by 0.124 g 4,5-dicyanoimidazole (DCI) as activator (1 mmol, 10 equiv). The reaction was stirred at room temperature for 10 min and then stopped by addition of 200 µl of t-butylhydroperoxide (1.6 mmol, 16 equiv). After 10 min stirring, all volatiles were removed under vacuum and the residue was taken up in 5 ml tetrabutylammonium fluoride (TBAF) solution (1 M in THF, 5 mmol, 50 equiv) under argon. After stirring overnight at r.t., full deprotection was obtained according to RP-TLC (ACN/H_2O 4 : 6). The THF was removed under vacuum and the residue was taken up with 3 ml of a 8 : 2 ACN/H_2O mixture, filtered through a 0.22 µm membrane filter, and purified by preparative RP-chromatography (elution from H_2O/ACN 9 : 1 to H_2O/ACN 1 : 1). The fractions containing the product were concentrated under vacuum to remove most of the ACN and lyophilized. Tetrabutylammonium cations were exchanged for sodium on an ion-exchange resin column (Ø = 1.5 cm, l = 14 cm, see general information) eluting with water. The eluate containing the product was lyophilized. An additional preparative RP-chromatography was finally carried out to remove excess salts. The product was obtained after lyophilization as pale yellow, hygroscopic solid (10 mg, 16.2 µmol, 16% yield). The purity of initiator nucleotide **3** was confirmed by HPLC (gradient: 5 → 40% buffer B in 25 min, 40 → 80% in 15 min, 80 → 100% B in 5 min).

^{1}H NMR (500 MHz, D_2O) δ 8.02 (s, 1H), 7.29 (m, 2H), 6.84 (m, 2H), 5.83 (d, J = 5.7 Hz, 1H), 5.44 (s, 1H), 4.63 (t, J = 5.7 Hz, 1H), 4.42 (t, J = 4.5 Hz, 1H), 4.29 (m, 1H), 4.15–4.07 (m, 6H), 3.97 (d, J = 11.4 Hz, 2H), 3.74 (d, J = 11.4 Hz, 2H), 1.78 (q, J = 7.5 Hz, 2H), 1.19 (q, J = 7.5 Hz, 2H), 0.91 (t, J = 7.5 Hz, 3H), 0.82 (t, J = 7.5 Hz, 3H). MALDI-TOF MS: m/z 626.2 $[M]^+$ (calculated for $[C_{26}H_{37} N_5O_{11}P]^+$ 626.22).

Preparation of aqueous solutions of initiator nucleotide 3

Initiator nucleotide **3** was dissolved in Milli-Q® water and filtered through a 0.22 µm membrane filter. Concentrations were determined by UV spectroscopy (Amersham Biosciences Ultrospec 2100 *pro* UV-vis spectrophotometer). Calibration curves were prepared using guanosine as model compounds for **3**. For guanosine, the absorbance at 260 nm of six aqueous solutions in the range 7.48–59.9 µM was measured. The coefficient of determination (R^2) for the calibration curves was 0,999.

Reaction of initiator nucleotide 5 (free aldehyde group) with N-nucleophiles

All HPLC analyses were performed using the gradient described for initiator nucleotide **3**.

(A) *Deprotection of the aldehyde group.* Initiator nucleotide **3** (20 µl, 40 nmol) was deprotected by addition of trifluoroacetic acid (TFA) (0.4 µl, 2% v/v final conc.). After 15 min the reaction mixture was quenched with 5.4 µl of a 1 M solution of $NaHCO_3$. The complete deprotection of the aldehyde group was confirmed by HPLC and by MALDI-TOF MS analysis (Table 1).

(B) *Reaction with benzyl amine.* One microlitre of a benzyl amine solution (0.4 M in 0.1 M phosphate buffer, pH 9.5, 10 equiv), 4 µl of a freshly prepared 1 M $NaCNBH_3$ solution and 10 µl of 0.1 M phosphate buffer at pH 9.5 were added to the deprotection reaction mixture prepared as described earlier. After 18 h, a second portion of benzyl amine solution (2 µl, 10 equiv) and of 1 M $NaCNBH_3$ solution (4 µl) were added and the reaction continued for an additional 10 h. HPLC analysis of the reaction mixture showed that the desired benzyl amine conjugate **6** was obtained in ~40% yield. The rest of the aldehyde **3** was reduced to the corresponding benzyl alcohol **9** as confirmed by MALDI-TOF MS (Table 1).

(C) *Reactions with benzoyl hydrazide and 6-hydrazino nicotinic acid.* These two reactions were performed at a 10 µM final concentration of initiator nucleotide **5** to closely reproduce the conditions used during the conjugation reactions with aldehyde-modified RNAs (see Preparation of RNA-peptide conjugates).

