Short, synthetic and selectively $^{13}$C-labeled RNA sequences for the NMR structure determination of protein–RNA complexes

Philipp Wenter, Luc Reymond, Sigrid D. Auweter, Frédéric H.-T. Allain, and Stefan Pitsch

ABSTRACT

We report an optimized synthesis of all canonical 2'-O-TOM protected ribonucleoside phosphoramidites and solid supports containing $^{[13C_5]}$-labeled ribose moieties, their sequence-specific introduction into very short RNA sequences and their use for the structure determination of two protein–RNA complexes. These specifically labeled sequences facilitate RNA resonance assignments and are essential to assign a high number of sugar–sugar and intermolecular NOEs, which ultimately improve the precision and accuracy of the resulting structures. This labeling strategy is particularly useful for the study of protein–RNA complexes with single-stranded RNA in solution, which is rapidly an increasingly relevant research area in biology.

INTRODUCTION

Unambiguous RNA resonance assignments and identification of a sufficient number of intermolecular NOEs is often one of the limiting factors preventing a rapid structure determination of protein–RNA complexes with NMR spectroscopy (1,2). Reasons for this limitation originate from the severe resonance overlap that is present in RNA compared to proteins especially within the sugar protons of the RNA that all resonate within one p.p.m. except for the anomeric proton (3). To overcome this problem, methods based on in vitro transcription have been developed to isotopically label RNA with $^{13}$C and $^{15}$N and to facilitate RNA resonance assignments and the structure determination of RNAs and protein–RNA complexes (4,5). Although, they have been successfully used to study RNA of 20 nt or more in size (4,5), the labeling methods to produce short RNA oligonucleotides (2 to 10 in size) by in vitro transcription failed (6). Yet, isotopic labeling of short RNAs would be very useful to confidently assign the resonances of short RNAs in complex with proteins and to improve the precision of the resulting structures. Recently, three structures of proteins in complex with short single-stranded RNAs have been determined with NMR spectroscopy, namely SF1 (7), TIS11d (6) and the Argonaute PAZ domain (8). All three were studied with unlabeled RNA making the resonance assignment of the RNA in complex very challenging, especially in the presence of an RNA with low sequence complexity like 5'-UUAAUUAAUU-3' used in the TIS11d-RNA complex (6).

Here, we report the synthesis of several short RNA sequences (4 to 7 nt) with $^{[13C_5]}$-labeled riboses at specific positions and their application for the resonance assignments and for the structure determination of two alternative-splicing factors, polypyrimidine tract binding protein (PTB) and feminizing locus on X (Fox-1), in complex with RNA. For this purpose, we evaluated and optimized the various reported synthetic methods (9–12) for the preparation of $^{[13C_5]}$-ribonucleosides from $^{[13C_6]}$-D-glucose. These intermediates were then converted into the corresponding 2'-O-(triisopropylsilyl)oxy)methyl (= 2'-O-TOM) protected ribonucleoside phosphoramidites, which allows an efficient preparation of RNA sequences under DNA coupling conditions (13,14). Chemical synthesis allows efficient $^{13}$C-labeling of very short RNAs and the possibility to introduce isotope-labeling at any position in the sequence, providing a fundamental advantage over isotope-labeling by in vitro transcription (1,2). Using chemically synthesized $^{13}$C-labeled RNA, RNA resonance assignment was greatly facilitated as several ambiguities could be resolved. Also a very high number of intermolecular and intramolecular sugar–sugar NOE cross-peaks could be assigned permitting to obtain structures of very high precision (15,16).

MATERIALS AND METHODS

General

Reagents and solvents (highest purity) from various suppliers, used without further purification, unless otherwise stated.

