An on-bead tailing/ligation approach for sequencing resin-bound RNA libraries

Anna Wiesmayr, Pierre Fournier and Andres Jäschke*

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

ABSTRACT

Nucleic acids possess the unique property of being enzymatically amplifiable, and have therefore been a popular choice for the combinatorial selection of functional sequences, such as aptamers or ribozymes. However, amplification typically requires known sequence segments that serve as primer binding sites, which can be limiting for certain applications, like the screening of on-bead libraries. Here, we report a method to amplify and sequence on-bead RNA libraries that requires not more than five known nucleotides. A key element is the attachment of the starting nucleoside to the synthesis resin via the nucleobase, which leaves the 3′-OH group accessible to subsequent enzymatic manipulations. After split-and-mix synthesis of the oligonucleotide library and deprotection, a poly(A)-tail can be efficiently added to this free 3′-hydroxyl terminus by *Escherichia coli* poly(A) polymerase that serves as an anchored primer binding site for reverse transcription. The cDNA is joined to a DNA adapter by T4 DNA ligase. PCR amplification yielded single-band products that could be cloned and sequenced starting from individual polystyrene beads. The method described here makes the selection of functional RNAs from on-bead RNA libraries more attractive due to increased flexibility in library design, higher yields of full-length sequence on bead and robust sequence determination.

INTRODUCTION

Over the past decades, nucleic acids have been shown to exhibit unique properties that depend on their folding into 3D structures, rather than on the primary sequence information. Such functional nucleic acids can either occur naturally, for example, as modulators of transcription or translation (riboswitches), or have been selected artificially as binding partners for small molecules or proteins (aptamers) or to catalyse reactions (ribozymes) (1–3). A broad range of potential applications in diagnostics, biotechnology and therapy has stimulated the search for such active nucleic acids, commonly performed via SELEX (Systematic Evolution of Ligands by Exponential Enrichment), where iterative enrichment of nucleic acids with tight binding properties or high catalytic activity is achieved over several rounds of selection (4–6). SELEX has been tremendously successful as it allows working with libraries of enormous complexity (typically 10^{14}–10^{15} different sequences), but it has some shortcomings: it is practically impossible to site-specifically incorporate individual chemical modifications; and only very few physical or chemical properties can be queried by the selection event.

Over the past years, alternative strategies have been developed to optimize the isolation process of active nucleic acids (7–10). One interesting approach exploits the split-and-mix technique, a well-known procedure from the field of combinatorial peptide chemistry (11), to generate OBOC (one-bead one-compound) libraries, in which each bead carries multiple copies of a single sequence. Yang *et al.* were the first to report on-bead screening of OBOC libraries for the isolation of oligonucleotide aptamers. These authors created on-bead DNA libraries in which specific phosphodiester groups were replaced by phosphorothioates and phosphorodithioates, and isolated an efficient aptamer against a transcription factor (12). Recently, our group reported the isolation of a fluorescence-enhancing RNA tag out of an on-bead OBOC RNA library (13). In addition to the inclusion of chemical modifications, the advantage of on-bead screening in comparison to SELEX lies in the possibility for direct read-out: while SELEX selects only for binding strength or catalytic attachment or detachment to/from a matrix, on-bead screening can directly report measurable physical properties like a fluorescence, luminescence or polarization signal. This makes the technique suitable for high-throughput screening: over a short period of time one can screen large libraries for a specific

---

*To whom correspondence should be addressed. Tel: +49 6221 54 4851; Fax: +49 6221 54 6430; Email: jaeschke@uni-hd.de

property, e.g. by making use of a FACS instrument. At lower throughput, even mass-spectrometric or chromatographic screening is possible, allowing for sophisticated screening criteria (14,15).

However, on-bead screening of OBOC oligonucleotide libraries, as well as SELEX, is dependent on the presence of conserved segments flanking the randomized region, serving as primer binding sites for amplification and sequence determination of isolated positive hits. As the total length of the RNA strands is limited by the coupling efficiency of solid phase synthesis, extended conserved regions restrict the number of variable nucleotides that can be incorporated, limit the size and—most importantly—bias the structure of the library. Several groups already described the severe limitations that are imposed by the introduction of fixed segments in a library sequence, including biased selection or difficulties with the truncation of isolated motifs. For example, Legiewicz *et al.* (16) showed that a pre-formed stable stem within fixed flanking sequences had a 5- to 10-fold negative effect on apparent motif abundance at all lengths during selection. Furthermore, primer binding sites have often been found to be involved in target recognition (17). In the two examples cited (12,13), the constant regions were even larger in size than the variable ones. Whereas various approaches have been proposed to shorten or remove primer binding sites in the context of SELEX (16,18–21), there is still a need for complementary methods using on-bead libraries that do not rely on extended constant regions.

Here, we describe a method for sequence determination of RNA on-bead libraries with the number of required constant nucleotides reduced down to five, thereby enabling the synthesis of almost completely unbiased libraries. This method uses a resin that leaves the 3′-OH group accessible for post-synthetic, post-screening enzymatic modifications. On-bead tailing, reverse transcription and adapter ligation allow amplification and sequence determination from individual resin beads.

MATERIALS AND METHODS

All reagents were purchased from Sigma-Aldrich, Roth or Proligo and used without further purification. Flash chromatography was carried out on silica gel 40–63 μm from J.T. Baker. NMR spectra were recorded on a Varian Mercury Plus 300 MHz spectrometer. ESI mass spectra were recorded on a Bruker micrOTOF-QII. Oligonucleotide synthesis was performed on an Expedite[TM] 8909 automated synthesizer using standard reagents and phosphoramidite chemistry [2′-O-TBDMS RNA Phosphoramidites, fast deprotection chemistry with tert-butylphenoxyacetyl (TAC)] from Sigma Aldrich Proligo. Deblock solution (3% v/v TCA in dichloromethane) was purchased from Roth. Unmodified primer and adapter oligonucleotides were purchased from IBA Göttingen. High-resolution agarose (Invitrogen) gels were stained with ethidium bromide and visualized by UV illumination using an AlphaImager[TM] 2200. The GeneRuler Ultra Low Range base pair ladder (Fermentas)

was used as a size marker. Absorption measurements were done with a Spectrophotometer Ultrospec 2100 pro (Amersham Biosciences). All bead manipulations were performed with a 2-μl pipette (Abimed) using a Nikon SMZ 1500 microscope.