Initiator nucleotide **3** (40 nmol) was dissolved in 1.6 ml H_2O and deprotected by addition of 0.4 ml of 10% TFA. The reaction mixture was quenched by addition of 0.64 ml of a 1 M solution of $NaHCO_3$. Conjugation reactions were performed by adding 1.36 ml of 0.3 M sodium acetate buffer (pH 5.5) followed by 8 µl of 0.25 M solution of benzoyl hydrazide (50 equiv) or by 40 µl of 0.05 M solution of 6-hydrazino nicotinic acid (50 equiv). After stirring overnight at r.t the reactions were analysed by HPLC. Conjugation products **7** and **8** were isolated and the identity confirmed by MALDI-TOF MS (Table 1).

T7 transcriptions: Preparation of 25mer 5′-G*GA GCU CAG CCU ACG AGC CUG AGC C-3′ and of 109mer 5′-G*GA GCU CAG CCU UCA CUG C (N)70 GG CAC CAC GG UCG GAU CCA C-3′

The incorporation of initiator nucleotide **3** at the 5′-end during transcription is indicated by G*. The dsDNA template for the 25mer transcript was prepared as previously reported (16). The dsDNA template for the 109mer RNA pool was obtained after polymerase chain reaction (PCR) amplification of a chemically synthesized ssDNA-pool having the following sequence: 5′-GGA GCT

Table 1. MALDI TOF analysis of compounds **3–9**

Compound	Molecular formula	$[M+H]^+$ calculated	$[M+H]^+$ found
3	$C_{26}H_{36}N_5O_{11}P$	626.22	626.1
5	$C_{19}H_{22}N_5O_{10}P$	512.12	512.1
6	$C_{26}H_{31}N_6O_9P$	603.20	603.3
7	$C_{26}H_{28}N_7O_{10}P$	630.17	630.1
8	$C_{25}H_{27}N_8O_{11}P$	647.16	647.2
9	$C_{19}H_{24}N_5O_{10}P$	514.13	514.3

CAG CCT TCA CTG C (N70) GGC ACC ACG GTC GGA TCC AC-3′.

Transcription reactions (50 µl) were performed in a buffer where final concentrations were: 80 mM 4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid (HEPES) (pH 7.5), 22 mM $MgCl_2$, 1 mM spermidine, 10 mM dithithreitol (DTT), 40 µg/ml bovine serum albumin (BSA). The concentration of DNA template was 0.5 µM. ATP, CTP and UTP were 4 mM, GTP 0.4 mM and initiator nucleotide **3** 4 mM. For radioactive labelling ~40 µCi of α-^{32}P-CTP were added. The reactions were started by addition of T7 RNA-polymerase (Fermentas, final concentration 8 U/µl) and incubated at 37°C overnight. The transcription reactions were terminated by adding 0.3 M sodium acetate buffer (pH 5.5) up to 400 µl and 1 ml cold absolute ethanol. The precipitated RNA was isolated by centrifugation and purified by HPLC (gradient: 5% buffer B for 5 min; 5 → 85% in 40 min) with detection by both UV (260 nm) and a radioflow detector (LB 509, Berthold Technologies). The identity of the aldehyde-initiated 25mer RNA was confirmed by MALDI-TOF MS: m/z 8348.4 $[M+H]^+$ (calculated for $[C_{254}H_{322}N_{98}O_{177}P_{25}]^+$ 8350.26). It should be noted that transcription from such templates always yields significant amounts of $(n+1)$mers due to untemplated addition of one nucleotide.

H_2N–HNA–Peg7–Gly–Pro–Arg–Gly–Phe–Peg7–Lys(biotin)–Phe–NH_2 (4)

General procedure (28). The synthesis was performed manually in a syringe equipped with a porous filter (5 ml, MultiSynTech) using Fmoc-protected Rink amide resin (80 mg, 0.05 mmol, loading 0.63 mmol/g, Novabiochem), Fmoc-protected monomers (Novabiochem) and *N,N*-dimethylformamide (DMF, amine-free, Roth) as solvent. Arginine was introduced as Fmoc-Arg(Pbf)-OH. Hydrazine was introduced by 6-Fmoc-hydrazino nicotinic acid (HNA) (Novabiochem). Polyethylene glycol linker was introduced as O-[2-(Fmoc-amino)-ethyl]-O′-[2-(diglycolyl-amino)ethyl]hexaethylene glycol (Fluka).

Synthetic cycle. (A) *Fmoc-deprotection* 2 × ~3 ml of a 20% piperidine solution in DMF, first 2 min and then 10 min, no washing in between; (B) *DMF wash* (5 × 3 ml); (C) *coupling* using 2.2 ml of coupling mixture over 20 min (see later) and (D) *DMF wash* (5 × 3 ml).