*To whom correspondence should be addressed. Tel: 0041 21 6939380; Fax: 0041 21 6939380; Email: Stefan.Pitsch@epfl.ch
*Correspondence may also be addressed to Frédéric H.-T. Allain. Tel: 0041 44 6333940; Fax: 0041 44 6331294; Email: allain@mol.biol.ethz.ch

© 2006 The Author(s).
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
[1,2,3,4,5,6-13C6] glucose (13C > 99%) from Spectra Stable Isotopes (Columbia, MD). Work up implies distribution of the reaction mixture between CH2Cl2 and saturated aqueous NaHCO3 solution, drying of the organic layer (MgSO4), and evaporation. TLC: pre-coated silica gel plates from Merck, stained by dipping into a solution of anisaldehyde (10 ml), H2SO4 (10 ml) and AcOH (2 ml) in EtOH (180 ml) and subsequent heating with a heat-gun. Column chromatography (CC): silica gel 60 (230–400 mesh) from Fluka. NMR (Bruker): chemical shift δ in p.p.m., relative to external standards (H– and 13C– and 31P: 85% aqueous H3PO4), coupling constants J in Hz; multiplicities (s) according to DEPT-spectra. ESI-MS (positive mode): SSQ 710 (Finnegan), measurements in MeCN/H2O/ACOH 50:50:1; MALDI-MS (positive mode): Axima CFR Plus (Kratos/ Shimadzu), matrix: 2,4,6-trihydroxyacetophenone (50 mg/ml), di-ammonium hydrogen citrate (15 mg/ml) in H2O/MeCN (1:1); relative intensity in % as indicated in brackets.

[1,2,3,4,5-13C5]-1,2-Di-O-acetyl-3,5-di-O-benzoyl-α-β-D-ribofuranose (1)
Prepared essentially according to Saito et al. (17), for experimental details see Supplementary Data.

[1',2',3',4',5'-13C5]-6-Allyloxy-2-amino-N2-isobutyryl-5'-O-(4,4'-dimethoxytrityl)-N2-isobutyrylguanosine (4)
A solution of 3 (3.44 g, 4.91 mmol) in CH2Cl2 (20 ml) was treated with Et3NH (1.28 ml, 12.3 mmol), PPh3 (296 mg, 1.13 mmol) and Pd(PPh3)4 (30 mg, 0.03 mmol). After 30 min at 20 °C, SiO2 (9 g) was added and the solvent was evaporated. The residue was purified by CC [10 g SiO2, CH2Cl2 (2% Et3N) → CH2Cl2/MeOH 95:5 (+2% Et3N)] to give 4 (2.90 g, 89%). TLC (CH2Cl2/MeOH 9:1) Rf 0.47. 1H-NMR (13C decoupled, 400 MHz, CDCl3): 1.13 (6H, d, J = 6.6, Me2CH); 2.62 (heptet, J = 6.8, Me2CHCO); 3.20 [d, J = 9.7, H–C(5')]; 3.37 [d, J = 10.2, H–C(5')]; 3.45 [br. s, HO]; 3.77 [t, 2 MeO]; 3.79 [br. s, HO]; 4.42 [m, H–C(3')]; 4.50 [m, H–C(5')]; 4.95 [s, H–C(2')]; 5.09 [d, J = 5.4, OCH2CH = CH2]; 5.33 [dd, J = 10.4, 8.1, OCH2CH = CH2]; 5.48 [dd, J = 17.3, 0.8 H, OCH2CH = CH2]; 5.94 [d, J = 5.2, H–C(1')]; 6.15 [m, OCH2CH = CH2]; 6.68–6.75 (m, 4 arom. H); 7.09–7.40 (m, 9 arom. H); 7.46 [br. s, HN–C(2)]; 8.13 [s, H–C(8)]. 13C-NMR (100 MHz, CDCl3): 19.4 (Me2CHCO); 37.2 (Me2CHCO); 55.2 (MeO); 64.2 [d, J = 42, C(5')]; 68.6 (OCH2CH = CH2); 72.2 [dd, J = 38, C(4')]; 77.1 [dd, J = 38, C(2')]; 87.4 [dd, J = 41, C(3')]; 88.8 (arom. C); 92.6 [d, J = 39, C(1')]; 113.5 (arom. CH); 119.2 (OCH2CH = CH2); 127.2, 128.2, 130.3, 130.4, 132.4 (5 arom. CH); 135.8, 135.9 (2 arom. C), 140.4 [C(8)]; 144.7 (arom. C); 151.5 [C(6)]; 152.0 [C(2)]; 158.8 (arom. C), 161.2 (arom. C); 175.7 (Me2CHCO), ESI-MS: 701.31 (100, [M + H]+).