Synthesis of 5′-O-(4,4′-dimethoxytrityl)-2′,3′-O-bis-(t-butyldimethylsilyl)-4-(1,2,4-triazolo)uridine 2

2′,3′,5′-protected uridine **1** (22,23) (300 mg, 388 μmol) and 1,2,4-triazole (428 mg, 6.20 mmol) were suspended in anhydrous acetonitrile (3.5 ml) and triethylamine (1.23 ml, 8.91 mmol) under argon, and after stirring for 5 min a clear, colourless solution was obtained. The flask was put at 0°C and POCl$_3$ (71.0 μl, 776 μmol) was added dropwise, whereupon a white solid precipitated. The ice bath was removed and the reaction mixture stirred overnight at room temperature. After the TLC control showed full conversion, ethyl acetate (100 ml) was added and the reaction mixture was washed with a 5% sodium bicarbonate solution (2 × 100 ml) and brine (2 × 100 ml), dried over anhydrous sodium sulphate and the solvents removed *in vacuo*. The raw product was purified by column chromatography (ethyl acetate/hexanes/triethylamine 4:6:1) to yield 295 mg (92%) of the title compound as a colourless foam.

^{1}H NMR (300 MHz; CDCl$_3$) δ 9.26 (s, 1H), 9.09 (d, 1H), 8.09 (s, 1H), 7.43–7.24 (m, 9H), 6.91–6.85 (m, 4H), 6.50 (d, J = 7.2 Hz, 1H), 5.85 (s, 1H), 4.36–4.19 (m, 3H), 3.94 (dd, J = 2.2 Hz, 11.1 Hz, 1H), 3.83 (s, 6H), 3.45 (dd, J = 1.4, 11.1 Hz, 1H), 0.94 (s, 9H), 0.75 (s, 9H), 0.34 (s, 3H), 0.19 (s, 3H), 0.00 (s, 3H), −0.09 (s, 3H). ^{13}C NMR (75.5 MHz; CDCl$_3$) δ 159.13, 158.80, 154.34, 153.76, 147.49, 143.40, 143.02, 134.84, 134.81, 130.36, 130.33, 128.63, 127.86, 113.14, 113.12, 94.41, 92.24, 87.28, 81.81, 75.81, 69.03, 60.16, 55.14, 25.77, 25.62, 17.97, 17.80, −4.00, −4.19, −5.25, −5.36. ESI MS: m/z 826.3 [M + H]$^+$ (calculated for [C$_{44}$H$_{60}$N$_5$O$_7$Si$_2$]$^+$ 826.4).

Preparation of the pre-loaded resin: coupling of 2 to amino-functionalized Tentagel

Compound **2** was dissolved in DMSO to a concentration of 330 mM (stock **2**). Ten milligrams (2.2 μmol) of Tentagel M NH$_2$ (Rapp Polymere; bead size 10 μm, loading capacity 220 μmol/g, 1.95×10^9 beads/g) and Tentagel N NH$_2$ (Rapp Polymere; bead size 90 μm, loading capacity 220 μmol/g, 2.86×10^6 beads/g) were loaded each into one eppendorf tube (1.5 ml) and swollen in DMF overnight. After spin filtration, the swollen beads were transferred to stock **2** diluted with DMSO (final concentration of **2** corresponds to 1, 2 and 10 eq. with respect to the amino groups on the resin) in a final volume of 100 μl. Addition of DMF to a 6% v/v concentration allowed for proper suspension of the beads. The coupling reaction was performed at 40°C for 10 h in a thermoshaker (Eppendorf) at 700 rpm. The beads were then transferred into a TWIST column (Glen Research, 1 μmol scale) and washed with 1 ml of following solvents (with a 2-ml syringe): DMSO, DMF, dichloromethane and acetonitrile. The extensive washing procedure is performed to quantitatively remove all

unreacted **2**. Notably, the 10 μm beads were more prone to agglomeration than the 90 μm beads and retained the chemicals more strongly; therefore, the volume of solvent during the washing procedure had to be doubled for the smaller beads. The unreacted amino groups were blocked by performing a capping step (5 min) with *tert*-butylphenoxyacetic anhydride (CAP A/CAP B, Proligo) on the automated synthesizer. The coupling yield of **2** to the resin was determined by absorption measurement at 498 nm of the cleaved DMT group of **2** after performing a deblock step (3% TCA in dichloromethane, 5 min).

RNA synthesis on pre-loaded resin

The pre-loaded resin was subjected to RNA synthesis (1 μmol scale) on an automated synthesizer (Expedite 8909). Since the Tentagel resin is swellable, the synthesis protocol for automated oligonucleotide synthesis differs significantly from conventional protocols used for the rigid beads of CPG or non-swellable polystyrene. The following changes to a conventional RNA synthesis protocol for CPG have been done: The resin was allowed to swell in DMF overnight prior to oligonucleotide synthesis. The amount of acetonitrile for all washing steps was doubled. The amount of deblock solution was retained to prevent depurination; however, the exposure time was increased 1.3-fold. The coupling reaction was performed for 10 min (recommended coupling time by Proligo when using 5-(3,5-*bis*(trifluoromethyl)phenyl)-1H-tetrazole (Activator 42) as an activator: 6 min), the volume of activator and monomer was increased 1.5-fold. We used two capping steps with an oxidizing step in between. In total, the volume of capping reagent was increased 1.4-fold, the exposure time 1.8-fold. The reaction time for the oxidation step was increased 1.6-fold, the delivered volume 1.2-fold. An additional washing step with anhydrous DMF was used after the oxidation step to thoroughly free the resin from remaining water. We achieved >99% coupling efficiency for a 30-mer RNA on 10 μm Tentagel beads. However, for 90 μm beads we had to further optimize the synthesis protocol, due to longer diffusion times into the larger beads: the volume of all reagents was retained, but all reaction times were increased by 10–20%. All coupling efficiencies were determined by the dimethoxytrityl cation assay (manual collection of cleaved DMT group after first and last coupling step and absorption measurement at 498 nm).

Deprotection was carried out in a 3:1 mixture of concentrated aqueous ammonia and EtOH for 2 h at room temperature, followed by thorough washing of the beads with EtOH/acetonitrile/water 3:1:1. The TBDMS protecting groups were removed by treatment with 1 M TBAF in THF for 24 h at room temperature, followed by thorough washing of the beads with EtOH/acetonitrile/water 3:1:1. The beads were dried under vacuum.