The coupling mixture contained: the respective monomer (protected amino acid, 0.25 mmol, 5-fold excess), 2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU) (78.7 mg, 0.49 mmol, 4.9-fold

excess) and *N*-hydroxybenzotriazole (HOBt) (37.5 mg, 0.49 mmol, 4.9-fold excess) in 2 ml DMF. Subsequently ethyldiisopropylamine (DIPEA) (0.1 ml, 0.59 mmol, 11.7-fold excess) was added followed by an activation period of 1 min.

Workup. The resin was dried under vacuum, then the product was deprotected and cleaved over 3 h at r.t. using 3 ml of a mixture containing TFA: H_2O: TIS = 95: 2.5: 2.5 (TIS = triisopropylsilane). The suspension was filtered and the resin washed with TFA (2×1.5 ml). The combined TFA solutions were poured into cold diethyl ether (Et_2O) (10 ml, $-30°C$) and the crude product recovered by centrifugation. The obtained pellet was washed with cold Et_2O (2×10 ml), dissolved in water, filtered, lyophilized and finally purified by HPLC. Gradient: 5% ACN $\rightarrow$ 55% ACN in 30 min, $R_t = 15.9$ min; MS ESI: m/z 1051.4 $[M + 2H]^{2+}$, 701.4 $[M + 3H]^{3+}$ (calculated for $[C_{95}H_{155}N_{21}O_{30}S]^{2+}$ 1051.05, $[C_{95}H_{156}N_{21}O_{30}S]^{3+}$ 701.03).

Preparation of RNA-peptide conjugates

25mer RNA or 109mer RNA pool (50 pmol) were deprotected by treatment with 2% TFA for 15 min at 20°C. Conjugation reactions were performed by adding 1.6 μl of 1 M $NaHCO_3$, 2.4 μl 0.3 M sodium acetate (pH 5.5) and 1.0 μl of a 2.5 mM solution of peptide **4** (2.5 nmol, 50 equiv) to a final volume of 10 μl (conc. RNA = 5 μM). The mixture was allowed to react for 6 h at 25°C. For ^{32}P-labelled probes, analysis of the reaction mixtures and product isolation was carried out by gel electrophoresis. For non-^{32}P-labelled 25mer RNA, the reaction was stopped by ethanol precipitation and the conjugate was isolated by HPLC (gradient: 5% buffer B for 5 min; 5 $\rightarrow$ 85% in 40 min). After lyophilization, the conjugate was redissolved in water and analysed by MALDI-TOF MS: m/z 10326.7 $[M + H]^+$ (calculated for $[C_{342}H_{459}N_{119}O_{205}P_{25}S]^+$ 10318.23).

Immobilization of RNA-peptide conjugates

Preparation of agarose immobilized streptavidin. One hundred microlitres of resin suspension (Immobilized Streptavidin, Pierce, 15–28 μg biotin/ml) was transferred to a Mobicol spin filter (MoBiTech, 10 μm bottom filter), centrifuged (12 000 rpm, 1 min), suspended in 200 μl immobilization buffer (HEPES 10 mM, NaCl 1 M, EDTA 5 mM, pH 7.2) containing 2 mg/ml total tRNA, incubated for 10 min at 25°C, centrifuged and washed twice with 200 μl immobilization buffer.

Immobilization of the conjugates. ^{32}P-labelled 25mer RNA conjugate (after purification by gel electrophoresis; 25 000 cpm) and 109mer RNA conjugate (after ethanol precipitation; 150 000 cpm) were dissolved in 200 μl immobilization buffer, loaded on a Streptavidin resin pretreated as described earlier and incubated for 30 min at 25°C.

Washing procedure. The unbound RNA was removed by centrifugation (''washing fraction'' No. 0) and the resin was washed 3–5 times with 200 μl washing buffer

(Tris-HCl 50 mM, urea 8 M, pH 7.4) and 3–5 times with 200 μl water until the measured radioactivity in the washing fractions reached background (<100 cpm, for the 25mer conjugate this corresponds to washing fractions 1–5, for the 109mer conjugate, to 1–6 in Figure 6a).

Stability test of immobilized conjugates. The resin was suspended in 200 μl immobilization buffer and washing steps as described earlier were repeated for the 25mer conjugate after 60 min and after overnight incubation (No. 6–10 and 11–18) and for the 109mer conjugate after overnight incubation (No. 7–13). In all cases the measured radioactivity in the flow-through was at background level (Figure 6a).

Protease digestion. The resin was suspended in 200 μl NH_4HCO_3 buffer (100 mM, pH 8.5), 6 μl trypsin (proteomics grade, Sigma, 1 mg/ml in 1 mM HCl, 1 : 16 (w/w) enzyme to substrate ratio) was added, and the mixture was incubated at 37°C for 4 h. The supernatant was removed by centrifugation and found to contain 85–90% of the immobilized radioactivity (for the 25mer conjugate fraction 19, for 109mer conjugate fraction 14). For the 109mer conjugate, additional washing steps were performed as described earlier (fractions 15–18), confirming that practically all RNA was released from the resin.

