J Biomol NMR (2014) 59:231–240
DOI 10.1007/s10858-014-9841-3

ARTICLE

Strategy for automated NMR resonance assignment of RNA: application to 48-nucleotide *K10*

Barbara Krähenbühl · Peter Lukavsky · Gerhard Wider

Received: 26 February 2014 / Accepted: 27 May 2014 / Published online: 5 June 2014

Abstract A procedure is presented for automated sequence-specific assignment of NMR resonances of uniformly [^{13}C, ^{15}N]-labeled RNA. The method is based on a suite of four through-bond and two through-space high-dimensional automated projection spectroscopy (APSY) experiments. The approach is exemplified with a 0.3 mM sample of an RNA stem-loop with 48 nucleotides, *K10*, which is responsible for dynein-mediated localization of *Drosophila fs(1)K10* mRNA transcripts. The automated analysis of the APSY data led to highly accurate and precise 3- to 4-dimensional peak lists. They provided a reliable basis for the subsequent sequence-specific resonance assignment with the algorithm FLYA and resulted in the fully automated resonance assignment of more than 80 % of the resonances of the ^{13}C–^{1}H moieties at the 1′, 2′, 5, 6, and 8 positions in the nucleotides. The procedure was robust with respect to numerous impurity peaks, low concentration of this for NMR comparably large RNA, and structural features such as a loop, single-nucleotide bulges and a non-Watson–Crick wobble base pairs. Currently, there is no precise chemical shift statistics (as used by FLYA) for RNA regions which deviate from the regular A-form helical structure. Reliable and precise peak lists are thus required for automated sequence-specific assignment, as provided by APSY.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9841-3) contains supplementary material, which is available to authorized users.

B. Krähenbühl · G. Wider (✉)
Institute of Molecular Biology and Biophysics, ETH Zurich, 8093 Zurich, Switzerland
e-mail: gsw@mol.biol.ethz.ch

P. Lukavsky
CEITEC – Central European Institute of Technology, Masaryk University, 62500 Brno, Czech Republic

Keywords Nucleic acids · NMR · Projection spectroscopy · APSY · Automated assignment · FLYA · Novel sampling methods

Introduction

Non-coding RNAs fulfill a wide range of biologically important functions, particularly in regulation of gene expression and catalysis. NMR plays an important role (Cléry et al. 2012) for investigations of these functions on a molecular level. A bottleneck for detailed structural and functional NMR studies of RNA and RNA–protein complexes is the assignment of the resonances. Sequence-specific assignment of resonances in RNA spectra can be a challenge due to the narrow and intertwined chemical shift ranges of the various atom groups in RNA molecules, and consequential signal overlap in NMR spectra. Isotope labeling with ^{13}C can partially alleviate this problem. However, often the dimensionality or the resolution of conventional NMR experiments is limited. The advent of novel sampling strategies (Billeter and Orekhov 2012), such as, e.g., sparse sampling (Orekhov et al. 2003), G-matrix Fourier transformation (GFT) (Kim and Szyperski 2003), projection-reconstruction spectroscopy (Kupce and Freeman 2003), or automated projection spectroscopy (APSY) (Hiller et al. 2005), has relieved the limitations on resolution and dimensionality by allowing the efficient acquisition of four or higher dimensional spectra with adequate resolution. The APSY method has proved to provide a range of benefits, which allow the use of the resulting data for efficient automated assignment of protein spectra (Hiller et al. 2008a, b; Krähenbühl et al. 2013; Krähenbühl and Wider 2012; Narayanan et al. 2010; Ranjan et al. 2011;

Fiorito et al. 2006). This fact strongly implicates that APSY might also be beneficial for RNA resonance assignments.

There is already a variety of high-dimensional NMR experiments for RNA, that have either been presented recently in connection with novel sampling strategies, or that were sampled conventionally with low resolution. Published experiments for RNA that were measured with novel sampling strategies are a 5D APSY-HCNCH experiment (Krähenbühl et al. 2012), a set of two 3D APSY-HCN experiments (Krähenbühl et al. 2012), and 4D HCCH COSY and HC(C)CH COSY experiments measured with GFT (Kim and Szyperski 2003; Atreya et al. 2012) or the filter diagonalization method (FDM) (Douglas et al. 2008). Conventionally sampled 4D experiments for RNA are, e.g., a 4D HMQC-NOESY-HMQC experiment (Nikonowicz and Pardi 1992). Further, there is a set of high-dimensional experiments using APSY that provides automated assignment for small RNAs (Krähenbühl et al. 2014). Here, we present a suite of partially novel high-dimensional NMR experiments (Fig. 1) for larger RNAs. When acquired using the APSY concept with uniformly [^{13}C, ^{15}N]-labeled RNA precise high-dimensional peak lists are obtained. APSY measures high-dimensional NMR experiments via 2-dimensional projection spectra (Kim and Szyperski 2003; Kupce and Freeman 2003), from which highly accurate and precise high-dimensional peak lists can be extracted by the algorithm GAPRO (Hiller et al. 2005). The precision of APSY peak lists can be 20 to 50 times higher than the spectral resolution of individual spectra (Krähenbühl et al. 2012; Hiller et al. 2007; Hiller and Wider 2012) since the positions of several projection peaks are utilized for the calculation of each final chemical shift. These peak lists provide an ideal basis for robust automated sequential assignment of RNA resonances, as we show using the FLYA algorithm (Aeschbacher et al. 2013; Schmidt and Güntert 2012). A flow chart of our assignment strategy is presented in the Supplementary Material (Fig. S1).