Poly(A)-tailing with *Escherichia coli* poly(A) polymerase

One single bead (pre-swollen in water for 3 h) was manually picked and transferred to 20 μl of the Poly(A)-tailing reaction mixture prepared according to the manufacturer (New England Biolabs). The reaction was carried out at 37°C for 20 min with subsequent heat inactivation of the enzyme at 65°C for 10 min.

Reverse transcription with anchored poly(T) primer

The RT reaction (Superscript II, Invitrogen) was carried out in the crude reaction mixture from the polyadenylation after addition of following components to a final volume of 30 μl: 1.5 μl of anchored poly(T) primer (5′-T_{23}GC-3′) (10 μM), 1.5 μl of dNTP mixture (10 mM each): 65°C, 5 min, quick chill on ice. Addition of 3 μl first strand buffer (5×), 3 μl DTT (100 mM): 42°C, 2 min. Addition of 1 μl Superscript II reverse transcriptase (200 U/μl): 50°C, 60 min. All steps were carried out in a PTC 100 (Biozym) cycler. The final concentrations of buffer ingredients in the one-pot reaction were as follows: Tris–HCl 58.3 mM, KCl 37.5 mM, $MgCl_2$ 8.2 mM, DTT 10 mM, NaCl 167 mM. Heat inactivation of the enzyme was omitted to prevent denaturation of the RNA/cDNA hybrid. The bead was manually picked out of the reaction solution and washed several times with water and transferred in a PCR tube containing 6 μl water.

Adapter ligation

After transferring the bead to a PCR tube containing 6 μl water, denaturation of the RNA/cDNA hybrid was performed at 90°C for 5 min, followed by a quick chill on ice. For the adapter ligation, following components were added to a final volume of 10 μl: 0.1 μl double-stranded adapter (3′-CCCAAATCACTCCCAATTATTCGCCGG CG-5′, 5′P-TTTAGTGAGGGTTAATAAGCGGCCGC GTCGTGACTGGGAGCGC-3′) (50 μM), 1 μl T4 DNA ligase buffer (Fermentas), water, 1 μl T4 DNA ligase (5 Weiss U/μl, Fermentas). The ligation reaction was performed at 16°C overnight. Alternatively, the ligation can be carried out within 5 min at 25°C using the NEB Quick Ligation Kit.

PCR amplification with anchored poly(T) primer and adapter specific primer

The PCR included following components for a 25 μl reaction: 18 μl H_2O, 2.5 μl PCR buffer (Rapidozym), 1 μl $MgCl_2$ (50 mM), 0.5 μl dNTP (10 mM each), 1 μl anchored poly(T) primer (5′-T_{23}GC-3′) + 1 μl adapter specific primer (5′-GCTTATTAACCCTCAC-3′) (10 μM), 0.5 μl template from crude adapter ligation mixture, 0.5 μl Taq DNA polymerase (5 U/μl, GenTherm Rapidozym). The PCR was run on a PTC 100 cycler (Biozym) using following program: 95°C for 3 min (1 cycle); 95°C for 1 min, 51°C for 1 min, 73°C for 1 min (30 cycles); 73°C for 7 min (1 cycle).

Blunt end cloning of PCR product

The PCR product was gel purified (2% agarose gel, Gel extraction kit Quiagen) and blunted following the manual (Quick Blunting Kit, NEB). The blunted PCR product was ligated in a blunted pKs (+) vector using the Quick Ligation Kit (NEB). After transformation in DH5 alpha bacteria and overnight incubation at 37°C, clones were analysed by colony PCR using the M13 primer binding

sites present in the vector (agarose gel, 2%). PCR products containing the insert were purified by PCR purification kit (Quiagen). Eighty nanograms of DNA of each sample were sent for sequencing.

Sequencing

Sequencing was done by SeqLab (Göttingen) using the M13 forward and reverse primer binding sites present in the PCR product. Due to the extended poly(A)/(T) segments, polymerase slippage occurs frequently, and only one sequencing direction (the direction in which the homopolymeric stretch is at the end of the read) produces exploitable results.

RESULTS

Synthesis of the pre-loaded resin

We aimed at generating a covalent linkage between the solid support and the first nucleoside monomer in a way that the oligonucleotide remains connected to the support after deprotection, and the 3′-OH terminus stays accessible for post-synthetic modifications. Since commonly used solid supports are bound to the first monomer via this very position, an alternative had to be found. The 4-triazolyl group has often been used as an activating group to introduce various substituents at the 4-position of uridine or deoxyuridine (24–26) especially by reaction with primary amino groups (27–29). Therefore, 4-triazolyluridine was an attractive option to attach a nucleoside to an amino-functionalized solid phase. The activated nucleoside **2** was synthesized in three steps

from uridine according to standard protocols (22,23) (Scheme 1).

