



Cite this: *Chem. Commun.*, 2014, 50, 10937

Received 21st June 2014,
Accepted 22nd July 2014

DOI: 10.1039/c4cc04719b

www.rsc.org/chemcomm

Enzymatic combinatorial nucleoside deletion scanning mutagenesis of functional RNA†

Katarzyna Wawrzyniak-Turek and Claudia Höbartner*

We describe a general and simple method to identify catalytically and structurally important nucleotides in functional RNAs. Our approach is based on statistical replacement of each nucleoside with a non-nucleosidic spacer (C3 linker, Δ), followed by separation of active library variants and readout of interference effects by analysis of enzymatic primer extension reactions.

Many examples of natural and artificial functional RNA motifs have been identified, including ribozymes, aptamers, riboswitches, small interfering RNAs, or protein-binding RNAs, that demonstrate the diversity of RNA functions beyond the transmission of genetic information.¹ Elucidating the structure and sequence requirements of non-coding nucleic acids is crucial for understanding their mechanisms of action. Furthermore, such knowledge facilitates design and engineering of RNA for practical applications in biomolecular chemistry and synthetic biology.

Numerous biochemical and biophysical methods are used to analyse nucleic acid architectures at various levels of resolution.² Recent additions to the traditional repertoire are geared towards high-throughput analyses on massively parallel arrays, or combine mutations and footprinting studies with computational methods.^{3,4}

Our laboratory has recently demonstrated combinatorial analysis methods (CoMA,⁵ dNAIM,⁶ NDS⁷) that allowed the simultaneous assessment of all possible single point mutants of deoxyribozymes, as well as the identification of catalytically important nucleotides and their functional groups. However, due to the method design, the reported approaches are only applicable for studying functional DNAs. They are based on the combinatorial solid-phase synthesis of DNA oligonucleotide libraries, in which the mutations are marked with a chemical tag: the hydroxyl group resembling the 2'-OH group of ribonucleotides. Mutations tagged in this way can be easily decoded

by alkaline hydrolysis. This is an asset for the analysis of functional DNAs, but, for obvious reasons, the same strategy cannot be applied to RNA.

In the present work we established an analogously versatile combinatorial approach for RNA that reveals functionally important nucleotides in a simple set of experiments that do not require sophisticated instrumentation. We used the three-carbon linker Δ as non-nucleosidic spacer unit to statistically replace ("delete") standard ribonucleotides within functional nucleic acids (Fig. 1). Following separation of the functionally active and inactive library variants, the spacer substitutions were decoded by primer extension reactions. Analysis of the primer extension pattern allowed identification of the essential and non-essential nucleotides. We demonstrated our new approach

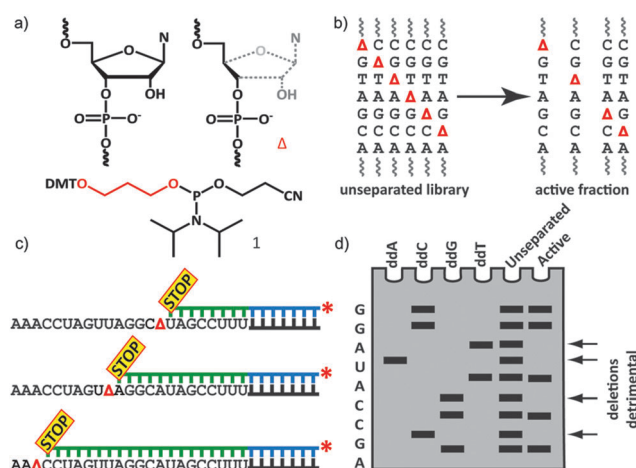


Fig. 1 Concept of enzymatic combinatorial nucleoside deletion scanning mutagenesis: (a) comparison of a standard nucleoside with non-nucleosidic spacer unit Δ (C3 linker) and structure of phosphoramidite **1** used for incorporation of Δ ; (b) schematic representation of a combinatorial library in which parent nucleosides are statistically replaced with Δ , and isolation of active fraction; (c) substituted positions are revealed by primer extension because polymerases cannot extend beyond Δ ; (d) analysis of the interference pattern by denaturing PAGE. Missing bands in the active fraction identify the essential nucleotides.

Research Group Nucleic Acid Chemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.

E-mail: claudia.hoebartner@mpibpc.mpg.de

† Electronic supplementary information (ESI) available: Experimental details and Fig. S1–S11. See DOI: 10.1039/c4cc04719b

with two well-known examples of functional RNAs, including a ribozyme and a protein-binding RNA. Our method was able to accurately identify the catalytically and structurally important areas.