[1',2',3',4',5'-13C5]-5'-O-(4,4'-dimethoxytrityl)-N2-isobutyrylguanosine (4)
A solution of 3 (3.44 g, 4.91 mmol) in CH2Cl2 (20 ml) was treated with Et3NH (1.28 ml, 12.3 mmol), PPh3 (296 mg, 1.13 mmol) and Pd(PPh3)4 (30 mg, 0.03 mmol). After 30 min at 20 °C, SiO2 (9 g) was added and the solvent was evaporated. The residue was purified by CC [10 g SiO2, CH2Cl2 (2% Et3N) → CH2Cl2/MeOH 95:5 (+2% Et3N)] to give 4 (2.90 g, 89%). TLC (CH2Cl2/MeOH 9:1) Rf 0.47. 1H-NMR (13C decoupled, 400 MHz, CDCl3): 0.81, 1.00 (2d, J = 6, Me2CHCO); 2.83 (br. q, J = 7, Me2CHCO); 3.19 [m, H–C(3')]; 3.40 (br. d, J = 6.7, HO); 3.47 [m, H–C(5')]; 3.75, 3.76 (2x, 2 MeO); 4.31 [m, H–C(4')]; 4.55 [m, H–C(3')]; 5.14 [s, H–C(2')]; 5.87 [d, J = 5.2, H–C(1')]; 6.75–6.85 (m, 4 arom. H); 7.10–7.50 (m, 9 arom. H); 7.78 [s, H–C(8)]; 9.46 [br. s, HN–C(2)]; 12.16 [br. s, H–N(1)]. 13C-NMR (100 MHz, CDCl3): 18.5, 18.7 (Me2CHCO); 35.8 (Me2CHCO); 55.2 (MeO); 63.7 [d, J = 43, C(5')]; 71.5 (t, J = 38), 73.9 (t, J = 39), 85.1 (t, J = 41); 132.0 (C(2'), C(3'), C(4')); 140.0 [d, J = 41, C(1')]; 113.2 (arom. CH); 121.0 [C(5')]; 126.9, 127.9, 128.1, 129.9, 130.0 (5 arom. CH); 135.7, 136.1 (2 arom. C); 138.9 [C(8)]; 144.7 (arom. C); 147.4 [C(4)]; 148.2 [C(2)]; 155.8 [C(6)]; 158.6 (arom. C); 179.7 (Me2CHCO), ESI-MS: 661.36 (100, [M + H]+' ).
A suspension of 1 (1.70 g, 3.80 mmol) and uracil (0.52 g, 4.64 mmol) in MeCN (19 ml) was treated with N,O-bis[(trimethylsilyl)acetamide (2.80 ml, 11.1 mmol) and stirred at 60°C for 30 min. Then, TMS-OTf (2.44 ml, 13.3 mmol) was added and the mixture was stirred 30 min at 60°C. After usual work up, the resulting intermediate was dissolved in 8 M MeNH2 in EtOH (19 ml) and stirred overnight at 20°C. The solvent was evaporated; the residue was co-evaporated with pyridine and dried overnight under vacuum (0.01 mbar). The solid was dissolved in pyridine (19 ml) and treated with (MeO)2TrCl (1.80 g, 5.32 mmol). A suspension of 18 g SiO2, C2H2Cl2 (40 g SiO2, C2H2Cl2 (40 g SiO2, C2H2Cl2 in 8 M MeNH2 in EtOH (1 ml) for 3 h at 35°C. After the assembly, the solid supports were treated with a 1:1 mixture of 12 M MeNH2 in H2O/8 M MeNH2 in EtOH (1 ml) for 3 h at 35°C. By centrifugation, the supernatant solution was separated from the solid support and evaporated, and the residue was dissolved in 1 M Bu4NF·3H2O solution in THF (1 ml). After 14 h at 20°C, 1 M Tris·HCl buffer (pH 7.4, 1 ml) was added and the solution concentrated to half its volume.

