2’-SCF₃ Uridine—A Powerful Label for Probing Structure and Function of RNA by ¹⁹F NMR Spectroscopy

Katja Fauster, Christoph Kreutz, and Ronald Micura*

Magnetic resonance methods that employ the ¹⁹F nucleus for cellular imaging or molecular structure and dynamics investigations become increasingly important for both in vitro and in vivo systems. Fluorine is hardly encountered in biomolecules and therefore provides excellent bioorthogonality. However, what is an advantage on the one hand, can become an obstacle on the other hand because proper methods for labeling are required. Although an elegant alternative that relies on “spy molecules”, which contain the fluoride sensor, has been reported recently, directly labeling of either the target or the interaction partner remains inevitable for the major success of applications.

Our group has a strong focus on ¹⁹F labeling of RNA in order to utilize the corresponding derivatives for structural and functional analysis. For instance, we disclosed a gene-regulation-determining, bistable sequence element in the preQ₁ class I riboswitch based on strategically positioned 5-F regulation-determining sequence element in the fact that three magnetically equivalent fluorine atoms allow majority of successful applications. In another example, we introduced ribose 2’-F atoms at specific nucleoside positions, allowing local monitoring of binding events and thus visualizing dynamic RNA–ligand interactions. Although being powerful, in all these cases, the reporter unit relied on a single fluorine atom, and thus limitations with respect to sensitivity could potentially be encountered.

Herein, we present a novel high-performance fluorine sensor for RNA, the ribose 2’-deoxy-2’-trifluoromethylthio unit (2’-SCF₃; Figure 1). The advantage of this label lies in the fact that three magnetically equivalent fluorine atoms allow ¹⁹F NMR experiments to be performed at micromolar concentrations. This labeling method thus constitutes a significant improvement compared to the above-mentioned single-atom labels, which require RNA concentrations in the low millimolar range; less material is needed and potential aggregation problems are minimized. Moreover, the 2’-SCF₃ group represents an isolated spin system, therefore proton decoupling (as, for example, required for 2’-F labels) is not necessary, thus making the label metrologically very straightforward. Similar to methyl groups, trifluoromethyl groups allow the prolongation of coherence lifetime based on transverse relaxation optimized spectroscopy (TROSY), accounting for an additional advantage in measurements of large RNA molecules or RNA–protein systems.

Originally, we considered to develop a new ¹⁹F label for RNA applications through trifluoromethylation. The selection of an appropriate nucleoside position is critical because several sites are to be excluded. For example, 5-trifluoromethyl uridine is chemically unstable during oligonucleotide deprotection as it can transform into a 5-cyano group. The 5 position of uridine can be functionalized by the sterically demanding 4,4,4-trifluoro-3,3-bis(trifluoromethyl)butyne residue, and this pyrimidine label with nine equivalent fluorine atoms was successful in ¹⁹F NMR spectroscopic analysis of DNA hybridization. However, we did not pursue such a concept in favor of a uniform labeling pattern at sites that are equivalent in all four standard nucleosides, preferably at the 2 position. In this sense, the logical follow-up consideration was trifluoromethylation of the 2’ hydroxy group of ribose to achieve 2’-OCF₃ labels. To our knowledge, a single study on 2’-OCF₃-modified oligonucleotides has been reported to date, and this refers to 2’-OCF₃ adenosine. Introduction of the modification was achieved via 2’-O-[(methylthio)thiocarbonyl]adenosine by treatment with pyridinium poly(hydrogen fluoride) (HF/pyridine) in the presence of 1,3-dibromo-5,5-dimethylhydantoin (DBH), however, yields were extremely low (22 %). In our own attempts, we were unable to increase the reported yields. Efforts to apply a new class of electrophilic trifluoromethylation reagents based on hypervalent iodine(III) derivatives (Togni reagents) for the zinc-mediated trifluoromethylation of the 2’-OH group of a 5’,3’-O-protected guanosine substrate failed. Furthermore, by using the first mentioned xanthate labeling reagent, we were unable to increase the reported yields. Efforts to apply a new class of electrophilic trifluoromethylation reagents based on hypervalent iodine(III) derivatives (Togni reagents) for the zinc-mediated trifluoromethylation of the 2’-OH group of a 5’,3’-O-protected guanosine substrate failed.
method,[13] we were able to generate minor amounts of 2'-OCF₃ uridine derivative. However, this pyrimidine nucleoside turned out to be unstable, its decomposition resulting in formation of 2,2'-anhydouridine.