For the enzymatic combinatorial nucleoside deletion scanning mutagenesis (enzymatic NDS), we prepared RNA libraries in which each nucleoside in the region of interest was statistically replaced by the non-nucleosidic spacer unit in such a way that there was on average only one substitution per strand (Fig. 1b). Such libraries were prepared by solid-phase synthesis using mixtures of standard RNA phosphoramidites with the non-nucleosidic phosphoramidite **1**. For example, in a 40 nt long region of interest, the desired one substitution per molecule can be achieved by approximately 5% spacer incorporation.⁵ To achieve this desired level of Δ substitution, we first examined the coupling efficiency of **1** in competition with RNA phosphoramidites (rN). We synthesized a series of pentamer oligonucleotides with single C3-spacer replacement using phosphoramidite mixtures containing 5–25% of **1**. The amount of Δ incorporation was subsequently analysed by anion-exchange HPLC (see ESI†). Based on these results we used phosphoramidite mixtures for synthesis of RNA libraries containing **1** and rN in a ratio of 5:95.

Enzymatic NDS mutagenesis was designed to take advantage of the fact that polymerase enzymes cannot extend beyond the synthetic abasic sites introduced by the C3 spacer (Δ) (Fig. 1c).⁸ The presence of Δ in the template strand causes RNA-dependent DNA polymerases to pause and abort the extension reaction. Therefore, after separation of the active and inactive library variants by appropriate means (such as gel electrophoresis or affinity chromatography), we could use the primer extension to decode the positions where the substitutions with Δ were tolerated. To optimize the primer elongation conditions, we first tested different reverse transcriptases on RNA templates that contained Δ substitutions at defined positions (see ESI†). We found that the reaction time is a critical parameter for analysis; short primer extension times of 2–10 min gave reliable and reproducible results. Both tested reverse transcriptases (M-MuLV and SSIII RT) terminated the primer extension by incorporation of the last nucleotide at the position preceding the acyclic modification. We also demonstrated that C3 spacer substitutions can be detected in a DNA template by DNA-dependent DNA polymerases (Klenow fragment and One Taq polymerase) (see ESI†). In contrast to reverse transcriptases, Klenow fragment DNA polymerase incorporated the last nucleotide directly opposite to the C3 linker (non-templated incorporation). One Taq DNA polymerase produced two abort bands (only partial addition of a non-templated nucleotide opposite Δ). This information is important for the correct assignment of the abort bands for analysis. In further experiments both M-MuLV and SSIII RT were used to analyse unseparated RNA libraries as well as the active RNA fractions (inactive fractions were occasionally included in the primer extension assay, but not used in the analysis).

Interference values were calculated as the ratio of band intensities resulting from the primer extension pattern in the active fraction and the unseparated library. The nucleosides for which the interference effect exceeded the value of 1.5 were considered to be functionally important for a given RNA.

Those nucleotides might be directly involved in catalysis or substrate recognition or play a role for correct folding and structure formation. Nucleotides that exhibited interference effects below 1.5 were considered non-essential and are good candidates for mutation and/or deletion without substantial impact on the nucleic acid function.

As a feasibility test for interference analysis of primer extension patterns, we used a combinatorial deletion library of a deoxy-ribozyme from our previous study that contained an analogous OH-modified C3 spacer.⁷ We performed the separation of the active and inactive library variants as previously described, and conducted the primer extension experiments according to our optimized methods (see ESI†). The results from the primer extension analysis were superimposable with the alkaline hydrolysis results,⁷ confirming that the enzymatic decoding by the DNA polymerase was able to provide a reliable readout of the interference pattern.

As a first example to demonstrate enzymatic combinatorial nucleoside deletion scanning mutagenesis on RNA, we chose to study the minimized hammerhead ribozyme (HHR) motif (Fig. 2a). This ribozyme catalyses a transesterification reaction, resulting in self-cleavage of the RNA strand. The HHR structure and consensus sequence are well characterized,⁹ providing a basis for comparison with our deletion mutagenesis results. The first step comprised the solid-phase synthesis of the combinatorial RNA library, containing the statistical C3 spacer substitutions