Analysis of digestion products. For gel electrophoretic analysis the supernatant obtained after digestion was lyophilized and resuspended in water before loading on 12 or 18% denaturing PAGE gels. The gels were run over a distance of 20 cm (1000 V).

Protease digestion of the 25mer RNA-peptide conjugate in solution

For the MALDI MS measurement of the digestion product, unlabelled purified 25mer RNA-peptide conjugate (300 pmol, 6 μM) was dissolved in 50 μl NH_4HCO_3 buffer (100 mM, pH 8.5) trypsin (0.1 mg/ml in 1 mM HCl, 1 : 20 (w/w) enzyme to substrate ratio) was added and the mixture was incubated at 37°C for 4 h. The enzyme was removed by extraction with the same volume of a phenol/chloroform/isoamyl alcohol 25 : 24 : 1 and subsequent ethanol precipitation. The digestion product was isolated by HPLC (gradient: 5% buffer B for 5 min; 5 $\rightarrow$ 85% in 40 min), lyophilized, redissolved in water and analysed by MALDI-TOF MS: m/z 9141.4 $[M + H]^+$ (calculated for $[C_{286}H_{373}N_{109}O_{190}P_{25}]^+$ 9147.63).

RESULTS AND DISCUSSION

General concept

The goal of this work was to develop a robust strategy that utilizes aldehyde groups for the mild conjugation of enzymatically synthesized RNA with larger, chemically synthesized peptides that may potentially carry various chemical modifications (Figure 1). This conjugation approach may serve to attach affinity tags, localization signals, proteolytic cleavage sites or other functions to RNA molecules, and thereby enable their directed delivery
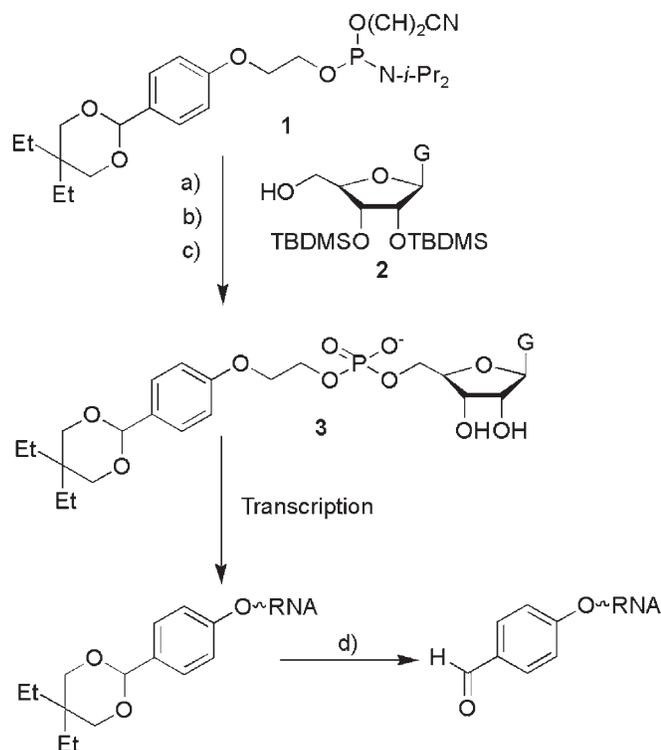
Figure 2. Synthetis of acetal-protected aldehyde–guanosine phosphate **3**. Reagents and conditions: (a) **2** (2.0 equiv), DCI (10 equiv), CH_3CN, room temperature, 10 min; (b) TBHP (16 equiv), room temperature, 10 min, volatiles removal; (c) TBAF (50 equiv), THF, room temperature, overnight, 16% yield over three steps; (d) 2% TFA in water, room temperature, 15 min, complete deprotection.

or triggered release. This concept uses aldehyde-derivatized initiator nucleotides that are selectively incorporated at the 5′-end of RNA transcripts. As a proof of concept, a complex multifunctional peptide derivative that allows conjugation, affinity purification, and proteolytic release of the RNA transcript was designed.

Design, synthesis and characterization of the building blocks

Initiator nucleotide **3** was designed and synthesized to introduce an aldehyde function into RNA transcripts according to Figure 2. Commercial phosphoramidite **1** (22) was therefore coupled with readily available 2′,3′-O-silyl-protected guanosine **2** (26,27) by activation with 4,5-dicyanoimidazole, followed by oxidation and deprotection, leaving the acetal protecting group intact. After purification by reversed phase chromatography, **3** was isolated in 16% total yield, and the purity and identity established by HPLC and MALDI mass spectrometry.