These observations prompted us to develop a novel concept for RNA labeling, namely with 2'-SCF₃ nucleosides, which appears rather unorthodox at first sight. Although such a label would most likely thermodynamically destabilize an RNA double helix (assuming that its behavior would be analogous to 2'-SCH₃ residues),[12] the many promising ¹⁹F NMR applications for probing structure and folding of RNA, binding of small molecules and RNA, or protein–RNA interactions, for which this label can be easily positioned in single-stranded regions, prompted us to pursue this goal.

We started our endeavors with 2'-deoxy-2'-mercaptopuridine 1 (Scheme 1) which is readily accessible in large amounts from 2,2'-anhydouridine and thiaoacetic acid, according to an early report.[13] Fortunately, the key step of our synthetic plan, the regioselective trifluoromethylation of the thiol group, was achieved in 80% yield using 3,3-dimethyl-1-(trifluoromethyl)-1,2-benziodoxole (1.2 equiv), CH₃OH, –78 °C → RT, 16 h, 80%; by DMT-Cl (1.1 equiv), DMAP (0.1 equiv), pyridine, RT, 16 h, 81%; c) (2-cyanoethyl)-N,N-diisopropylchlorophosphoramidite (1.3 equiv), N-ethylisopropylamine (10 equiv), CH₂Cl₂, RT, 2.5 h, 93%.

Scheme 1. Synthesis of the 2'-deoxy-2'-trifluoromethylthio (2'-SCF₃) uridine building block 4 for RNA solid-phase synthesis. Reagents and conditions: a) 3,3-dimethyl-1-(trifluoromethyl)-1,2-benziodoxole (1.2 equiv), CH₃OH, –78 °C → RT, 16 h, 80%; b) DMT-Cl (1.1 equiv), DMAP (0.1 equiv), pyridine, RT, 16 h, 81%; c) (2-cyanoethyl)-N,N-diisopropylchlorophosphoramidite (1.3 equiv), N-ethylisopropylamine (10 equiv), CH₂Cl₂, RT, 2.5 h, 93%.

Confirming liquid-chromatography (LC) electrospray-ionization (ESI) mass spectrometry (MS). Synthesized RNA sequences containing 2'-SCF₃ uridine labels are listed in Table 1 in the Supporting Information. Noteworthy, the 2'-SCF₃ label was completely stable under repetitive oxidative conditions (20 mM aqueous iodine solution) required during RNA solid-phase synthesis for transformation of P III to P V. Therefore, no special adaptations of the standard synthesis cycle (as, for example, required for 2-SeCH₃-modified RNA)[19] were necessary to provide high-quality crude products (Figure 2).

The efficient synthetic access to 2'-SCF₃-modified RNA encouraged us to evaluate the new label in ¹⁹F NMR applications. In the following, we present three examples:

1) probing of the secondary structure of bistable RNA sequences;
2) verification of RNA–protein interactions; and
3) attesting rationally designed riboswitch modules.