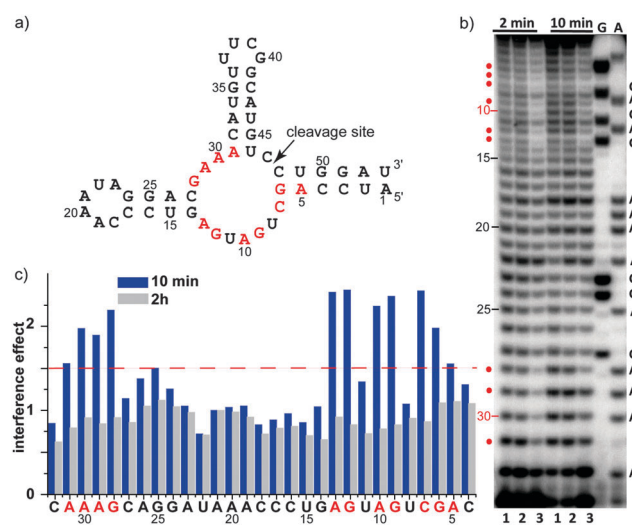


Fig. 2 (a) Secondary structure of the minimized hammerhead ribozyme motif. Nucleotides depicted in red proved to be essential for efficient catalysis. (b) Denaturing sequencing gel separation of the primer extension products generated by M-MuLV from the unseparated library (lane 1) and catalytically active fractions after 2 h (lane 2) and 10 min (lane 3) cleavage reactions. Primer extension for each sample was performed for 2 min and 10 min (label on top of gel). Sequencing ladders (G, A) were obtained by using corresponding ddNTPs (Sanger sequencing). (c) Bar graph showing the interference values for both active fractions (from 10 min and 2 h cleavage reactions) analysed by 2 min primer extension. Nucleotides showing interference effects higher than 1.5 are critical for the ribozyme function. Longer reaction times for DNA cleavage can compensate the negative effect of the mutation/deletion (compare interference values of the 10 min and 2 h separation reactions, blue and grey data).

within the HHR motif at nt positions 1–31. The substrate-containing region from position 32–53 remained unmodified. The HHR-catalysed cleavage reaction was performed in the presence of 10 mM MgCl₂ (Tris HCl pH 8.0) for various incubation times (10 min to 2 h). The active (*i.e.* cleaved) and inactive (un-cleaved) fractions were separated by denaturing gel electrophoresis (see ESI†). Subsequently, the active fractions were isolated and subjected to primer extension in comparison to the unselected library. Primer elongation was performed by M-MuLV and SSIII RT for 2–30 min (Fig. 2b and ESI†). In agreement with the results from the reaction optimization, longer incubation times for primer extension produced less significant differences in band intensities (see ESI† for the full analysis). Results from 2 min primer elongation reactions were used to calculate the interference effects (Fig. 2).

Analysis of the data revealed that nucleotides 7 (C), 9–10 (GA), 12–13 (GA) and 28–31 (GAAA) exhibited high interference effects (Fig. 2c), indicating that they are important for catalysis. This finding is consistent with results from previous studies on hammerhead ribozymes.¹⁰ In addition, strong interference effects were found at positions 5 and 6 (AG) which are located in a stem region in close proximity to the cleavage site. Most likely the substitution with the C3 spacer at these positions caused a destabilizing effect on the catalytic centre and/or hindered the substrate recognition. Interestingly, longer reaction times for the HHR-catalysed cleavage compensated the negative effect of the Δ substitutions (compare the interference values of the 10 min and 2 h RNA-catalysed cleavage reaction, Fig. 2c). This result indicated that the presence of Δ did not completely abolish the activity of the ribozyme but severely affected the reaction rates. As expected, single-nucleotide substitutions within the stem-loop (nt 14–27) were well tolerated (even at short reaction times), confirming that the method is suitable to distinguish critical from non-critical nucleotides.

The enzymatic combinatorial nucleoside deletion approach can also be used to study RNA–protein interactions. Here, we analysed the hairpin II of U1 snRNA (Fig. 3a) for which 7 nucleotides in the loop (AUUGCAC) are known to be essential for the recognition by the U1A protein and 2 loop closing base-pairs (G–C) are important for structural reasons.¹¹ The synthesised RNA library contained 17 mutagenized nucleotides (positions 1–17) and 14 non-substituted nucleotides (positions 18–31) as primer binding site for analysis. The separation of the active and inactive library variants was achieved by incubation of the RNA library with the U1A protein and subsequent separation of the bound and unbound fractions by the gel mobility shift assays on non-denaturing polyacrylamide gels (see ESI†). The analysis of the primer extension results revealed 9 nucleotides (AUUGCAC in the loop and CC in the stem) that exhibited strong interference effects (higher than 1.5), indicating that those nucleotides are required for the interaction of the U1A protein with the U1 snRNA. This result is in agreement with the above mentioned previous studies on the U1 snRNA motif.¹¹