In principle, **3** could now be deprotected to unmask the aldehyde function, and the resulting product then be used in transcription. However, as aldehyde-modified nucleotides have been reported to react with active site nucleophiles of polymerases under Schiff base formation (29), we decided to use protected **3** as substrate for the enzyme, and deprotect after transcription.

To establish the conjugation chemistry with rigorous analytics, a small amount of **3** was deprotected to give **5** and then reacted with various nucleophiles (Figure 3). The reaction mixtures were analysed by HPLC and the identity of the products confirmed by mass spectrometry (Table 1). Benzyl amine was found to give stable reaction product **6** only under reductive amination conditions in the presence of sodium cyanoborohydride. Due to the reducing conditions, **5** was also converted to the corresponding benzylalcohol derivative **9**, limiting the total yield for the conjugation reaction (see Supplementary Data). In contrast, benzoyl hydrazide and 6-hydrazino-nicotinic acid yielded quantitatively the respective coupling products (**7** and **8**) under non-reducing conditions. Therefore we chose a hydrazine-modified peptide for efficient tagging of *in vitro* transcribed RNA.

Peptide derivative **4** carries a hydrazine moiety (6-hydrazino-nicotinic acid, HNA) at one end, a biotin tag at the other and a trypsin-cleavable sequence in between, all connected via heptaethylene glycol spacers (PEG_7): HNA–PEG_7–Gly–Pro–Arg–Gly–Phe–PEG_7–Lys(biotin)–Phe (exact structural formula: see Supplementary Figure S4). The molecule was efficiently assembled by Fmoc solid phase peptide synthesis (SPPS) using commercially available monomers, purified by HPLC and characterized by mass spectrometry.

Enzymatic incorporation of the initiator nucleotide

Initiator nucleotide **3** was added to standard transcription reactions yielding 25mer and 109mer transcripts, the latter actually being a combinatorial library with a randomized 70nt domain in the center. As the addition of the protected aldehyde function had almost no effect on the electrophoretic mobility, transcriptions were analyzed by HPLC, where the aromatic residue caused an increase in retention time by 7.5 min (25mer) and 3 min (109mer), respectively (Figure 4). At the optimized ratio of **3** to competing guanosine triphosphate (GTP) of 10:1 (final conc. **3** = 4.0 mM), 70–75% of all RNA molecules were found to be 5′-end modified both for a 25mer and for a 109mer.

Conjugates were then deprotected by a short treatment with 2% aqueous trifluoroacetic acid (TFA). This was found to leave the RNA essentially intact, as primer elongation studies by reverse transcriptase on the TFA-treated 109mer pool showed no reduction in the amount of full-length product in comparison to the untreated reference (see Supplementary Material).

Conjugation reaction and characterization of the conjugates

The deprotected aldehyde-modified RNAs were coupled to **4** within 6 h at 25°C using a 50-fold excess of peptide. The reaction mixtures were analysed by gel electrophoresis and MALDI-TOF MS. The RNA was found to be quantitatively conjugated as shown by the significant reduction in electrophoretic mobility caused by the attachment of the peptide (Figure 5).