Figure 3 depicts a 32 nt long RNA sequence (5) that exists in slow conformational exchange of two distinct secondary structures (5' and 5''), as confirmed by comparative imino proton NMR spectroscopy[16] using the truncated reference hairpin 5a. When we labeled this RNA with 2'-SCF₃ at uridine-26 (6), the label lies within a 4 nt loop of fold 6' while it is located in a 7 nt internal bulge of fold 6''. The assignment of secondary structures 6' and 6'' by ¹⁹F NMR spectroscopy is depicted in Figure 3B and provides a ratio of about 55:45 in

tography on a Sephadex G25 column, and RNA sequences were purified by anion-exchange chromatography under strong denaturing conditions (6 M urea, 80 °C; Figure 2). The molecular weights of the purified RNA molecules were
RNA with an apparent domain (U1A-RBD) of this protein binds to its cognate diesterases, as has been shown for the 2 modification is likely to improve resistance against phosphodiesterases in vivo studies. In this context, we should mention that the conserved sequence of AUUGCAC (Figure 4). Hairpin We synthesized the stem–loop RNA molecules comprise the recognition sequence for the small nuclear ribonucleoprotein U1A (Figure 4). [19] The RNA-binding motif of this protein with eight SCF3, labeled analogues. (C) Assignment of folds 6 and 6′ of RNA 6 by 19F NMR spectroscopy. D) Same as C, but in E. coli cell lysate.

Figure 3. Structure probing of a bistable RNA. A) Unmodified RNA,[17] secondary structure model of full-length (5) and reference (5a) RNA (left); imino proton NMR spectra (right). B) Same as (A), but for 2-SCF3, labeled analogues. C) Assignment of folds 6 and 6′ of RNA 6 by 19F NMR spectroscopy. D) Same as C, but in E. coli cell lysate.

Figure 4. Characterization of an RNA–protein interaction. A) left: RNA stem–loop 7 with 2′-SCF3 label (red) and the recognition sequence (blue) for the U1A protein (cyan), Pymol model generated from 3CUL (Protein Data Bank); right: 19F NMR spectra of RNA and U1A RBD[19] mixed at different ratios as indicated. B) Same as (A), but with label (red) at different position. Conditions: [RNA] = 0.3 mM, [protein] = 0.3 mM, [Na2HPO4] = 10 mM, pH 6.0, H2O/D2O = 9:1, 298 K.

favor of 6. The imino proton NMR spectra of modified (6) and unmodified (5) sequences are nearly identical, thus demonstrating that the equilibrium position has not been influenced by the label. This tendency was confirmed for a second bistable RNA (see the Supporting Information, Figure 1) and hence underscores the applicability of the label for secondary-structure probing. Even more satisfying was the observation that in E. coli lysate at a very low RNA concentration of 10 μM, the two folds were readily detectable (Figure 3D), showing the potential of the 2′-SCF3 label for in vivo studies. In this regard, we should mention that the modification is likely to improve resistance against phosphodiesterases, as has been shown for the 2′-SCF3 counterparts.[19]

As a second example, we demonstrate the utility of the 2′-SCF3 label for the verification of RNA–protein interactions. We synthesized the stem–loop RNA molecules 7 and 8, which comprise the recognition sequence for the small nuclear ribonucleoprotein U1A (Figure 4).[19] The RNA-binding domain (U1A-RBD) of this protein binds to its cognate RNA with an apparent Kd of about 2 × 107 M−1. We positioned the 2′-SCF3 moieties within the 10 nt loop, close to the conserved sequence of AUUGCAC (Figure 4). Hairpin 7 showed a major 19F NMR resonance at −39.90 ppm and a minor one up-field in the signal flank, reflecting an additional conformational population in slow exchange, most likely because of a different microenvironment in the loop (Figure 4A). When one equivalent of U1A-RBD was added, a new uniform signal at −39.70 ppm was detected, representative for the high-affinity, conformationally well-defined RNA–protein complex. Likewise, when the label was shifted one nucleotide downstream in stem–loop 8, the same behavior was observed (Figure 4B), demonstrating that the responsiveness of the label is not restricted to a single site, and that it can even be positioned more distant from the interaction site.