Desalting and purification of RNA sequences S1–S6
The RNA sequences S1–S6 were assembled from the 13C5-labeled phosphoramidites 11–14 and the corresponding 2’-O-TOM protected, unlabeled phosphoramidites on a Gene Assembler Plus (Pharmacia) in a 1 μmol scale according to (14). After the assembly, the solid supports were treated with a 1:1 mixture of 12 M MeNH2 in H2O/8 M MeNH2 in EtOH (1 ml) for 3 h at 35°C. By centrifugation, the supernatant solution was separated from the solid support and evaporated, and the residue was dissolved in 1 M Bu4NF·3H2O solution in THF (1 ml). After 14 h at 20°C, 1 M Tris·HCl buffer (pH 7.4, 1 ml) was added and the solution concentrated to half its volume.


The remaining solution from above (1 ml) was applied on a NAP-10 cartridge (Pharmacia) and eluted with H2O. Two fractions were collected: fraction 1 (1.5 ml) and fraction 2 (1.5 ml). Fraction 2 was again applied to a NAP-10 cartridge and eluted with H2O. The first 1.5 ml solution was combined with fraction 1 and purified by anion exchange–high-performance liquid chromatography (HPLC): Pharmacia Source 15Q column (4.5 × 100 mm), flow 2 ml/min, eluent A: 50 mM NaHCO3, eluent B: 1 M NaOH, detection at 260 nm, elution at 20°C, gradient 0–45% B in 40 min, 1 injection. The fractions containing pure product were pooled and subjected to lyophilization. In order to completely remove the remaining NaHCO3, 0.5 ml H2O were added to the residue, followed by lyophilization (this procedure was carried out twice). The residue was dissolved in H2O (0.5 ml), treated with NaHCO3 (5 mg) and lyophilized twice. Final desalting on a NAP-10 cartridge was performed as described above, resulting in pure RNA sequences S1–S4 as Na+ salts.
Their amount was determined spectrophotometrically. S1 (2 syntheses, 0.68 mg, 0.58 μmol, 29% yield), MALDI-MS (negative mode): 1182 a.m.u. [M+H]+; S2 (2 syntheses, 0.66 mg, 0.56 μmol, 28% yield), MALDI-MS (negative mode): 1172 a.m.u. [M+H]+; S3 (2 syntheses, 1.27 mg, 0.85 μmol, 43% yield), MALDI-MS (negative mode): 1493 a.m.u. [M+H]+; S4 (2 syntheses, 1.26 mg, 0.85 μmol, 43% yield), MALDI-MS (negative mode): 1483 a.m.u. [M+H]+.

Desalting and purification of RNA sequences

\[ r_{[13C5]}^{[\text{U}]\text{Gp}}^\text{[13C5]} \text{Cp}\text{A} \text{p}[_{13C5]}^\text{Gp}\text{U} \] (S5) and
\[ r_{[13C5]}^{[\text{U}]\text{Gp}}^\text{[13C5]} \text{Cp}\text{A} \text{p}[_{13C5]}^\text{Gp}\text{U} \] (S6)

The remaining solution from above (1 ml) was applied on a NAP-10 cartridge (Pharmacia) and eluted with H2O. The first 1.5 ml solution was purified by anion exchange–HPLC: Pharmacia Source 15Q column (4.5 × 100 mm), flow 2 ml/min, eluent A: 12.5 mM Tris–HCl (pH 7.5); eluent B: 12.5 mM Tris–HCl (pH 7.5), 1 M NaCl; detection at 260 nm, gradient 0–50% B in 45 min, two injections. Fractions containing pure product were treated with 1 M aqueous Et3N·H2CO3 to a final 0.1 M concentration and applied to a Sepak cartridge [Waters, conditioned by washing with CH2Cl2,2 0°C].