Due to the presence of the biotinyl group, the RNA conjugates interacted efficiently with streptavidin and
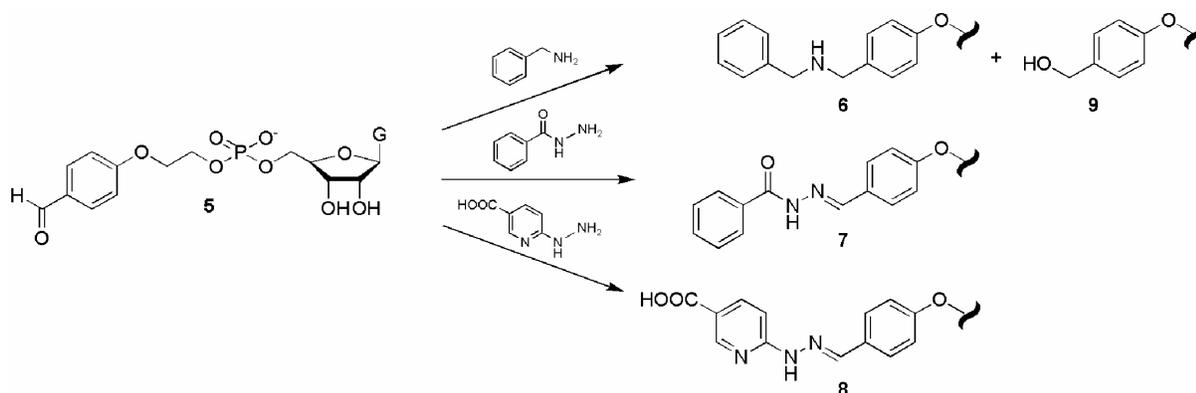
Figure 3. Model reactions performed to establish the conjugation strategy. Benzyl amine, benzoyl hydrazide and 6-hydrazino-nicotinic acid were coupled to unprotected initiator nucleotide **5.** The conjugation with benzyl amine under reductive conditions resulted in the benzyl alcohol derivative **9** as side product.
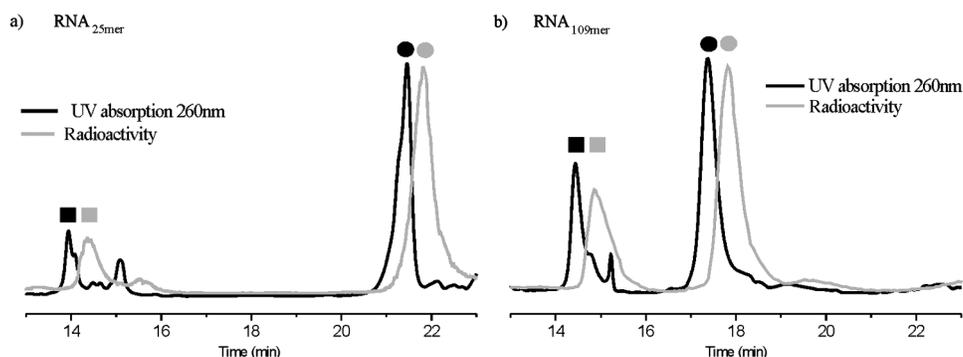
Figure 4. Selective modification of RNA. Incorporation of **3** into a 25mer (a) and 109mer pool (b) RNA by *in vitro* transcription. Transcription mixtures were analysed by HPLC. UV absorption (black), radioactivity (grey); radioactivity traces are ∼1 min shifted due to the connection configuration of the radioflow detector. (black squares) Non-initiated transcripts; (black circles) aldehyde-initiated transcripts.
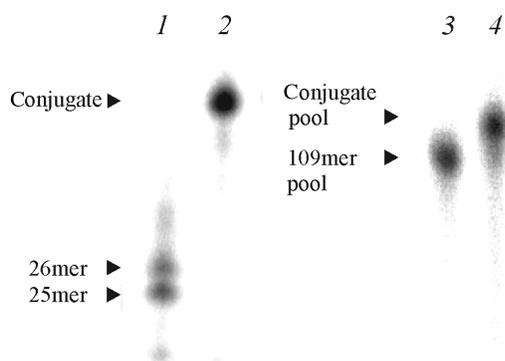
Figure 5. Conjugation of aldehyde-derivatized 25mer (left) and 109mer pool (right) RNAs (^{32}P-labelled) with **4**. Electrophoretic analysis of conjugation mixtures. Lane *1* and *3*: reaction mixtures containing 10 μM aldehyde-RNA$_{25mer}$ or aldehyde-RNA$_{109mer}$ without peptide **4**; lane *2* and *4*: reaction mixtures containing 10 μM aldehyde-RNA$_{25mer}$ or aldehyde-RNA$_{109mer}$ plus 50 equiv of **4**. In lane *1* 25mer and 26mer transcripts produced by T7 RNA polymerase are resolved into two bands.

could be immobilized on streptavidin agarose. Immobilized ^{32}P-labelled RNA-peptide conjugates were shown to be stable over 24 h, as no radioactivity was observed in the eluates during successive washing of the resin, and the amount of radioactivity immobilized on the resin remained constant during this time (Figure 6a). While the immobilization experiment with the 25mer was done with HPLC-purified conjugate and resulted in near-quantitative immobilization (left panel), the 109mer sample was conjugated following a quick protocol that uses the raw conjugation mixture without purification. Therefore, a lower amount of ^{32}P got immobilized in that latter case (right panel).