For a third example, we designed a novel riboswitch module that consists of only 27 nucleotides and verified its function using the 2′-SCF3 labeling concept (Figure 5). In the emerging field of synthetic biology, such modules are of growing interest to engineer gene-regulation systems,[20] but currently their number is rather limited and refer to only few small-molecule ligands, such as theophylline, tetracyclin, or neomycin.[21] Here, we employed a known tobramycin-sensitive aptamer recognition sequence[22] for the design of a novel switchable RNA module. In the free form, this functional RNA 9 exists in an extended stem–loop conformation (9′; Figure 5). Once tobramycin (tob) is added, it captures the minor conformation of the RNA (9′) that comprises a characteristic 14 nt recognition loop to form a high-affinity complex in the nanomolar range. This implies a ligand-induced rearrangement of the secondary structure, and thus provides the typical characteristics of a riboswitch.[23]

With the focus on 2′-SCF3 labeling and 19F NMR spectroscopy, we showed that the new label allows the straightforward assignment and quantification of the different RNA conformations (9′, 9″, 9″-tob) involved for this riboswitch module (Figure 5D).

One issue that remains to be addressed in more detail is positioning of the label within single-stranded RNA regions. This is advisable because the attachment of 2′-SCF3 groups thermodynamically destabilizes RNA double helices, comparable to their 2′-SCH3 counterparts.[19] UV melting profile analysis of two exemplary hairpins, 5′-GAAGGGCAAC-
CUUCG and the corresponding modified RNA 5′-GAAGGGCAACC(2′-SCF3-U)UCCG, showed a Δ∆G°298K of 1.9 kcal mol⁻¹ (see the Supporting Information, Figure 2 and Table 2). The reason for the destabilization is most likely the preference for the C2'-endo conformation of the modified nucleoside. This theory was supported experimentally and by MD simulations for 2′-SCH3 moieties.[12,18] To provide evidence for a comparable behavior of 2′-SCF3 functionalities, we synthesized a short RNA strand, 5′-UGU(2′-SCF3-U)GC, and determined 2J (H1'–H2') coupling constants by 2D 1H,1H-DQF COSY NMR experiments (see the Supporting Information, Figure 3). For the 2′-SCF3 uridine, a value of 9.9 Hz was determined, accounting for a population of 98% of C2'-endo ribose conformation, which is indeed a strong indication that this modification would cause interference if forced into a C3'-endo conformation, as it demands an A-form RNA double helix.

Taken together, the 19F NMR applications for 2′-SCF3-modified RNA molecules introduced here make this labeling concept a compelling new tool for probing of RNA structure and function, in particular when protein or small-molecule interaction partners are involved. Moreover, we stress that the chemical synthesis of this label is extremely robust and can be performed on large scale. Incorporation of the label into RNA is compatible with standard solid-phase synthesis and deprotection protocols, and thus highly convenient. Additionally, the potential expansion of the labeling concept to all four standard nucleosides holds promise for flexible and widespread applications in order to explore structure and dynamics of biologically relevant RNA sequences through in vitro and in vivo 19F NMR spectroscopic methods.

**Keywords:** nucleosides · riboswitch · RNA · solid-phase synthesis · trifluoromethylation

Received: September 3, 2012
Published online: November 19, 2012

[1] a) J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wüthrich, *Science* **2012**, 335, 1106–1110; b) L. Liu, J. J. Byeon, I. Bahar, A. M. Gronenborn, *J. Am. Chem. Soc.* **2012**, 134, 4229–4235; c) P. Wadhwani, E. Strandberg, N. Heidenreich, J. Bürc, S. Fanghänel, A. S. Ulrich, *J. Am. Chem. Soc.* **2012**, 134, 6512–6515; d) S. Temme, F. Bönner, J. Schrader, U. Flogel, *WIREs Nanomol. Nanobiotechnol.* **2012**, 4, 329–343; e) A. Kiviniemi, P. Virta, *J. Am. Chem. Soc.* **2010**, 132, 8560–8562; f) C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Sronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl, G. J. Pielak, *J. Am. Chem. Soc.* **2010**, 132, 321–327; g) R. Moumne, M. Pasco, E. Prost, T. Lecourt, L. Micouin, C. Tisné, *J. Am. Chem. Soc.* **2010**, 132, 13111–13113; h) S. Cobb, C. Murphy, J. Fluorine Chem. **2009**, 130, 132–143; i) G. Papeo, P. Giordano, M. G. Brasca, F. Buzzo, D. Caronni, F. Ciprandi, N. Mongelli, M. Veronesi, A. N. Vepulski, C. Dalvít, *J. Am. Chem. Soc.* **2007**, 129, 5665–5672; j) M. Hennig, L. G. Scott, E. Sperling, W. Berme, J. R. Williamson, *J. Am. Chem. Soc.* **2007**, 129, 14911–14921; k) M. Oliènjczak, Z. Gdaniec, A. Fischer, T. Grabarkiewicz, L. Bielecki, R. W. Adamiañ, *Nucleic Acids Res.* **2002**, 30, 4241–4249; l) C. Hammann, D. G. Norman, D. M. J. Lilley, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 5503–5508; m) B. Luy, J. P. Merino, J. Biomol. NMR **2001**, 20, 39–47; n) B. Reif, V. Wittmann, H. Schwalbe, C. Griesinger, K. Werners, K. Jahn-Hoffmann, J. Engels, W. Berme, *Helv. Chim. Acta* **1997**, 80, 1952–1971; o) W. C. Chu, J. Horowitz, *Nucleic Acids Res.* **1989**, 17, 7241–7252.