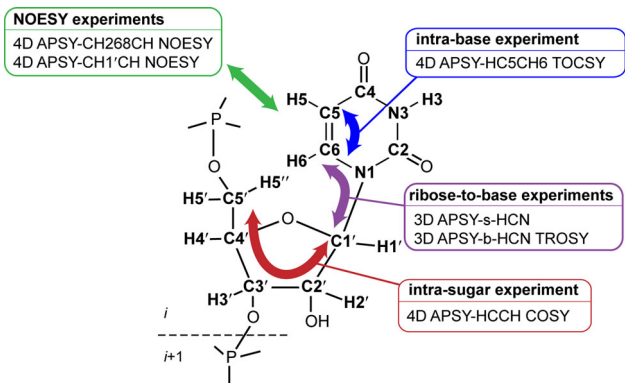

Fig. 1 Illustration of a set of 3- and 4-dimensional APSY experiments that allows automated sequence-specific assignment of non-labile NMR resonances of uniformly [^{13}C, ^{15}N]-labeled large RNAs. The base in this exemplary representation is uracil. The *dashed line* separates the nucleotides i and $i + 1$. All nuclei that are covered by any of the presented APSY experiments are indicated by bold letters using the IUPAC rules for nucleic acids (IUPAC-IUB 1983); the remaining nuclei are illustrated in regular letters. The *red arrow* symbolizes experiments for intra-sugar correlations represented here by a 4D APSY-HCCH COSY experiment. The *purple arrow* illustrates the ribose-to-base correlations achieved for large RNA with a set of two 3D APSY-HCN experiments. The *blue arrow* indicates the 4D APSY-HC5CH6 TOCSY experiment used to establish the intra-pyrimidine correlations. The *green arrow* symbolizes through-space NOESY experiments that deliver intra-nucleotide as well as sequential and non-sequential inter-nucleotide correlations; we propose two selective NOESY APSY experiments: one starting on the ribose CH1′ group, and the other on the base CH2/6/8 groups

As proof of principle the APSY based assignment strategy was applied to the 48-nucleotide RNA stem-loop, *K10*, responsible for dynein-mediated localization of *Drosophila fs(1)K10* mRNA transcripts (Bullock et al. 2010; Serano and Cohen 1995). Since the goal of our strategy is to use only a single sample, *K10* was dissolved in D_2O sacrificing the signals of exchangeable protons; we worked with a concentration of approximately 0.3 mM. Unfortunately, the sample was inhomogeneous and further degraded during the measurements (see Fig. 2) which unintentionally provided a challenging test for our procedure. Despite these problems we were able to achieve more than 80 % resonance assignment of the ^{13}C–^{1}H moieties at the 1′, 2′, 5, 6, and 8 positions in the nucleotides.

Materials and methods

All NMR experiments were measured at a temperature of 25 °C on one sample of 0.3 mM *K10* with 48 nucleotides and a molecular weight of 17 kDa. The total spectrometer time used was 10.7 days. Uniformly [^{13}C, ^{15}N]-labeled *K10* RNA was prepared by in vitro transcription from linearized plasmid DNA using phage T7 RNA polymerase and in-house prepared [^{13}C, ^{15}N]-labeled NTPs as described previously (Bullock et al. 2010). The RNA was purified by weak anion-exchange chromatography under native conditions (Easton et al. 2010). The final sample contained *K10* in an initial concentration of 0.35 mM (as determined by PULCON (Wider and Dreier 2006) in a volume of 300 µl D_2O in a Shigemi® NMR tube, with 10 mM sodium phosphate buffer at pH 6.0 and 0.25 mM EDTA-d12 (^{2}H-labeled). *K10* has a loop with 8 nucleotides, two single-nucleotide bulges (C35 and A39), a wobble base pair (G6-U43), and a region with five consecutive U-A base pairs; its secondary structure is sketched in the center of the illustration in Fig. 3. These structural irregularities have an impact on automated resonance assignment, but the presented strategy can cope with these challenges.