A solid support suitable for reaction with **2** should not only bear a primary amino group, but also be compatible with solid-phase oligonucleotide synthesis as well as with on-bead screening procedures. Controlled pore glass (CPG), the most commonly used support for nucleic acid synthesis, lacks the mechanical stability and homogeneity in size that is required for on-bead screening procedures (30). In contrast, polystyrene resins are available in different bead sizes with a uniform particle size distribution and a high mechanical stability. We therefore opted for the polystyrene-based, swellable Tentagel resin. This solid support is chemically compatible with all used reagents, and it is equally swellable in all solvents used. This type of resin has frequently been used in the preparation of combinatorial libraries of peptides (30) and has also been applied to the synthesis of OBOC DNA libraries (31). The coupling of **2** was performed on Tentagel-NH_2 with two different bead sizes: 10 and 90 μm. After allowing the Tentagel-NH_2 to pre-swell in DMF, the beads were suspended in solutions of **2** in DMSO at various concentrations. Addition of 6% DMF facilitated a homogeneous suspension of the beads. After incubation at 40°C overnight and several washing steps, unreacted amino groups were blocked by capping with *tert*-butylphenoxyacetyl acetic anhydride. The coupling yield of **2** to the resin and the respective loading capacity of each resin were determined by cleavage of the dimethoxytrityl (DMT) group and measurement of its absorption at 498 nm (Figure 1). The 90 μm beads showed better coupling yields than the 10 μm ones.
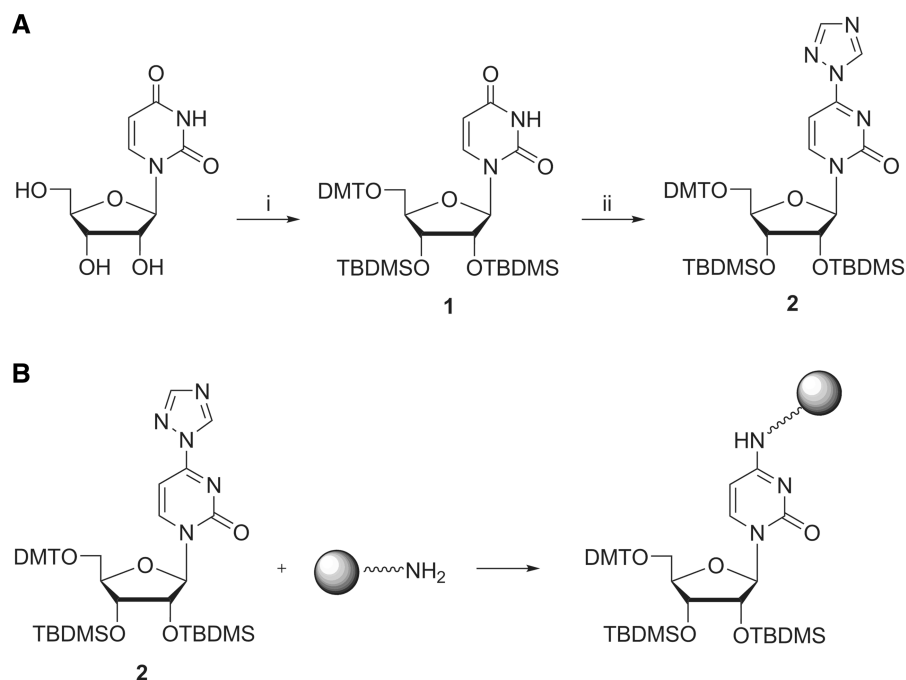
Scheme 1. (**A**) Synthesis of the activated uridine analogue **2**: (i) 1. DMT-Cl, pyridine; 2. TBDMS-Cl, imidazole, DMF; (ii) 1,2,4-triazole, $POCl_3$, NEt_3, acetonitrile. (**B**) Coupling scheme of protected 4-triazolyluridine (**2**) to amino-functionalized resin (for conditions, see 'Materials and Methods' section). The covalent linkage is formed between the base moiety of **2** and the amino-groups of the resin, thereby leaving the 3′-OH accessible for subsequent modifications.
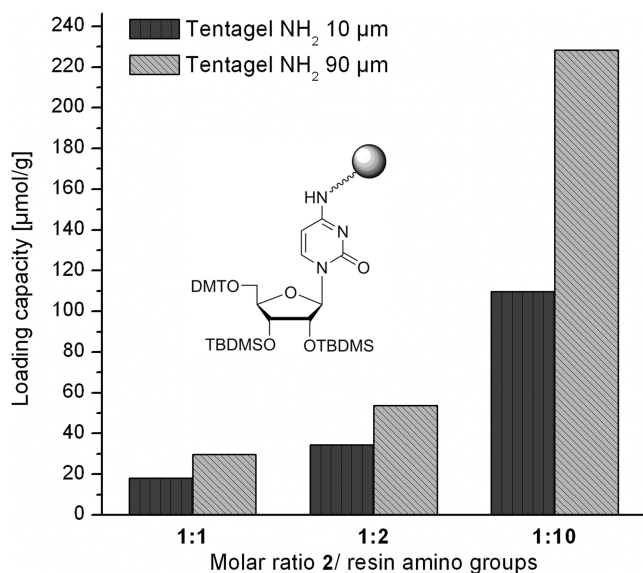
Figure 1. Loading capacity (μmol/g) of pre-loaded Tentagel resins (10 and 90 μm bead size) after coupling to **2** as determined by absorption measurement (498 nm) of cleaved DMT groups. The different values were obtained by variation of the molar ratio **2**/resin amino groups. The nominal loading capacity of both resins before coupling was 220 μmol/g.

By varying the molar ratio of **2** to resin amino groups, coupling yields could be obtained ranging from 8.3% (Tentagel 10 μm, 1:1 ratio) to quantitative coupling (Tentagel 90 μm, 1:10 ratio). It is therefore possible to adjust the loading capacity of the resin to a desired value depending on the application. For example, a high loading capacity may be beneficial for obtaining high signals during the screening process, while a low loading capacity can be desirable for preventing misfolding, improper binding of a target molecule (32) or quenching of fluorophores (33).

RNA synthesis on pre-loaded resin

The pre-loaded resin was applied in the solid phase synthesis of RNA. Since the Tentagel resin is swellable, the synthesis protocol differs considerably from conventional protocols used for CPG. It was necessary to increase the reaction times in order to allow the reagents to diffuse inside the swollen beads. Increasing the volume of reagents and acetonitrile during the washing steps also led to better coupling efficiencies. Furthermore, it was essential to introduce an additional washing step with anhydrous DMF after the oxidation step, which uses aqueous iodine. This procedure removed residual water still present on the resin, thereby preventing hydrolysis of the phosphoramidites in the following coupling step. Compared to standard RNA synthesis on CPG time for synthesis on Tentagel is increased 1.6-fold and cost 1.8-fold. In our opinion, the advantages of Tentagel, including the decreased unspecific binding of hydrophobic substances compared to polystyrene resin and the increased wettability in aqueous medium exceed the disadvantages of increased synthesis time and costs. Using this optimized protocol, we synthesized five different sequences on five pre-loaded resins varying in their loading capacity and bead size (Table 1). After coupling of the first monomer (rG), the cleaved DMT fraction was manually collected. The absorption measurement at 498 nm showed in all five cases a coupling yield of ~99%, demonstrating the high accessibility and reactivity of the resin-linked nucleoside. After completion of RNA synthesis, the average coupling yield was ~99% as determined by the DMT cation assay. After ammonia treatment and TBAF deprotection of the 2′- and 3′-OH, the solid support thus prepared was ready for post-synthetic modifications.