Preparation of the [13C5]-ribose-labeled 2′-OMT protected ribonucleoside phosphorimidates 11–14. Abbreviations: Ac = acetyl, Bz = benzoyl, Ibu = isobutyl, CE = cyanoethyl, DMT = (4,4′-dimethoxytrityl), TOM = (trispropylsilyl)oxymethyl. Reagents and conditions: (a) Adapted from Saito et al. (17), detailed procedure in Supplementary Data: 1. FeCl3, MgSO4, acetone, 20°C; 2. pyridinium dichromate, Ac2O, CH2Cl2, reflux; 3. H3O+, THF, 20°C; 4. NaBH4, THF/EtOH 1:1, 20°C; 5. BzCl, pyridine, 20°C; 6. Ac2O, AcOH, H2SO4, 20°C. (b) 6-Chloro-N3-isobutyluracil-2-amine, N,O-bis(trimethylsilyl)acetamide (BSA), Me3SiOTf, 1,2-dichloroethane, 65°C. (c) 1. Allyl alcohol, DABCO, DBU, 20°C; 2. NaOH, THF/MeOH/H2O, 0°C; 3. DMT-Cl, pyridine, 20°C; (d) Pd(Ph3P)4, HNEt2, PPh3, CH2Cl2, 20°C; (e) Synthesis of 5. 1. N3-benzoyladename, BSA, SnCl4, 1,2-dichloroethane, 65°C; 2. NaOH, THF/MeOH/H2O, 0°C; 3. DMT-Cl, pyridine, 20°C; synthesis of 6. 1. uracil, BSA, Me3SiOTF, MeCN, 60°C; 2. MeNH2, EtOH, 20°C; 3. DMT-Cl, pyridine, 20°C; (f) Bu3SnCl2, ‘Pr2NEt, TOM-Cl, 1,2-dichloroethane, 80°C according to (14). (g) 1. Ac2O, DMAP, pyridine 25°C; 2. 4-chlorophenyl phosphorodichloridate, 1H-1,2,4-triazole, ‘Pr2NEt, MeCN, 4°C → 20°C; 3. aqueous NH3, dioxane/McCN, 20°C; 4. NaOH, THF/MeOH/H2O, 4°C. 5. Ac2O, DMF, 20°C, according to (18). (h) 2-Cyanoethylidiospropylphosphorimidochloridite, ‘Pr2NEt, CH2Cl2, 20°C, according to (14).
distance constraints from the unlabeled RNA sequence was performed in a restrained simulated annealing run using AMBER 7.0, following the same protocol that was employed for the structure calculation of the Fox-1 complex with the entire set of constraints; see (15).

RESULTS AND DISCUSSION

Preparation of the $^{13}$C$_5$-ribose-labeled ribonucleoside phosphoramidites

For the preparation of the four canonical, $^{13}$C$_5$-ribose-labeled 2'-O-TOM protected ribonucleoside phosphoramidites and solid supports, we adapted and optimized existing reaction conditions with the purpose of minimizing the number of time-consuming purification steps (Scheme 1). The four-step synthesis of the protected $^{13}$C$_5$-ribose derivative 1 from $^{13}$C$_5$-labeled D-glucose (40% yield) was carried out essentially according to Saito et al. (17) and is described in the Supplementary Data. The protected adenosine and uridine derivatives 5 and 6 were obtained by nucleosidation of 1 with N$_{0}$-benzoyladenosine (→ 5) and uracil (→ 6), respectively, under optimized Vorbrüggen conditions, followed by partial deacylation and dimethoxytritylation. The guanosine derivative 6 was prepared according to a modified method originally described for the synthesis of the corresponding ribopyranosylguanosine phosphoramidite by Pittsch et al. (19). In the presence of Me$_3$SiOTf and in 1,2-dichloroethane, 1 reacted smoothly and completely regioselectively with in situ trimethylsilylated N$_{0}$-isobutyl-6-chloro-2-aminopurine to the purine derivative 2. Without purification, it was treated with 1,4-diazabicyclo[2,2,2]octane (DABCO) and 1,8-diazabicyclo[5,4,0]undec-7-ene (DBU) in allyl alcohol, resulting in the corresponding O$^a$-allyl derivative which was partially deacylated with NaOH in THF/MeOH/H$_2$O and then dimethoxytritylated with (MeO)$_2$TrCl in pyridine (→ 3, 65% based on 1). Cleavage of the O$^a$-allyl group was achieved by treating the intermediate 3 with Et$_3$NH in the presence of [Pd(PPh$_3$)$_4$], resulting in formation of the guanosine derivative 4 (89% yield).