The protease cleavage site was found to be readily accessible to the enzyme when immobilized to streptavidin agarose. Eighty-five to ninety percent of the immobilized RNA was released from the resin by treatment with trypsin (proteomics grade, 4 h at 37°C, Figure 6a). Gel-electrophoretic analysis of the eluate revealed one sharp band with a mobility between the unmodified RNA and the RNA-peptide conjugate (Figure 6b), indicating that trypsin cleavage occurred specific without side reactions, and in particular, without detectable RNA degradation. Additionally, the digestion product of the 25mer RNA-conjugate was isolated by HPLC, characterized by mass spectrometry and found to have the correct mass for the expected cleavage at the C-terminal side of the arginine residue.
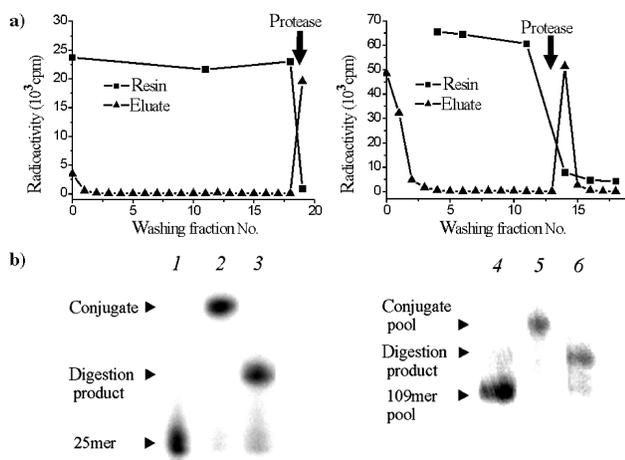
Figure 6. Immobilization and stability of ^{32}P-labelled RNA-peptide conjugates. (a) Purified RNA$_{25mer}$ conjugate (left) and crude conjugation reaction mixture of RNA$_{109mer}$ conjugate (right) were immobilized on streptavidin resin; the resin was washed until background levels of radioactivity were reached; and stability over 24 h in aqueous buffer at 25°C was investigated by monitoring the radioactivity in the washing fractions (black triangles) and on the resin (black squares). Proteolytic cleavage with trypsin is indicated by the arrows. (b) Gel electrophoretic analysis of digestion products: (*1*) RNA$_{25mer}$, (*2*) RNA$_{25mer}$ conjugate, (*3*) digestion product (eluted in fraction No. 19), (*4*) RNA$_{109mer}$, (*5*) RNA$_{109mer}$ conjugate, (*6*) digestion product (eluted in fraction No. 14).

CONCLUSION

In conclusion we have developed a simple and reliable strategy for the temporary tagging of long RNAs at their 5′-ends, where temporary refers to the possibility of removing the introduced tags by proteolytic cleavage. This strategy is potentially versatile since practically any RNA sequence can be synthesized by *in vitro* transcription, and an almost infinite variety of peptide sequences carrying linkers and tags is accessible by solid phase synthesis. Aldehyde-based modification strategies have already been shown to be compatible with a variety of other postsynthetic derivatization schemes, and the direct enzymatic incorporation of an aldehyde group represents an orthogonal method to modify long RNAs that enriches the portfolio of conjugation and modification strategies (30). While our main motivation for this work is the use of such conjugates in the isolation of ribozymes with protease activity from combinatorial RNA–peptide conjugate libraries, the strategy described here may find future applications for the delivery of functional RNA molecules to specific cellular locations (31) by attaching suitable localization signals or affinity tags, and subsequent release by cellular enzymes. Furthermore, it may provide a tool for the isolation of RNA-binding proteins (32), followed by mild enzymatic release of the complexes.