The goal of our strategy is to use a single sample for the assignment procedure; thus the question arises whether

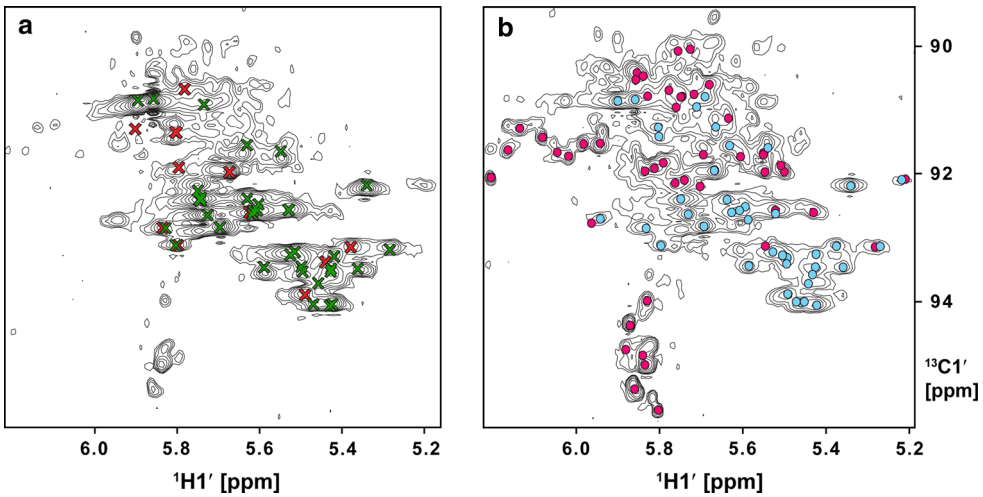

Fig. 2 Expansions of two constant-time 2D [^{1}H, ^{13}C]-HSQC spectra of the *K10* sample: the spectral region contains the correlation peaks of the H1′-C1′ nuclei. The spectra were acquired with identical parameters and conditions; but **a** was measured at the beginning of the APSY experiment series, and **b** at the end 1 month later. The sample was a total of 2.5 weeks in the spectrometer at 25 °C, and in the remaining time stored at 4 °C. In **a**, the *green crosses* indicate correct automated assignments by the method presented in this work, and *red crosses* assignments that are not in agreement with the manual assignment (Bullock et al. 2010); please note that from the expected 34 *green* and 14 *red crosses* (compare to Fig. 3) not all are clearly visible due to overlap. All *dots* in **b** identify peaks that were contained in the 4D APSY-HCCH COSY peak lists (second ^{13}C–^{1}H group): the *cyan dots* denote *K10* correlation peaks, whereas the *magenta dots* denote real RNA correlation peaks that belong to impurities. A comparison of **a** and **b** reveals that the number of impurity peaks, as well as their intensity, increased significantly during the measurements, along with the attenuation of the *K10* peaks, documenting an ongoing degradation process

H_2O or D_2O should be preferred as solvent. Measurements in H_2O have the advantage to provide correlation information also for exchangeable protons of the ^{1}H–^{15}N groups in the bases of RNA. Since the resonance frequencies of the ribose protons are often close to the resonance frequency of water protons, the latter cannot be selectively excited, and the water signal can only be efficiently suppressed by gradient selection of the desired coherence transfer pathway. For larger RNAs such a suppression scheme reduces the sensitivity due to additional relaxation, and saturation transfer via the exchangeable protons. Our strategy thus uses a sample with RNA dissolved in D_2O. If the assignment of imino and amino signals is required, e.g. for a high resolution structure determination, we propose to use conventional methods for these nuclei (Bullock et al. 2010).

The APSY experiments were developed and measured on three different BRUKER AVANCE III spectrometers: a 600 MHz spectrometer equipped with a cryogenic ^{1}H/^{13}C/^{15}N/^{31}P quadruple resonance probe, as well as a 700 MHz and a 900 MHz spectrometer equipped with ^{1}H/^{13}C/^{15}N triple resonance cryogenic probes. The spectrometers were controlled with the software Topspin 3.1 (Bruker, Karlsruhe, Germany). Details on the setup of APSY experiments are given in the Supplementary Material. In the following we present an overview of the novel APSY experiments for RNA. Figure 1 illustrates the connectivity patterns of the individual experiments used for the automated assignment strategy.