In addition, our analysis showed that mutations at some of the critical loop nucleotides had less severe effects than others. When low protein concentration was used for the separation of the active and inactive fractions, strong interference effects

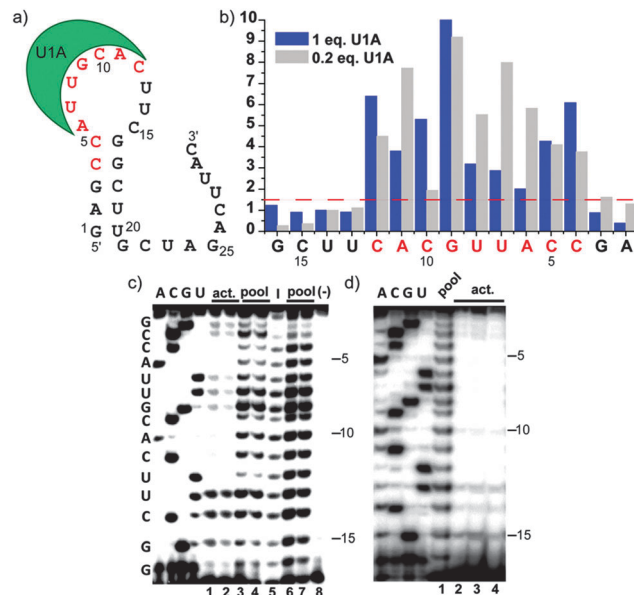


Fig. 3 (a) Hairpin II of the U1 snRNA extended by a primer binding sequence. Seven nucleotides in the loop region (AUUGCAC) are recognised by the U1A protein. Two C–G stem closing base-pairs are essential for the correct structure formation. (b) Summary of the interference results for 2 different separation conditions (1 and 0.2 equivalents of protein over RNA). (c) and (d) Primer extension gels for the separated fractions and unselected library. Extension products were produced by M-MuLV. Lanes A, C, G, U – sequencing ladders; (c) separation was done with 1 eq. of protein. Lane 1, 2 – active fraction from 2 independent experiments (act.); lane 3, 4, 6, 7 – unselected library (pool); lane 5 – inactive fraction (I); lane 8 – primer extension on the unmodified RNA (–) control. (d) Separation of active fraction (act.) was done with 0.5 eq. (lane 2), 0.2 eq. (lane 3) and 0.1 eq. (lane 4) of the protein; lane 1 – unselected library (pool).

were observed at all seven positions. However, when the selection was done in the presence of higher protein concentration, the nucleotides 6–8 (AUU) displayed lower interference effects compared to the other critical positions. This shows that the negative effect caused by the substitution with Δ can be partially compensated, and implies differential effects of individual nucleotides on the overall affinity. Similar effects of different RNA/protein ratios were reported when separation of RNA mutants was done with immobilized U1A protein.³ In contrast, nucleotides 4 and 5 (CC), which form the loop-closing base-pairs, exhibited strong interference effects under all separation conditions, as the intact base pairs are needed for formation of the high affinity binding-competent hairpin structure.

In conclusion, we demonstrated that our new strategy enables simultaneous identification of all essential areas of a given nucleic acid motif in an experimentally simple procedure. All needed reagents, including phosphoramidite **1**, are commercially available. In addition, the separation of functionally active and inactive library variants upon interaction with their cognate substrates is performed in solution (*i.e.* no immobilisation on a solid support needed, thus avoiding potential surface effects). Furthermore, our study showed that the magnitude of the interference effects at some positions strongly depends on the experimental parameters used to separate the active from inactive library variants. This opens up the possibility to gain

deeper insights into mutation tolerance of individual nucleotides by performing the analyses under various conditions. For example, variation of metal ions and their concentrations¹² can identify cofactor binding sites, or variation of reaction times for the separation step can yield specific kinetic information.

The three carbon spacer unit used in our study has been successfully applied in the past as nucleoside replacement to investigate sequence and structure requirements of ribozymes,¹³ ribonucleoproteins,¹⁴ siRNAs,¹⁵ hairpins¹⁶ and deoxy-ribozymes.¹⁷ Here, we report for the first time its application in a combinatorial way. This approach is more efficient and less time-consuming than systematic replacement of individual nucleotides and separate analysis of each individual mutant. The comprehensive datasets elucidate all critical nucleotides within RNA motifs and pinpoint non-essential positions. This information contributes to deeper understanding of underlying mechanisms and might further promote the design of functional derivatives.