Post-synthetic introduction of primer binding sites for amplification and sequencing

To amplify unknown RNA sequences, it is necessary to introduce a first primer binding site at the 3′-terminus to enable reverse transcription, and then a second primer binding site at the 3′-end of the respective cDNA to perform PCR. Our strategy consists of adding a homopolymeric tail to the 3′-terminus of the resin-bound RNA sequence by *E. coli* poly(A) polymerase, to allow for the use of a poly(T) primer during reverse transcription (Figure 2A). For the introduction of the second primer binding site at the 3′-end of the respective cDNA, a specific problem had to be taken into account: solid phase RNA synthesis always yields incomplete sequences due to non-quantitative coupling efficiencies. This leads to beads that carry shorter, abortive sequences next to the full-length product. To ensure that only the full-length product is amplified, we introduced three cytidines at the 5′-end of the RNA strands, which during RT are converted to three deoxyguanosines at the 3′-end of the cDNA. These deoxyguanosines serve as anchor for the

Table 1. List of pre-loaded resins subjected to RNA synthesis

No.	Resin (μm)	Loading capacity (μmol/g)	Sequence (5′–3′)
1	Tentagel-C* 10	18	**CCC**AUUAGGUCAGUAACUCAGU**GC*-SS**
2	Tentagel-C* 10	110	**CCC**AGUAUAGAGGAAGUCAGGUAUAAGU**GC*-SS**
3	Tentagel-C* 90	30	**CCC**AAUGGAACCAUAUCUAUACCUGAGU**GC*-SS**
4	Tentagel-C* 90	228	**CCC**UGAUAUUCCUUUGUCUGAGU**GC*-SS**
5	Tentagel-C* 10	35	**CCC**GUUGAUCCCACAAGUCAUAGGUAAUA**GC*-SS**

Resins with different bead sizes (10 μm, 90 μm) and different loading capacities were used to synthesize five sequences that only share the conserved nucleotides (bold). 'C*' indicates the cytidine-like nucleoside that is formed after coupling of **2** to the amino-functionalized resin. '–SS' indicates the covalent linkage of RNA to the solid support.
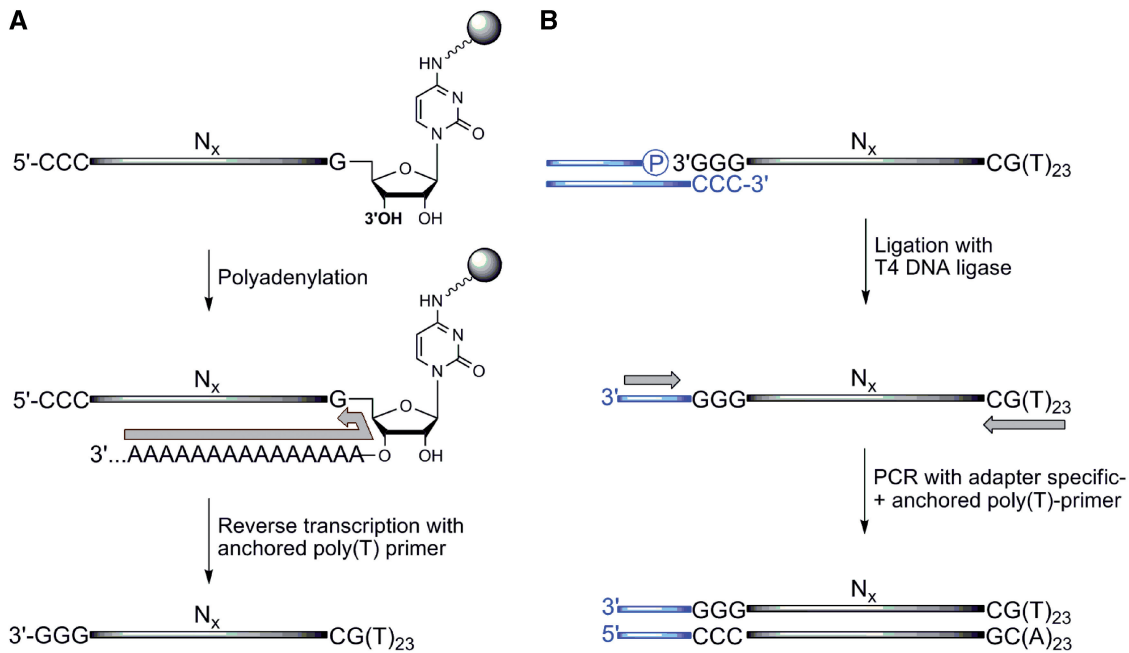
Figure 2. Post-synthetic introduction of primer binding sites for amplification. (**A**) Polyadenylation of the 3′-terminus of RNA on bead. After introduction of the poly(A)-tail, reverse transcription can be performed using an anchored poly(T) primer (grey arrow). (**B**) Adapter ligation to the 3′-end of the generated cDNA with T4 DNA ligase. The 3′-CCC overhang of the double-stranded adapter (in blue) hybridizes only to full length product exhibiting three guanosines at the 3′-terminus. Subsequent amplification by PCR using an adapter-specific and an anchored poly(T)-primer (grey arrows) yields a product that can be subjected to blunt end cloning and sequencing.

3′-overhang of a double-stranded DNA-adapter which in turn is ligated by T4 DNA ligase (Figure 2B). Hence, only the full length sequences will carry the second primer, and will therefore be amplified. The three additional cytidines have a double role: to allow for the introduction of the second primer, and to exclude abortive sequences. The PCR product obtained by using the two introduced primers can then be subjected to blunt end cloning and sequencing.

As analytical monitoring options are limited for resin-bound RNA, we first optimized the combination of tailing and reverse transcription in solution using a known RNA sequence (data not shown). We found that an anchored primer of 23 thymidines and two constant nucleotides that match the first 2 nt at the 3′-end of the oligonucleotide was ideally suited to produce a single-length reverse transcript starting from the heterogenous poly(A) tail. Finally, this process was applied to RNA being attached to Tentagel beads: it remained to be shown whether both the poly(A) polymerase and the reverse-transcriptase would tolerate the presence of the solid phase in close proximity to the RNA to be modified. To investigate the general applicability of this methodology, we applied it to the five different RNA sequences synthesized on five different pre-loaded resins (10 and 90 μm, high loading and low loading; Table 1). One sequence (No. 5) was designed to include a C-triplet within the variable region, to test if such a triplet would interfere with the adapter ligation. The following steps were applied to three to five single beads out of every sample. Each bead was processed individually since the sequence determination during on-bead screening
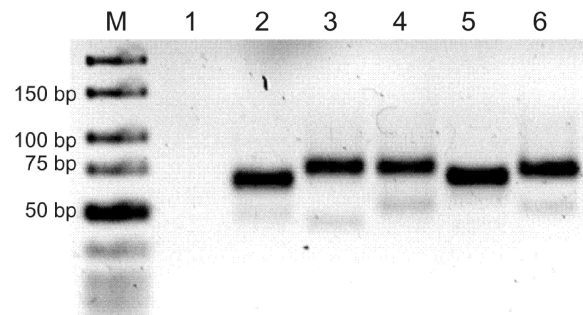