For sequential assignments, correlations between neighboring sugar moieties are required. Through-bond correlations require a transfer via ^{31}P nuclei in the phosphodiester group of the oligonucleotide backbone. ^{31}P has a conformation-dependent scalar coupling of 3–20 Hz to the C3′, C4′ or C5′ carbon nuclei of the same or the previous residue. Correlation experiments connect ^{31}P to only one of its neighboring atoms, so that sequence-specific through-bond assignment requires the unambiguous connection via ^{31}P resonances which are distributed only in a very small chemical shift range of 0.5–1.2 ppm. Further, ^{31}P experiences strong chemical shift anisotropy (CSA) relaxation, which broadens its signals making their separation even more difficult. As expected from these considerations, connections via ^{31}P proved ambiguous for the case of *K10* and will be for most RNAs with a size in this range. Thus, our strategy relies on through-space experiments (based on NOESY) to establish connections between neighboring nucleotides and on through-bond experiments within individual nucleotides.

NOESY experiments

A set of two selective 4D APSY-CHCH NOESY experiments was developed based on existing conventional 4D

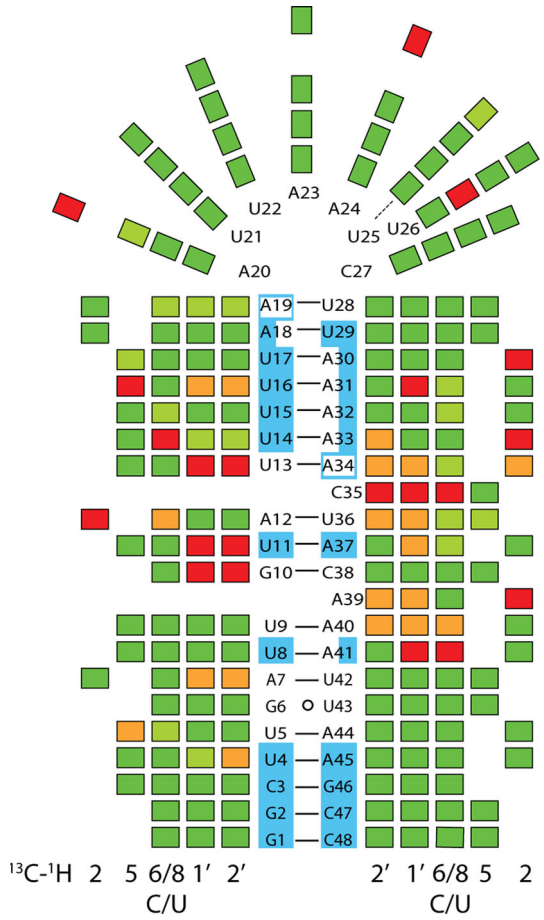

Fig. 3 Illustration of the completeness of the automated assignment based on a series of high-dimensional APSY experiments (see Fig. 1). Horizontally, the groups of 3 to 4 *green* or *red squares* indicate different CH groups of individual nucleotides; at the *bottom* the standard atom numbers are given. In the *center*, the secondary structure of *K10* is outlined. The *horizontal lines* in the center connect Watson–Crick base pairs, whereas the *circle* connects the non-Watson–Crick wobble base pair G6-U43. For nucleotides with a *full cyan background* very precise sequence-specific chemical shift statistics exists (Aeschbacher et al. 2013). For the adenosine residues with a partial *cyan background* there is no precise C2 chemical shift statistics, and for the adenosine residues with a cyan frame only base-type specific information (details explained in (Aeschbacher et al. 2013)). All nucleotides with a *white background* were mapped to the general FLYA statistics, which does not consider neighboring nucleotides, and is considerably less precise. *Green rectangles* denote CH groups that were correctly assigned by the automated strategy compared to the previous manual assignment (Bullock et al. 2010); among them, the CH groups marked with *lime green rectangles* are those whose quality value lies below 50 % for one or both nuclei (explained in "Materials and methods"). *Red and orange rectangles* denote CH groups for which the assignment differed from the manual one which is assumed correct (see text): the *red ones* have FLYA quality values above 50 %, and the *orange ones* quality values below 50 %. *Missing rectangles* signify that the base of this residue does not contain the CH group of this column (only *C* and *U* have a proton attached to C5, and only *A* a proton attached to *C2*)

CHCH NOESY experiments (Clore et al. 1991; Vuister et al. 1993). Pulse sequence schemes (Fig. S2) and experimental details are provided in the Supplementary Material. The selectivity of these experiments concerns the first ^{13}C–^{1}H group: the C2/6/8 resonances in the bases, or in a separate APSY experiment the C1′ resonances in the ribose. These resonances are selectively inverted in the first ^{1}H-^{13}C polarization transfer period. The subsequent ^{13}C evolution includes two selective C5 or C2′ inversion pulses, respectively; one for homonuclear decoupling, and one for compensation of Bloch–Siegert phase shifts. The 4D APSY-CH268-CH NOESY experiments was measured on *K10* as 57 2D projection spectra within 45 h, and the 4D APSY-CH1′-CH NOESY experiment as 59 projections within 66 h, both of them on the 900 MHz spectrometer. All NOESY experiments were measured with a mixing time of 150 ms.