ACKNOWLEDGEMENTS

REFERENCES

1. Meisenheimer,K.M., Meisenheimer,P.L. and Koch,T.H. (2000) Nucleoprotein photo-cross-linking using halopyrimidine-substituted RNAs. *Methods Enzymol.*, **318**, 88–104.
2. Keppler,B.R. and Jarstfer,M.B. (2006) A high-throughput assay for a human telomerase protein-human telomerase RNA interaction. *Anal. Biochem.*, **353**, 75–82.
3. Ujvari,A. and Luse,D.S. (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol. Biol.*, **13**, 49–54.
4. Di Giusto,D.A., Wlassoff,W.A., Giesebrecht,S., Gooding,J.J. and King,G.C. (2004) Enzymatic synthesis of redox-labeled RNA and dual-potential detection at DNA-modified electrodes. *Angew. Chem. Int. Ed.*, **43**, 2809–2812.
5. Blount,K.F. and Tor,Y. (2003) Using pyrene-labeled HIV-1 TAR to measure RNA-small molecule binding. *Nucleic Acids Res.*, **31**, 5490–5500.
6. Kurreck,J. (2003) Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.*, **270**, 1628–1644.
7. Venkatesan,N., Kim,S.J. and Kim,B.H. (2003) Novel phosphor-amidite building blocks in synthesis and applications toward modified oligonucleotides. *Curr. Med. Chem.*, **10**, 1973–1991.
8. Kurschat,W.C., Müller,J., Wombacher,R. and Helm,M. (2005) Optimizing splinted ligation of highly structured small RNAs. *RNA*, **11**, 1909–1914.
9. Moore,M.J. (1999) Joining RNA molecules with T4 DNA ligase. *Methods in Molecular Biology*. Humana Press, Totowa, New Jersey, **118**, 11–19.
10. Tor,Y. and Dervan,P.B. (1993) Site-specific enzymic incorporation of an unnatural base, N6-(6-aminohexyl)isoguanosine, into RNA. *J. Am. Chem. Soc.*, **115**, 4461–4467.
11. England,T.E. and Uhlenbeck,O.C. (1978) 3′-Terminal labeling of RNA with T4 RNA ligase. *Nature*, **275**, 560–561.
12. Hausch,F. and Jäschke,A. (1997) Libraries of multifunctional RNA conjugates for the selection of new RNA catalysts. *Bioconjug. Chem.*, **8**, 885–890.
13. Proudnikov,D. and Mirzabelkov,A. (1996) Chemical methods of DNA and RNA fluorescent labeling. *Nucleic Acids Res.*, **24**, 4535–4542.
14. Pitulle,C., Kleineidam,R.G., Sproat,B. and Krupp,G. (1992) Initiator oligonucleotides for the combination of chemical and enzymic RNA synthesis. *Gene*, **112**, 101–105.
15. Huang,F. (2003) Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res.*, **31**, e8/1–e8/8.
16. Fiammengo,R., Musilek,K. and Jäschke,A. (2005) Efficient preparation of organic substrate-RNA conjugates via in vitro transcription. *J. Am. Chem. Soc.*, **127**, 9271–9276.
17. Zhang,L., Sun,L., Cui,Z., Gottlieb,R.L. and Zhang,B. (2001) 5′-sulfhydryl-modified RNA: initiator synthesis, in vitro transcription, and enzymatic incorporation. *Bioconjug. Chem.*, **12**, 939–948.
18. Fusz,S., Eisenführ,A., Srivatsan,S.G., Heckel,A. and Famulok,M. (2005) A ribozyme for the aldol reaction. *Chem. Biol.*, **12**, 941–950.
19. Schlatterer,J.C. and Jäschke,A. (2006) Universal initiator nucleotides for the enzymatic synthesis of 5′-amino- and 5′-thiol-modified RNA. *Biochem. Biophys. Res. Commun.*, **344**, 887–892.
20. Hermanson,G.T. (ed.) (1996) *Bioconjugate Techniques*. San Diego: Academic Press.
21. Zatsepin,T.S., Stetsenko,D.A., Gait,M.J. and Oretskaya,T.S. (2005) Use of carbonyl group addition-elimination reactions for synthesis of nucleic acid conjugates. *Bioconjug. Chem.*, **16**, 471–489.
22. Podyminogin,M.A., Lukhtanov,E.A. and Reed,M.W. (2001) Attachment of benzaldehyde-modified oligodeoxynucleotide probes to semicarbazide-coated glass. *Nucleic Acids Res.*, **29**, 5090–5098.
23. Tsukiji,S., Pattnaik,S.B. and Suga,H. (2004) Reduction of an aldehyde by a NADH/Zn2+-dependent redox active ribozyme. *J. Am. Chem. Soc.*, **126**, 5044–5045.
24. Timofeev,E.N., Kochetkova,S.V., Mirzabekov,A.D. and Florentiev,V.L. (1996) Regioselective immobilization of short oligonucleotides to acrylic copolymer gels. *Nucleic Acids Res.*, **24**, 3142–3148.

25. Fraenkel-Conrat,H. and Steinschneider,A. (1968) Stepwise degradation of RNA; periodate followed by aniline cleavage. *Methods Enzymol.*, **12**, 243–246.
26. Kotch,F.W., Sidorov,V., Lam,Y.-F., Kayser,K.J., Li,H., Kaucher,M.S. and Davis,J.T. (2003) Water-mediated association provides an ion pair receptor. *J. Am. Chem. Soc.*, **125**, 15140–15150.
27. Zhu,X.-F., Williams,H.J. and Scott,A.I. (2000) Facile and highly selective 5′-desilylation of multi-silylated nucleosides. *Perkin*, **1**, 2305–2306.
28. Kirin,S.I., Noor,F., Metzler-Nolte,N. and Mier,W. (2007) Manual solid-phase peptide synthesis of metallocene–peptide bioconjugates. *J. Chem. Educ*, **84**, 108–111.
29. Riva,M., Schaffner,A.R., Sentenac,A., Hartmann,G.R., Mustaev,A.A., Zaychikov,E.F. and Grachev,M.A. (1987) Active site labeling of the RNA polymerases A, B, and C from yeast. *J. Biol. Chem.*, **262**, 14377–14380.
30. Olejnik,J., Ludemann,H.C., Krzymanska-Olejnik,E., Berkenkamp,S., Hillenkamp,F. and Rothschild,K.J. (1999) Photocleavable peptide-DNA conjugates: synthesis and applications to DNA analysis using MALDI-MS. *Nucleic Acids Res.*, **27**, 4626–4631.
31. Kubo,T., Zhelev,Z., Bakalova,R., Ohba,H., Doi,K. and Fujii,M. (2005) Controlled intracellular localization and enhanced antisense effect of oligonucleotides by chemical conjugation. *Org. Biomol. Chem.*, **3**, 3257–3259.
32. Carlson,C.B., Stephens,O.M. and Beal,P.A. (2003) Recognition of double-stranded RNA by proteins and small molecules. *Biopolymers*, **70**, 86–102.