The introduction of the 2'-O-TOM protecting group into nucleosides 4, 5 and 6 was carried out according to our standard procedure (14), with Bu$_3$SnCl$_2$, Pr$_2$NEt and TOM-Cl in 1,2-dichloroethane at 80$^\circ$C (detailed procedure in Supplementary Data). The product mixture was separated into the 2'-O-TOM protected products 7–9, the unreacted starting materials, and the 3'-O-TOM protected regioisomers. The latter were deprotected with Et$_3$NF in MeCN, the products were combined with the unreacted starting materials and subjected again to alklylation with TOM-Cl. By this procedure, the 2'-O-TOM protected nucleosides 7–9 were obtained in fair yields of 42–50% (Scheme 1). From the 2'-O-TOM protected uridine 9, the corresponding cytidine derivative 10 was obtained according to our published method (18), in a yield of 65%. Under standard conditions, the four 2'-O-TOM protected nucleosides 7–10 were converted with 2-cyanoethyl diisopropylphosphoramidodichloride/Pr$_2$NEt in CH$_2$Cl$_2$ into their corresponding phosphoramidite building blocks (Scheme 1). The solids supports were prepared according to (14) by first synthesizing the 2'-O-(nitrophenylethane)dioate from the 3'-O-tom nucleosides of adenosine, uridine and cytosine and the 3'-O-(nitrophenylethanthione)dioate from the 2'-O-TOM guanosine derivative. These activated esters were immobilized on 600 Å (aminoalkyl)-functionalized controlled pore glass (CPG) with loadings from 25 to 35 μmol/g (a detailed description for the preparation of phosphoramidites 7–14 and the corresponding solid supports is given in the Supplementary Data).

Chemical synthesis of the short $^{13}$C-labeled RNA oligonucleotides

The short and partially labeled RNA sequences S1–S6 (Table 1) were assembled in 1 μM batches from conventional 2'-O-TOM protected ribonucleoside phosphoramidites and solid supports, and the corresponding 13C-labeled building blocks 11–14, respectively, on an oligonucleotide synthesizer according to our standard protocols (14). The isolation and purification protocols of the shortest 4 and 5mer oligonucleotides S1–S4 had to be modified, because of the small size and small number of negative charges present (for a detailed procedure see Materials and Methods). Finally, all RNA sequences were converted into their Na$^+$ salts and characterized by MALDI-MS (Table 1).

NMR study of the $^{13}$C$_5$-sugar labeled labeled UUCU and CUU in complex with PTB

The oligo(CU) sequences S1–S4 (Supplementary Table) containing $^{13}$C$_5$-sugar labeled nucleotides at various positions were used for structure determination of PTB in complex with its target RNA sequence. We had to use these very short tetramers and pentamers, because PTB RBD1, RBD2 and RBD4 recognize only 3 nt and RBD3 only 5 nt (16). As illustrated in the $^{13}$C HSQC spectrum of S3 bound to PTB RBD34 H134A, complete or partial $^{13}$C-labeling of the sugar moieties of these short RNAs allows resolving the

| Table 1. Preparation of selectively $^{13}$C-labeled RNA sequences |
| --- | --- | --- | --- |
| Synthesis$^a$ | Sequence | Scale | Yield | Purification$^b$ and MALDI-MS$^c$ Condition |
| | | (μmol) | (%) | | [M+H$^+$]$_{calc}$ | [M+H$^+$]$_{found}$ |
| S1: 5'-CUCU-3' | 2 | 0.58 | 29 | A | 1182 | 1182 |
| S2: 5'-CUCU-3' | 2 | 0.56 | 28 | A | 1172 | 1172 |
| S3: 5'-CUCU-3' | 2 | 0.85 | 43 | A | 1493 | 1493 |
| S4: 5'-CUCU-3' | 4 | 0.85 | 43 | A | 1483 | 1483 |
| S6: 5'-CUGU-3' | 4 | 2.05 | 51 | B | 2196 | 2197 |
| UGCAUGU-3' | 4 | 1.56 | 39 | B | 2202 | 2201 |

$^a$Preparation and deprotection of RNA-sequences according to Pittsch et al.(14); underlined letters in sequences indicate $^{13}$C$_5$-ribose-labeled nucleosides.