Through-bond experiments

The intra-nucleotide correlations were derived with a set of through-bond experiments (Fig. 1). The intra-ribose correlations were established with a 4D APSY-HCCH COSY experiment using the pulse sequence of the GFT 4D HCCH COSY experiment published recently (Atreya et al. 2012). 65 2D projection spectra were measured within 61.5 h on the 900 MHz spectrometer. For the ribose-to-base connection, a set with two 3D APSY-HCN experiments (Krähenbühl et al. 2012) that correlate H6/8-C6/8-N1/9 and H1′-C1′-N1/9, respectively, was measured at 700 MHz. The ribose-to-base 3D APSY-HCN experiment was measured with 60 projections within 39 h. The intra-base 3D APSY-HCN experiment was measured in its TROSY version (Krähenbühl et al. 2012; Pervushin et al. 1997) with 60 projections within 26 h; the ^{13}C dimension was subsequently calibrated to optimally match the CH6/CH8 peaks in the CH68CH NOESY experiment in order to compensate for the shift of the resonances in the TROSY version. For intra-base correlations in pyrimidines, a 4D APSY-HCCH TOCSY pulse sequence that includes selective pulses to establish the $^{1}H6$–$^{13}C6$–$^{13}C5$–$^{1}H5$ magnetization transfer pathway was developed. This 4D APSY-HC6CH5 TOCSY experiment was measured on *K10* with 41 projections within 18.5 h on the 600 MHz spectrometer. The pulse sequence for this pyrimidine-selective 4D APSY-HCCH TOCSY experiment is provided in the Supplementary Material (Fig. S3). Experimental details for the through-bond experiments are provided in the Supplementary Material.

Processing and analysis

The APSY datasets were serially processed with the macro 'manageapsy' that is part of the software Topspin, which controls Bruker spectrometers. From the resulting 2D

projection spectra the full high-dimensional peak lists were derived with the elaborate algorithm GAPRO (geometric analysis of projections) (Hiller et al. 2005). Peak picking and the GAPRO analysis were also steered by 'manageapsy'. Each of these steps took only a couple of minutes of computational time on a standard desktop computer, summing up to ~1 h of processing and analysis time for all experiments. The subsequent automated assignment procedure was conducted with the algorithm FLYA (implemented in the software CYANA), as described in the following paragraph. Further analysis steps that were required to quantify, evaluate and verify the results of the APSY experiments and the automated assignment were performed with scripts written for MATLAB (R2010b, The MathWorks, Natick, WA, USA). This included a program that allows plotting and interleaving the APSY peak lists of different experiments to ensure their quality and precision. Furthermore, a program was written that partially eliminates impurity peaks from the APSY peak lists. These impurity signals (visible in Fig. 2) were identified via their increasing intensity in a set of ^{1}H-^{13}C HSQC spectra that were recorded between the APSY experiments at different times with identical parameters. However, this intermediate filtering step was not required for the final automated assignment since the FLYA algorithm eliminated the corresponding spin systems. Verification of the final sequence-specific assignment was also performed with a MATLAB program, as is discussed below. All MATLAB programs are available from the authors.

Automated resonance assignment

For the automated sequence-specific assignment, the algorithm FLYA (Schmidt and Güntert 2012; Lopez-Mendez and Güntert 2006) was used that is part of the program package CYANA (Güntert 2004; Güntert et al. 1997) (©Peter Güntert). Recently, the application of FLYA for resonance assignments of unlabeled stem-loop RNAs with a size of 21–42 nucleotides has been published (Aeschbacher et al. 2013): the authors used sets with a ^{1}H–^{1}H TOCSY, a ^{1}H–^{1}H NOESY, and a natural abundance ^{1}H–^{13}C HSQC spectrum, and achieved automated or assisted resonance assignment of RNA by mapping their chemical shifts to very narrow statistical values that consider the base type of the two preceding and the following residue. This statistics was based on a resonance analysis restricted to the ^{13}C–^{1}H moieties (CH) at the 2/5/6/8/1′ positions of the central nucleotide of Watson–Crick base-paired triplets as well as Watson–Crick base-paired termini; only poor statistics could be obtained for C2 in adenosines. In *K10,* out of 48 nucleotides only 16 fulfill the criteria for this narrow statistics to be applicable; they are colored in cyan in Fig. 3. Another 6 adenosines benefit from the detailed statistics for CH 6/8/1′ and H2, but not for C2; they are colored half in cyan in Fig. 3. For two further adenosines (in cyan-framed boxes in Fig. 3) that precede irregularities only the base type is considered; consequentially, only 2 out of 18 adenine C2 resonances benefit from precise chemical shift statistics. The other 24 nucleotides of *K10* were mapped to the general statistics, which does not consider neighboring nucleotides, and is thus considerably less specific.