Figure 3. PCR products obtained from different on-bead sequences (Table 1) after introduction of primer binding sites. M, base pair ladder. Lane 1, negative control without template. Lane 2, RNA No. 1; Lane 3, RNA No. 2; Lane 4, RNA No. 3; Lane 5, RNA No. 4; Lane 6, RNA No. 5. All PCR products show the expected length.

procedures has to be carried out from a single bead. A bead picked under a light microscope was subjected to polyadenylation and subsequent reverse transcription in a one-pot reaction. Combining two steps in one reaction vessel required buffer adjustment but decreased the number of pipetting steps and thereby reduced the risk of losing the bead. After isolation, the bead was suspended in water in a fresh vial and heated to 95°C to denature the RNA/cDNA hybrid. The adapter ligation was then performed with a ~50-fold excess of double-stranded adapter over cDNA using T4 DNA ligase. For subsequent amplification, 5% of the crude ligation mixture was subjected to PCR using an adapter-specific primer and an anchored poly(T) primer. Irrespective of sequence, length, bead

Table 2. Sequencing results (**A**), including template sequence (bold) and introduced primer sequences, in comparison to synthesized RNA sequences (**B**)

No.	Sequence alignment
1	**A** . . .TTGCTTATTAACCCTCACTAAAC**CCCATTAGGTCAGTAACTCAGTGC**AAAAAAAAAAAAAAAAAAAAAAAA. . .
	B CCCAUUAGGUCAGUAACUCAGUGC
2	**A** . . .TTGCTTATTAACCCTCACTAAA**CCCAGTATAGAGGAAGTCAGGTATAAGTGC**AAAAAAAAAAAAAAAAAAAAAAAA. . .
	B CCCAGUAUAGAGGAAGUCAGGUAUAAGUGC
3	**A** . . .TTGCTTATTAACCCTCACTAAA**CCCAATGGAACCATATCTATACCTGAGTGC**AAAAAAAAAAAAAAAAAAAAAAAA. . .
	B CCCAAUGGAACCAUAUCUAUACCUGAGUGC
4	**A** . . .TTGCTTATTAACCCTCACTAAA**CCCTGATATTCCTTTGTCTGAGTGC**AAAAAAAAAAAAAAAAAAAAAAAA. . .
	B CCCUGAUAUUCCUUUGUCUGAGUGC
5	**A** . . .TTGCTTATTAACCCTCACTAAA**CCCGTTGATCCCACAAGTCATAGGTAATAGC**AAAAAAAAAAAAAAAAAAAAAAAA. . .
	B CCCGUUGAUCCCACAAGUCAUAGGUAAUA**GC**

The number (1–5) corresponds to the respective RNA sequence, see Table 1.

size and loading capacity, all five reactions yielded a PCR product of the expected length (Figure 3). No differences between individual beads that carry the same sequence could be detected. Even sequence 5, containing an internal C-triplet, produced a full-length PCR product and none of the shorter sequences that would have been expected in case of hybridization of the adapter to the internal C-triplet. Importantly, the efficiency of the method is insensitive to the folding of the RNA molecules. In principle, folding could block sequence segments involved in tailing and adapter ligation. In spite of different folding patterns (prediction by mfold, data not shown), all five sequences were successfully obtained after sequencing. Hence, the method described here seems to be applicable regardless of the folding pattern. In all PCR reactions (except for RNA 4), a single, faint by-product has been observed in some cases that might be caused by adapter amplification. Subsequent gel purification of the PCR product, however, quantitatively removed the by-product, and blunt end cloning and sequencing yielded in all five cases the expected sequences as shown in Table 2, thus demonstrating that this strategy is a reliable and versatile method for the sequencing of unknown on-bead RNA sequences. It should be noted that gel purification of the PCR product is mandatory for a clean sequencing. Omitting this step, e.g. using the crude PCR product and performing T/A cloning, yielded several sequences with extended deletions.

DISCUSSION

Here, we describe a new format of oligonucleotide combinatorial libraries for ligand, catalyst and tag discovery. Split-and-mix synthesis of one-bead one-compound libraries ensures that each physical entity (bead) contains multiple copies of one sequence. The Tentagel resins in combination with the strategy for sequence determination used here allow for large library sizes; 500 mg of 10 μm resin contain 10^9 beads, permitting complete randomization of 15 nt ($4^{15} \approx 10^9$), compared to 10^6 beads or 10 randomized nucleotides reported previously (13). Considering that 13- and 14-nt aptamers are known to bind small organic molecules with low micromolar

affinity (34,35) and that a simple pentanucleotide accelerates an aminoacylation reaction 25-fold (36), OBOC RNA libraries such as the one described here should allow searching for a wide variety of functional oligonucleotides. The small number of nucleotides required for sequence determination, combined with the highly efficient synthesis, gives more freedom to the rational design of libraries in comparison to previous studies (12,13); it does so without compromising the purity of the library. The randomized positions do not have to be synthesized as one contiguous stretch, they can rather be presented in a precise structural context, e.g. of a bulge, a pseudoknot, a quadruplex structure or a three-way junction that can be defined by constant elements designed into the sequence. Of particular interest is the search for properties that cannot be selected for by SELEX approaches; e.g. autofluorescent or fluorescence-enhancing tags which are urgently needed for RNA cell biology (37). Using fluorescence-activated cell sorters, a library of 10^9 beads can be screened for very specific optical properties in less than a day. The fact that one bead of the smallest size (10 μm) with the lowest loading (18 μmol/g) still contains 9 fmol or 5 billion copies of one molecular species constitutes a dramatic advantage over the SELEX situation, where one relies on the interaction of individual molecules, and allows for a much wider choice of analytical options.