$^c$Matrix: 2,4,6-trihydroxyacetophenone, di-ammonium citrate.
strong proton resonance overlap in the sugar region of the RNA in complex with PTB (Figure 1A). The RNAs containing only labeled uracils (S2 and S4) were essential to confirm the resonance assignments of the bound RNAs (Figure 1B) and to unambiguously derive intermolecular NOEs between the sugar moiety and the protein side chains. Isotope-labeling of the RNAs was essential to this structure determination as 50 and 67% of the intermolecular NOEs involved RNA sugar resonances in the complex with RBD3 and RBD4 of PTB, respectively (16).

NMR study of the 13C5-sugar labeled UGCAUGU in complex with Fox-1

The partially 13C-labeled RNA oligonucleotides S5 and S6 were used to determine the structure with Fox-1 (16). The two RNAs were essential to unambiguously assign the RNA resonances and to collect a high number of intra- and intermolecular NOE-derived constraints for UGCAUGU in complex with Fox-1. Without isotopically labeled RNA, the resonance assignment of the 7 nt RNA in complex was difficult and not unambiguous. While the proton resonances within each individual nucleoside could be linked using 2D NOESY and 2D TOCSY spectra, it was not possible to unambiguously place each individually assigned nucleotide spin-systems in the RNA sequence. In particular, we were doubtful about which spin-systems were corresponding to U5 and U7. With the 2D 13C HSQC spectra of the partially 13C-labeled RNA S6 and S7 in complex with Fox-1, these ambiguities could be solved, as S6 contains a 13C5-labeled U5 and S7 a 13C5-labeled U7 (Figure 2A). Importantly, with the particular labeling scheme present in S6 and S7, in which 13C5-labeled and unlabeled nucleotides are present in alternation, unusual internucleotide sugar–sugar NOEs could be assigned using an unlabeled RNA, the remaining 34 could only be resolved using labeled S6 and S7 due to ambiguity or resonance overlap. Similarly, of the 149 intermolecular distance constraints, which were used to calculate the Fox-1 complex structure, 83 originate from the sugar protons of the RNA and 57 of these could not have been assigned without the help of the labeled oligonucleotides (15). A structure calculation without the distance constraints from the 13C-labels resulted in a set of structures in which the RNA is significantly less defined: the RMSD of the heavy atoms of the RNA is 1.02 Å as compared to 0.55 Å in the case when all NOE distance constraints were included (Figure 2C). While each nucleotide is affected, the effect is strongest at the 5' and 3' terminal ends of the RNA. In particular, nucleotide U7 does not converge to one particular conformation (Figure 2C). Furthermore, the relative orientation of protein and RNA is less precise (Figure 2C). The complex structure of Fox-1 using all constraints is extraordinarily well defined, allowing a detailed analysis of the individual hydrogen bond, van der Waals and electrostatic interactions at the protein–RNA interface (15). In the structures obtained with the reduced set of constraints, an equal level of precision cannot be reached. Especially the intermolecular contacts to the RNA backbone are more heterogeneous. For example, a salt bridge between lysine 156 and the phosphate between G6 and U7 is observed in 28 of 30 structures (Figure 2C, left). In contrast, this salt bridge is observed in only 11 of the 30 structures obtained with the smaller number of NOE distance constraints (Figure 2C, middle). Hence, these NOEs were critical to determine with high precision the very unusual structure that the RNA adopts when bound to the RBD of Fox-1.