The input for FLYA is prepared using the C++ program Chess2FLYA (Chemical shift statistics to FLYA) (Aeschbacher et al. 2013) which creates the file with the chemical shift statistics. The high-dimensional peak lists derived from the APSY experiments could be used with FLYA/CYANA without any modification of the algorithm. The current CYANA version provides the option to define any type of experiment with the corresponding magnetization transfer pathways and their respective probabilities, along with any selectivity; currently up to 8-dimensional experiments can be handled by CYANA. The pathway definitions for the APSY RNA experiments are provided in the Supplementary Material. The sequence-specific resonance assignment relies on the sequential information that is contained in through-space APSY experiments. An evaluation of this data requires a prediction of NOESY peak lists based on a predicted tertiary structure. FLYA offers an option for tertiary structure prediction, but we used RNAComposer (Popenda et al. 2012), which consistently produced significantly higher completeness of the assignments. For the automated assignment of *K10* resonances with APSY peak lists, matching thresholds of 0.02 ppm for ^{1}H, 0.15 ppm for ^{13}C, and 0.05 ppm for ^{15}N were set for FLYA. The FLYA output contains a proposed assignment for all RNA resonances in CH-groups: the most important criterion to evaluate their quality is the percentage of chemical shift values from 50 independent runs of FLYA that agree (within the thresholds) with the consensus value. If this value lies above 80 %, the assignment is considered very reliable by FLYA; for the case of *K10,* most assignments were already correct when this quality value was above 50 %. The duration of the FLYA calculations was approximately 10–40 min, depending on the exact input data and parameters; the preparation steps with Chess2FLYA and the RNAComposer took a few seconds each.

Conventional K10 assignment

The output of the automated resonance assignment with APSY was analyzed and verified by comparing it to the assignment that had been performed with conventional experiments assisted by spectral simplification through site-specific deuteration of pyrimidine bases (Bullock et al. 2010).

The chemical shifts of the two assignments were compared with each other with a MATLAB program which plotted the results for visual verification of the matching. The APSY-based FLYA assignments for individual ^{13}C–^{1}H-groups were considered correct when their ^{1}H and ^{13}C chemical shifts laid within a threshold ellipse in the 2D ^{1}H–^{13}C space with axes lengths 0.4 ppm and 0.12 ppm for the ^{13}C and ^{1}H dimensions, respectively, compared to the manual assignment. The seemingly high threshold value for ^{1}H was required due to slightly different sample conditions. However, a two-dimensional elliptical threshold is more stringent than using the same values as simple cutoffs.

Results

The automated assignment of the NMR resonances of the uniformly [^{13}C, ^{15}N]-labeled RNA *K10* with 48 nucleotides could be accomplished with a completeness of more than 80 % for the ^{13}C–^{1}H moieties at the 5, 6, 8, 1′ and 2′ positions (CH5/6/8/1′/2′) in the nucleotides. The basis for this achievement were high precision peak lists that resulted from high-dimensional APSY experiments developed to assign RNA: two selective 4D APSY-CHCH NOESY starting on CH2/6/8 or CH1′, respectively, a 4D APSY-HCCH COSY for ribose resonances, and a 4D APSY-HCCH TOCSY correlating CH6 with CH5 in pyrimidines. These experiments were complemented with a set of two published 3D APSY-HCN experiments (Krähenbühl et al. 2012) for sugar-to-base correlation or intra-base correlation, respectively. The measurements and the parameters are described in the "Materials and methods" section, as well as in the Supplementary Material.

The GAPRO analysis of the CH2/6/8-selective 4D APSY-CHCH NOESY experiment with *K10* produced a 4D CH268CH peak list with 454 correlation peaks. The CH1′-selective 4D APSY-CHCH NOESY experiment was processed and analyzed separately for the signals with a transfer that ended on base resonances, and the signals that were detected on the ribose protons. As a result, a 4D CH1′CH268 peak list with 119 peaks and a CH1′CH ribose peak list with 47 peaks were obtained. The pyrimidine-selective 4D APSY-HCCH TOCSY experiment led to a CH6CH5 peak list with 42 peaks. The 3D APSY-HCN experiments resulted in peak lists with 147 H6-C6-N1 or H8-C8-N9 peaks, respectively, for the intra-base experiment, and with 99 H1′-C1′-N1/9 peaks for the ribose-to-base experiment. All peak lists contained significantly more resonances than expected for *K10* due to the heterogeneity of the sample. The additional correlations correspond to real peaks in the spectra stemming from progressing hydrolysis in the hairpin loop and single-nucleotide bulge of *K10*. The heterogeneity of the sample and the consequent implications for the automated assignment strategy with APSY experiments are commented in the "Discussion".