This work was supported by the DFG (International Research Training Grant IRTG 1422, Metal Sites in Biomolecules) and the Max Planck Society. We thank Jana Schmitzova and Vlad Pena (Macromolecular Crystallography group at MPIIbpc) for help with the preparation of U1A protein and lab members of the Nucleic Acid Chemistry group for helpful discussions.

Notes and references

- 1 T. R. Cech, *Cold Spring Harbor Perspect. Biol.*, 2012, **4**, a006742.
- 2 K. M. Weeks, *Curr. Opin. Struct. Biol.*, 2010, **20**, 295–304; I. Krstic, B. Endeward, D. Margraf, A. Marko and T. F. Prisner, *Top. Curr. Chem.*, 2012, **321**, 159–198; T. Carlomagno, *J. Magn. Reson.*, 2014, **241**, 126–136; S. P. Ryder and S. A. Strobel, *Methods*, 1999, **18**, 38–50.
- 3 D. Ray, H. Kazan, E. T. Chan, L. Pena Castillo, S. Chaudhry, S. Talukder, B. J. Blencowe, Q. Morris and T. R. Hughes, *Nat. Biotechnol.*, 2009, **27**, 667–670.
- 4 J. D. Buenrostro, C. L. Araya, L. M. Chircus, C. J. Layton, H. Y. Chang, M. P. Snyder and W. J. Greenleaf, *Nat. Biotechnol.*, 2014, **32**, 562–568; J. M. Tome, A. Ozer, J. M. Pagano, D. Gheba, G. P. Schroth and J. T. Lis, *Nat. Methods*, 2014, **11**, 683–688; W. Kladwang, C. C. VanLang, P. Cordero and R. Das, *Nat. Chem.*, 2011, **3**, 954–962.
- 5 F. Wachowius, F. Javadi-Zarnaghi and C. Höbartner, *Angew. Chem., Int. Ed.*, 2010, **49**, 8504–8508.
- 6 F. Wachowius and C. Höbartner, *J. Am. Chem. Soc.*, 2011, **133**, 14888–14891.
- 7 B. Samanta and C. Höbartner, *Angew. Chem., Int. Ed.*, 2013, **52**, 2995–2999.
- 8 M. Takeshita, C. N. Chang, F. Johnson, S. Will and A. P. Grollman, *J. Biol. Chem.*, 1987, **262**, 10171–10179; P. A. Kupfer, C. Crey-Desbiolles and C. J. Leumann, *Nucleic Acids Res.*, 2007, **35**, 6846–6853.
- 9 C. Hammann and D. M. Lilley, *ChemBioChem*, 2002, **3**, 690–700.
- 10 D. E. Ruffner, G. D. Stormo and O. C. Uhlenbeck, *Biochemistry*, 1990, **29**, 10695–10702; C. C. Sheldon and R. H. Symons, *Nucleic Acids Res.*, 1989, **17**, 5679–5685; A. Peracchi, L. Beigelman, N. Usman and D. Herschlag, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 11522–11527.
- 11 D. Scherly, W. Boelens, W. J. van Venrooij, N. A. Dathan, J. Hamm and I. W. Mattaj, *EMBO J.*, 1989, **8**, 4163–4170; D. E. Tsai, D. S. Harper and J. D. Keene, *Nucleic Acids Res.*, 1991, **19**, 4931–4936; C. Oubridge, N. Ito, P. R. Evans, C. H. Teo and K. Nagai, *Nature*, 1994, **372**, 432–438; M. J. Law, A. J. Rice, P. Lin and I. A. Laird-Offringa, *RNA*, 2006, **12**, 1168–1178.
- 12 F. Javadi-Zarnaghi and C. Höbartner, *J. Am. Chem. Soc.*, 2013, **135**, 12839–12848.
- 13 S. Schmidt, L. Beigelman, A. Karpeisky, N. Usman, U. S. Sorensen and M. J. Gait, *Nucleic Acids Res.*, 1996, **24**, 573–581.
- 14 D. J. Williams and K. B. Hall, *J. Mol. Biol.*, 1996, **257**, 265–275.
- 15 T. C. Efthymiou, B. Peel, V. Huynh and J. P. Desaulniers, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 5590–5594.
- 16 W. Pils and R. Micura, *Nucleic Acids Res.*, 2002, **28**, 1859–1863.
- 17 B. Wang, L. Cao, W. Chiuman, Y. Li and Z. Xi, *Biochemistry*, 2010, **49**, 7553–7562.