CONCLUSION AND PERSPECTIVES

In conclusion, we present a short and efficient synthetic access to 2′-O-TOM protected ribonucleoside phosphoramidites containing 13C-labeled sugar moieties. The site-specific
Figure 2. (A) $^{13}$C HSQC spectra of ~1 mM Fox-1 in complex with S6 (red) or S7 (blue). (B) Left: section of a 2D NOESY spectrum of ~1 mM Fox-1 (unlabeled) in complex with UGCAUGU (unlabeled) showing the extreme overlap in the RNA sugar resonances. Middle left: the same section of a $\omega_1$, $\omega_2$-filtered 2D NOESY spectrum of ~1 mM Fox-1 (uniformly $^{13}$C-labeled) in complex with UGCAUGU (unlabeled). In this spectrum, NOE cross-peaks arise only from intra-RNA correlations; overlap is reduced but assignments remain ambiguous. Middle right: the same section of a $\omega_1$-filtered, $\omega_2$-edited 2D NOESY spectrum of ~1 mM Fox-1 (unlabeled) in complex with S6. NOE cross-peaks arise from correlations of protons attached to $^{13}$C($\omega_2$) with protons attached to $^{12}$C($\omega_1$). Right: the same section of a $\omega_1$-filtered, $\omega_2$-edited 2D NOESY spectrum of ~1 mM Fox-1 (unlabeled) in complex with S7. NOE cross-peaks arise from correlations of protons attached to $^{13}$C($\omega_2$) with protons attached to $^{12}$C($\omega_1$). Underlined NOE assignments are displayed in the structure of the complex with green dashed lines in (C). Overlap and ambiguity are significantly reduced in the spectra with S6 and S7 and many internucleotide cross-peaks can be assigned due to the alternating labeling scheme. (C) Left: overlay of the 30 final structures of Fox-1 in complex with UGCAUGU calculated using the entire set of NOE distance constraints [PDB code: 2ERR] (15). Middle: overlay of the 30 final structures of Fox-1 in complex with UGCAUGU calculated using only those NOE distance constraints that could be assigned with unlabeled RNA oligonucleotides. The structures are superimposed on the heavy atoms of the RNA; the protein backbone is shown in black, the heavy atoms of the RNA are in red, and the side chains of F126 and K156 are in green and blue, respectively. Right: sections of one representative structure of the Fox-1-UGCAUGU complex calculated using the entire set of NOEs. The protein backbone is depicted in ribbon representation (gray), aromatic protein side chains are shown in black, and the RNA is shown in red. Protons are displayed in gray.
introduction of these $^{13}$C-labeled nucleotides into RNA sequences permitted a straightforward and unambiguous resonance assignment of their complexes with PTB and Fox-1. The analysis of sets of RNA sequences with alternating labeling patterns allowed the assignments of numerous unusual sugar–sugar and intermolecular NOEs, which substantially improved the precision and accuracy of the resulting structures. Use of chemically synthesized $^{13}$C-labeled RNA sequences is likely to facilitate the study of single-stranded RNA binding proteins in complex with their target RNAs (21), particularly RNA binding proteins targeting low complexity or repeat sequences that are numerous in posttranscriptional gene regulation (22–24).

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

This work was supported by the EPFL, the NCCR ‘Structural Biology’ and the Swiss National Science Foundation (Grant No. 2000-06890) to S.P. and by the grants from the Swiss National Science Foundation and the ETH Zurich through the NCCR Structural Biology and by the Roche Research Fund No. 2000-06890) to S.P. and by the grants from the Swiss National Science Foundation and the ETH Zurich to F.H.T.A. F.H.T.A. is an NCCR Structural Biology and by the Roche Research Fund.

**REFERENCES**

None declared.

**CONFLICT OF INTEREST STATEMENT.** None declared.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

This work was supported by the EPFL, the NCCR ‘Structural Biology’ and the Swiss National Science Foundation (Grant No. 2000-06890) to S.P. and by the grants from the Swiss National Science Foundation and the ETH Zurich through the NCCR Structural Biology and by the Roche Research Fund for Biology at the ETH Zurich to F.H.T.A. F.H.T.A. is an EMBO Young Investigator. Funding to pay the Open Access publication charges for this article was provided by the EPFL.

**CONFLICT OF INTEREST STATEMENT.** None declared.

**REFERENCES**