The peak lists obtained from the six APSY experiments for RNA were taken as input for the automated resonance assignment calculations with the algorithm FLYA as discussed in "Materials and methods" and the Supplementary Material. The resulting resonance assignment was compared to the existing manual assignment and documented in Fig. 3 (the chemical shift lists can be obtained from the authors). In Fig. 3 all individual CH moieties at the atom positions 2, 5, 6, 8, 1′ and 2′ of the 48 nucleotides are represented by colored rectangles. The green or lime green rectangles indicate coinciding resonance assignments from the manual and the automated method, and the red or orange ones differing assignments. The CH groups denoted with lime green or orange rectangles have quality values below 50 % for the ^{1}H and/or the ^{13}C resonance (see "Materials and methods"). For the red and orange assignments, it is assumed that the manual assignment is correct since it is based on different samples and measurements at different temperatures and last but not least a high resolution structure was calculated based on the manual assignment. When considering only identical assignments correct, the following assignment completeness was obtained for the APSY assignment: 92 % (23 out of 25) for CH5 and CH6 in pyrimidines, 87 % (20 out of 23) for CH8 in purines, and 71 % (34 out of 48) for CH1′/2′. The assignment of the CH2 chemical shifts in adenosine was correct for 11 out of 18 (61 %), since for these resonances no intra-nucleotide correlations were available in the suite of APSY experiments; CH2 group assignment was thus solely NOESY-based. In addition, FLYA did not produce a well-defined C2 chemical shift range for 16 out of 18 adenines. Details of these assignments are discussed in the "Automated sequential resonance assignment" section of the following "Discussion and conclusions".

Discussion and conclusions

APSY experiments for RNA

The strategy for automated sequence-specific assignments of resonances of non-labile nuclei in RNA is based on partially novel APSY experiments. The resulting precise high-dimensional peak lists served as input for the software FLYA which connected the spin systems to the final sequence-specific assignment (Fig. S1 of the Supplementary Material shows a flow chart of the strategy). For large RNAs, as in our example the 48-nucleotide *K10*, neighboring nucleotides cannot be connected via the intervening

^{31}P nucleus in the phosphodiester group of the oligonucleotide backbone (see "Materials and methods"), and thus NOESY-type experiments have to be used.

APSY NOESY experiments

Conventional 4D CHCH NOESY experiments for RNA have been of limited use due to insufficient resolution. When applied to *K10,* also APSY versions of these experiments suffered from overlap of signals. Since several correlation peaks share common ^{13}C–^{1}H groups, many projections could not separate them appropriately. The 4D APSY NOESY experiments were thus designed to be selective for the starting ^{13}C–^{1}H moiety. This leads to a significant reduction of overlap for two reasons: first, the overall signal density is reduced significantly, and second, the ^{13}C neighbors can be decoupled selectively, thus allowing longer evolution times resulting in higher resolution. The ^{13}C–^{1}H groups CH1′ in the sugar moiety or in the bases CH2/6/8 (in adenines/pyrimidines/purines), respectively, were chosen as start nuclei. This procedure is favorable since their carbon resonances can be selectively excited, they show a wider chemical shift dispersion compared to other CH groups, and their protons correlate to protons of the neighboring nucleotides, thus providing information to establish sequential connectivities.

Intra-ribose correlation experiments

The limited resolution of the ribose resonances is the biggest challenge for a complete assignment. A 4D APSY-HCCH TOCSY experiments revealed the same problem with the separation of shared ^{13}C–^{1}H groups as the non-selective NOESY experiments above. Starting selectively on CH1′ could not relieve the situation sufficiently, and the best result was achieved with a 4D APSY-HCCH COSY experiment that is based on a (4,3)D HCCH GFT experiment; the related (4,3)D HC(C)CH GFT experiment, when implemented as an APSY experiment, led to spectral crowding, since the pathway selection was not unique (Atreya et al. 2012). The 4D APSY-HCCH COSY provided more reliable CH1′-CH2′ correlations than a COSY or TOCSY experiment starting on CH1′ selectively, since it included both the 1′-2′ and the 2′-1′ magnetization transfer pathways. The experiment included also all other 4D HCCH ribose correlation pathways. However, many resonances could not uniquely be resolved with the limited ^{13}C constant-time evolution periods of 8 ms imposed by ^{13}C–^{13}C scalar couplings. In addition, signals from partially hydrolyzed and degraded RNA affect the ribose region of the spectra most: some degradation products in RNA samples are not uncommon (not only for our *K10* sample). We hence excluded the CH3′/4′/5′/5″ resonances from our assignment strategy. The H4′/5′/5″ resonances do usually not provide information for the sequential connection via through-space experiments. This is different for H3′, which is usually close (~ 3.5 Å) to sequential H6/8 protons; if the sample condition and signal resolution allows their inclusion for an RNA under investigation, it is recommended to do so. However, the sequential information is also available via the correlations of H2′, and partially H1′, to the sequential base protons, or also via base–base correlations.