Compared to other classes of polymers, sequence determination is highly developed for nucleic acids. Previous reports on sequence determinations from DNA OBOC libraries used either Maxam-Gilbert chemistry (31) (which is indeed suitable for such short oligonucleotides, but cannot be easily applied when chemical modifications are incorporated into the library), mass spectrometric sequencing (38) (requiring cleavable linkers) or Sanger sequencing using extended primer binding sites included in the sequence (12). Here, we investigated whether tailing and adapter ligation strategies developed for the discovery of non-coding RNAs (39–41) could be adapted to sequence OBOC libraries without large known regions. This is not a trivial problem, as many enzymes do not tolerate the presence of bulky solid particles attached to their substrates (42). Our strategy requires two on-bead enzymatic reactions: poly(A)-tailing and reverse

transcription. The results demonstrate that *E. coli* poly(A) polymerase accepts the 3′-hydroxyl group of the RNA even though the nucleobase of this very nucleotide is attached to a polystyrene bead, and even though this is done via the exocyclic amino group that is involved in Watson–Crick pairing of the complementary nucleotide. Also, the reverse transcriptase produced in all cases sufficient amounts of cDNA to allow for ligation to the double-stranded adapter and amplification in a subsequent PCR reaction. In comparison to our previously published study (13), the method presented here decreases the structural bias of the library and increases flexibility in library design at the same time. Using the new method for sequence determination, one can now synthesize and screen on-bead libraries of dramatically increased diversity compared to published approaches (12,31) and thereby increase the probability to isolate oligonucleotides with the desired characteristics.

The presence of an enzymatically and chemically accessible 3′-end in immobilized oligonucleotides prepared by standard 3′ → 5′ phosphoramidite synthesis may be attractive for many other applications, from labelling (43–46) to microarrays (47,48) and genotyping (49). It should be noted that the overall structure of the RNA-bound beads after tailing bears a strong resemblance to the clonally clustered amplicons that are generated in deep sequencing (50,51): multiple copies of one unknown sequence attached to a bead or to a localized surface area, each with a known primer binding site at its 3′-end. It is therefore conceivable that the OBOC RNA libraries described here can be directly sequenced (52) with next-generation sequencing instrumentation, without any PCR, cloning and Sanger sequencing. The combination of high-throughput FACS sorting with next generation sequencing would provide an attractive pipeline for the discovery of functional RNAs.

REFERENCES

1. Famulok,M., Hartig,J.S. and Mayer,G. (2007) Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. *Chem. Rev.*, **107**, 3715–3743.
2. Joyce,G.F. (2007) Forty years of in vitro evolution. *Angew. Chem. Int. Ed.*, **46**, 6420–6436.
3. Weigand,B.S., Zerressen,A., Schlatterer,J.C., Helm,M. and Jäschke,A. (2006) Catalytically active RNA molecules: tools in organic chemistry. In: Klussmann,S. (ed.), *The Aptamer Handbook - Functional Oligonucleotides and Their Applications*. Wiley-VCH, Weinheim, Germany, pp. 211–227.
4. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
5. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage-T4 DNA-polymerase. *Science*, **249**, 505–510.
6. Robertson,D.L. and Joyce,G.F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.
7. Blank,M., Weinschenk,T., Priemer,M. and Schluesener,H. (2001) Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels - selective targeting of endothelial regulatory protein pigpen. *J. Biol. Chem.*, **276**, 16464–16468.
8. Misono,T.S. and Kumar,P.K.R. (2005) Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. *Anal. Biochem.*, **342**, 312–317.
9. Mendonsa,S.D. and Bowser,M.T. (2004) In vitro evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.*, **126**, 20–21.
10. Lou,X.H., Qian,J.R., Xiao,Y., Viel,L., Gerdon,A.E., Lagally,E.T., Atzberger,P., Tarasow,T.M., Heeger,A.J. and Soh,H.T. (2009) Micromagnetic selection of aptamers in microfluidic channels. *Proc. Natl. Acad. Sci.*, **106**, 2989–2994.
11. Lam,K.S., Salmon,S.E., Hersh,E.M., Hruby,V.J., Kazmierski,W.M. and Knapp,R.J. (1991) A new type of synthetic peptide library for identifying ligand-binding activity. *Nature*, **354**, 82–84.
12. Yang,X., Bassett,S.E., Li,X., Luxon,B.A., Herzog,N.K., Shope,R.E., Aronson,J., Prow,T.W., Leary,J.F., Kirby,R. *et al.* (2002) Construction and selection of bead-bound combinatorial oligonucleoside phosphorothioate and phosphorodithioate aptamer libraries designed for rapid PCR-based sequencing. *Nucleic Acids Res.*, **30**, e132/131–e132/138.
13. Wiesmayr,A. and Jäschke,A. (2011) Isolation and characterization of fluorescence-enhancing RNA tags. *Bioorg. Med. Chem.*, **19**, 1041–1047.
14. Trapp,O. (2008) Gas chromatographic high-throughput screening techniques in catalysis. *J. Chromatogr. A*, **1184**, 160–190.
15. Reetz,M.T. (2001) Combinatorial and evolution-based methods in the creation of enantioselective catalysts. *Angew. Chem. Int. Ed.*, **40**, 284–310.
16. Legiewicz,M., Lozupone,C., Knight,R. and Yarus,M. (2005) Size, constant sequences, and optimal selection. *RNA*, **11**, 1701–1709.
17. Seelig,B. and Jäschke,A. (1999) A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.*, **6**, 167–176.
18. Shtatland,T., Gill,S.C., Javornik,B.E., Johansson,H.E., Singer,B.S., Uhlenbeck,O.C., Zichi,D.A. and Gold,L. (2000) Interactions of Escherichia coli RNA with bacteriophage MS2 coat protein: genomic SELEX. *Nucleic Acids Res.*, **28**, e93.
19. Jarosch,F., Buchner,K. and Klussmann,S. (2006) In vitro selection using a dual RNA library that allows primerless selection. *Nucleic Acids Res.*, **34**, e86.
20. Wen,J.-D. and Gray,D.M. (2004) Selection of genomic sequences that bind tightly to Ff gene 5 protein: primer-free genomic SELEX. *Nucleic Acids Res.*, **32**, e182.
21. Pan,W., Xin,P. and Clawson,G.A. (2008) Minimal primer and primer-free SELEX protocols for selection of aptamers from random DNA libraries. *Biotechniques*, **44**, 351–360.
22. Chirakul,P. and Sigurdsson,S.T. (2003) Stereospecific syntheses of 3′-deuterated pyrimidine nucleosides and their site-specific incorporation into DNA. *Org. Lett.*, **5**, 917–919.
23. Sheu,C., Kang,P., Khan,S. and Foote,C.S. (2002) Low-temperature photosensitized oxidation of a guanosine derivative and formation of an imidazole ring-opened product. *J. Am. Chem. Soc.*, **124**, 3905–3913.
24. Reese,C.B. and Skone,P.A. (1984) The protection of thymine and guanine residues in oligodeoxyribonucleotide synthesis. *J. Chem. Soc., Perkin Trans*, **1**, 1263–1271.
25. Xu,Y.Z., Zheng,Q. and Swann,P.F. (1992) Synthesis of DNA containing modified bases by post-synthetic substitution. Synthesis of oligomers containing 4-substituted thymine: O4-alkylthymine, 5-methylcytosine, N4-dimethylamino-5-methylcytosine, and 4-thiothymine. *J. Org. Chem.*, **57**, 3839–3845.