[2] T. Lombès, R. Moumne, V. Larue, E. Prost, M. Catala, T. Lecourt, F. Darcel, L. Micouin, C. Tisné, *Angew. Chem.** 2012**, 124, 9668–9672; *Angew. Chem. Int. Ed.** 2012**, 51, 9530–9534.

[3] a) C. Kreutz, K. Kählig, R. Konrat, R. Micura, *J. Am. Chem. Soc.* **2005**, 127, 11585–11599; b) D. Gruber, H. Moroder, R. Micura, *J. Am. Chem. Soc.* **2008**, 130, 17230–17231; c) D. B. Puffer, C. Kreutz, U. Rieder, M.-O. Ebert, J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wüthrich, *Science* **2010**, 329, 8950–8958; d) B. E. Kiviniemi, P. Virta, *J. Am. Chem. Soc.* **2012**, 134, 11928–11931.

[4] U. Rieder, C. Kreutz, R. Micura, *Proc. Natl. Acad. Sci. USA* **2010**, 107, 10804–10809.

[5] C. Kreutz, H. Kählig, R. Konrat, R. Micura, *Angew. Chem.** 2006**, 118, 3528–3532; *Angew. Chem. Int. Ed.** 2006**, 45, 3450–3453.

[6] A. Studer, *Angew. Chem.** 2012**, 124, 9082–9090; *Angew. Chem. Int. Ed.** 2012**, 51, 8950–8958.

[7] a) J. C. Markley, P. Chirakul, D. Sologub, S. T. Sigurdsson, *Biosorg. Med. Chem. Lett.* **2001**, 11, 2453–2455; b) W. H. Gmeiner, T. R. Pon, J. W. Lown, *J. Org. Chem.* **1991**, 56, 3602–3608.

[8] N. B. Barhate, R. N. Barhate, P. Cekan, G. Drobsny, S. T. Sigurdsson, *Org. Lett.* **2008**, 10, 2745–2747.

[9] N. Nishizono, Y. Sumita, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **1998**, 26, 5067–5072.

[10] R. Müller, K. Stanek, D. Stolz, R. Aardoom, K. Niedermann, A. Togni, *Angew. Chem.** 2009**, 121, 4396–4400; *Angew. Chem. Int. Ed.** 2009, 48, 4332–4336.

[11] C. Thibaudeau, N. Nishizono, Y. Sumita, A. Matsuda, J. Chattopadhyaya, *Nucleosides Nucleotides Nucleic Acids** 1999*, 18, 1035–1053.