26. MacMillan,A.M. and Verdine,G.L. (1990) Synthesis of functionally tethered oligodeoxynucleotides by the convertible nucleoside approach. *J. Org. Chem.*, **55**, 5931–5933.

27. Pochet,S., Huynh-Dinh,T. and Igolen,J. (1985) Synthèse d'oligonucléotides liés de facon covalente à un support solide. *Tetrahedron Lett.*, **26**, 627–630.

28. Caprioara,M., Fiammengo,R., Engeser,M. and Jäschke,A. (2007) DNA-based phosphane ligands. *Chem. Eur. J.*, **13**, 2089–2095.

29. Fournier,P., Fiammengo,R. and Jäschke,A. (2009) Allylic amination by a DNA–diene–iridium(I) hybrid catalyst. *Angew. Chem. Int. Ed.*, **48**, 4426–4429.

30. Lam,K.S., Lebl,M. and Krchnak,V. (1997) The "one-bead-one-compound" combinatorial library method. *Chem. Rev.*, **97**, 411–448.

31. Seliger,H., Bader,R., Hinz,M., Rotte,B., Astriab,A., Markiewicz,M. and Markiewicz,W.T. (1997) Synthetic oligonucleotide combinatorial libraries - tools for studying nucleic acid interactions. *Nucleosides Nucleotides*, **16**, 703–710.

32. Chen,X., Tan,P.H., Zhang,Y. and Pei,D. (2009) On-bead screening of combinatorial libraries: reduction of nonspecific binding by decreasing surface ligand density. *J. Comb. Chem.*, **11**, 604–611.

33. Yan,B., Martin,P.C. and Lee,J. (1998) Single-bead fluorescence microspectroscopy: detection of self-quenching in fluorescence-labeled resin beads. *J. Comb. Chem.*, **1**, 78–81.

34. Anderson,P.C. and Mecozzi,S. (2005) Identification of a 14mer RNA that recognizes and binds flavin mononucleotide with high affinity. *Nucleic Acids Res.*, **33**, 6992–6999.

35. Anderson,P.C. and Mecozzi,S. (2005) Unusually short RNA sequences: design of a 13-mer RNA that selectively binds and recognizes theophylline. *J. Am. Chem. Soc.*, **127**, 5290–5291.

36. Turk,R.M., Chumachenko,N.V. and Yarus,M. (2010) Multiple translational products from a five-nucleotide ribozyme. *Proc. Natl. Acad. Sci.*, **107**, 4585–4589.

37. Darzacq,X., Yao,J., Larson,D.R., Causse,S.Z., Bosanac,L., de Turris,V., Ruda,V.M., Lionnet,T., Zenklusen,D., Guglielmi,B. *et al.* (2009) Imaging transcription in living cells. *Annu. Rev. Biophys.*, **38**, 173–196.

38. Rulka,A. and Markiewicz,W.T. (2010) *IRT 2010 - XIX International Round Table on Nucleosides, Nucleotides and Nucleic Acids.* Lyon, France.

39. Hafner,M., Landgraf,P., Ludwig,J., Rice,A., Ojo,T., Lin,C., Holoch,D., Lim,C. and Tuschl,T. (2008) Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods*, **44**, 3–12.

40. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858–862.

41. Hüttenhofer,A., Brosius,J. and Bachellerie,J.-P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, **6**, 835–843.

42. Kress,J., Zanaletti,R., Amour,A., Ladlow,M., Frey,J.G. and Bradley,M. (2002) Enzyme accessibility and solid supports: Which molecular weight enzymes can be used on solid supports? An investigation using confocal Raman microscopy. *Chem. Eur. J.*, **8**, 3769–3772.

43. Schmitz,G.G., Walter,T., Seibl,R. and Kessler,C. (1991) Nonradioactive labeling of oligonucleotides in vitro with the hapten digoxigenin by tailing with terminal transferase. *Anal. Biochem.*, **192**, 222–231.

44. Igloi,G.L. (1996) Substrate properties of fluorescent ribonucleotides in the terminal transferase-catalyzed labeling of DNA sequencing primers. *Biotechniques*, **21**, 1084–1092.

45. Martin,G. and Keller,W. (1998) Tailing and 3′-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA*, **4**, 226–230.

46. Paredes,E. and Das,S.R. (2011) Click chemistry for rapid labeling and ligation of RNA. *Chembiochem*, **12**, 125–131.

47. Kurg,A., Tonisson,N., Georgiou,I., Shumaker,J., Tollett,J. and Metspalu,A. (2000) Arrayed primer extension: Solid-phase four-color DNA resequencing and mutation detection technology. *Genet. Test.*, **4**, 1–7.

48. Pullat,J., Kusnezow,W., Jaakson,K., Beier,M., Hoheisel,J.D. and Metspalu,A. (2008) Arrayed primer extension on in situ synthesized 5′ –> 3′ oligonucleotides in microchannels. *New Biotechnol.*, **25**, 133–141.

49. Steemers,F.J., Chang,W., Lee,G., Barker,D.L., Shen,R. and Gunderson,K.L. (2006) Whole-genome genotyping with the single-base extension assay. *Nat. Methods*, **3**, 31–33.

50. Margulies,M., Egholm,M., Altman,W.E., Attiya,S., Bader,J.S., Bemben,L.A., Berka,J., Braverman,M.S., Chen,Y.-J., Chen,Z. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

51. Shendure,J. and Ji,H. (2008) Next-generation DNA sequencing. *Nat. Biotechnol.*, **26**, 1135–1145.

52. Ozsolak,F., Platt,A.R., Jones,D.R., Reifenberger,J.G., Sass,L.E., McInerney,P., Thompson,J.F., Bowers,J., Jarosz,M. and Milos,P.M. (2009) Direct RNA sequencing. *Nature*, **461**, 814–818.