[12] a) A. Fraser, P. Wheeler, P. Dan Cook, Y. S. Sanghvi, *J. Heterocycl. Chem.* **1993**, 30, 1277–1287; b) D. Venkateswarlu, K. E. Lind, V. Mohan, M. Manoharan, D. M. Ferguson, *Nucleic Acids*...
[13] a) M. Imazawa, T. Ueda, T. Ukita, Tetrahedron Lett. 1970, 11, 4807–4810; b) K. J. Divakov, A. Mottoh, C. B. Reese, Y. S. Sanghvi, J. Chem. Soc. Perkin Trans. 1 1990, 969–974; c) L. Le Hir de Fallois, J.-L. Decout, M. Fontecave, J. Chem. Soc. Perkin Trans. 1 1997, 2587–2596; d) L. Zhu, P. S. Lukeman, J. W. Canary, N. C. Seeman, J. Am. Chem. Soc. 2003, 10178–10179.

[14] a) I. Kieltsch, P. Eisenberger, A. Togni, Angew. Chem. 2007, 119, 768–771; Angew. Chem. Int. Ed. 2007, 46, 754–757; b) S. Capone, I. Kieltsch, O. Flögel, G. Lelais, A. Togni, D. Seebach, Helv. Chim. Acta 2008, 91, 2035–2056.

[15] a) S. Pitsch, P. A. Weiss, J. Jenny, A. Stutz, X. Wu, Heli. Chim. Acta 2001, 84, 3773–3795; b) R. Micura, Angew. Chem. 2002, 114, 2369–2373; Angew. Chem. Int. Ed. 2002, 41, 2265–2269; c) F. Wachowius, C. Hobartner, ChemBioChem 2010, 11, 469–480.

[16] C. Hobartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang, A. Polonskaia, A. Serganov, R. Micura, J. Am. Chem. Soc. 2005, 127, 12035–12045.

[17] a) C. Hobartner, M.-O. Ebert, B. Jaun, R. Micura, Angew. Chem. 2002, 114, 619–623; Angew. Chem. Int. Ed. 2002, 41, 605–609; b) C. Hobartner, H. Mittendorfer, K. Breuer, R. Micura, Angew. Chem. 2004, 116, 4012–4015; Angew. Chem. Int. Ed. 2004, 43, 3922–3925.

[18] a) C. Oubridge, N. Ito, P. R. Evans, C. H. Teo, K. Nagai, Nature 1994, 372, 432–438; b) A. R. Ferré, D’Amaré, Methods 2010, 52, 159–167.

[19] C. Hobartner, R. Micura, J. Mol. Biol. 2003, 325, 421–431.

[20] a) A. L. Chang, J. J. Wolf, C. D. Smolke, Curr. Opin. Biotechnol. 2012, 23, 679–688; b) H. Gu, K. Furukawa, R. R. Breakey, Anal. Chem. 2012, 84, 4935–4941.

[21] a) S. Hanson, G. Bauer, B. Fink, B. Suess, RNA 2005, 11, 503–511; b) M. Wieland, A. Benz, B. Klaus, J. S. Hartig, Angew. Chem. 2009, 121, 2753–2756; Angew. Chem. Int. Ed. 2009, 48, 2715–2518; c) E. Duchtart-Fernier, J. E. Weigand, O. Ohlen-schläger, S. R. Schmidtke, B. Suess, J. Wornert, Angew. Chem. 2010, 122, 6352–6355; Angew. Chem. Int. Ed. 2010, 49, 6216–6219.

[22] L. Jiang, D. J. Patel, Nat. Struct. Biol. 1998, 5, 769–774.

[23] a) R. R. Breakey, Mol. Cell 2011, 43, 867–879; b) A. Haller, M. F. Soulrière, R. Micura, Acc. Chem. Res. 2011, 44, 1339–1348; c) C. D. Eichhorn, J. Feng, K. C. Suddala, N. G. Walter, C. L. Brooks III, H. M. Al-Hashimi, Nucleic Acids Res. 2012, 40, 1345–1355; d) J. Rinnenthal, J. Buck, J. Ferne, A. Wacker, B. Fürtig, H. Schwalbe, Acc. Chem. Res. 2011, 44, 1292–1301.

[24] A. I. Haziri, C. J. Leumann, J. Org. Chem. 2012, 77, 5861–5869.