Ribose-to-base correlation experiments

The 5D APSY-HCNCH experiment proved not sufficiently sensitive for an application with *K10*: only about half the expected signals were observed. The combination of two 3D APSY-HCN (Krähenbühl et al. 2012) experiments proved more successful. The efficiency of HCN experiments depends on base type and conformation as discussed e.g. in (Fiala et al. 1998, 2000). Not all correlations were contained for various reasons: (1) three dimensions could not resolve all resonances in an RNA of this size; (2) the intra-base 3D APSY-HCN experiment requires TROSY to reduce relaxation; however, its efficiency can differ for individual bases; (3) the 3D APSY-HCN experiments were measured last, and the sample contained at this stage only about 65 % of the initial *K10* concentration, thus 0.23 mM instead of the initial 0.35 mM, determined by changes in peak integrals of separated peaks in [^{1}H, ^{13}C]-HSQC spectra. The accompanying degradation products further aggravated peak overlap. Nevertheless, the two 3D APSY-HCN experiments provided precious ribose-to-base correlation information, which was complemented by the intra-nucleotide ribose-to-base NOE correlations from the through-space APSY experiments.

Intra-base experiments

The remaining non-exchangeable protons in the bases are H5 for pyrimidines and H2 for adenines. In pyrimidines, CH6 and CH5 could be connected with a fully selective 4D APSY-HCCH TOCSY experiment (H6-C6-C5-H5), which delivers robust results. For CH2 groups in adenines, through-bond experiments that establish correlations with CH8 could not be used for an RNA of the size of *K10*: a direct correlation experiment (Marino et al. 1994) was not sensitive enough, and a TROSY-relayed set that connects via C4/C5 (Simon et al. 2001) did not provide the required precision and resolution, neither in the conventional nor in the APSY version. Thus CH2 resonances are only assigned based on through-space correlations, which can be ambiguous. E.g. the CH2 resonances of A32–A34 have NOESY correlation peaks to the CH1′ resonances of the diagonal residues U14–U16; but they are systematically assigned to A31-A33, since the same correlations could also appear for base pairs (Fig. 3).

The *K10* RNA sample

The assignment strategy with high-dimensional APSY experiments was successfully applied with the stem-loop RNA *K10* containing 48 nucleotides. Such an RNA is considered large for detailed studies by liquid-state NMR. In addition, the rather low average concentration of 0.3 mM and the heterogeneity (see Fig. 2) made this RNA a challenging, real-life application. The successful automated sequence-specific assignment was therefore not only a proof of concept, but a case that required the method to be robust with respect to challenges caused by the sample. The increasing heterogeneity during the experiments required the automated assignment strategy to distinguish the peaks of *K10* from the additional peaks. At the end of the measurements, the latter constituted a substantial part of the spectra, both in number and intensity. APSY constitutes a powerful filter against spectral artifacts, but can naturally not distinguish impurity peaks of degradation products from the desired ones, since they have the same spectroscopic features. Further, the appearance of new peaks, and the changes in intensity and chemical shift of existing peaks can deteriorate the inherent precision of APSY peak lists. Despite all these obstacles, our automated assignment strategy proved successful in the application with *K10* (Fig. 3). The details of the resonance assignment are discussed in the next section.

The sample conditions are clearly unfortunate. However, the consequences are conservative in the sense that we get less assignments and more wrong assignments than with a perfect sample. Further, most larger biologically active RNAs will have more irregular secondary structure with larger internal and hairpin loops containing UpA and CpA dinucleotide sequences which tend to hydrolyze over time (Hosaka et al. 2000), thus giving rise to additional peaks. In addition, structural heterogeneity (equilibria of different folds) will also be an issue with larger systems again giving rise to additional subsets of resonances. And, we also envision future applications in cellular lysates or fractionated lysates where RNases can degrade the sample over time due to wearing out RNase inhibitors. Thus our sample is a robust test for such larger systems, where we can assist secondary structure determination (based on sequential assignments) and use quick APSY-derived assignments to map interactions with proteins and other biological ligands (Wüthrich et al. 1991).

Automated sequential resonance assignment

As basis for the discussion of our novel automated resonance assignment approach with APSY experiments, we need to briefly recapitulate an assignment method that was recently published for Watson–Crick base-paired RNA (Aeschbacher et al. 2013). This method partially assigns resonances of unlabeled RNA with up to 42 nucleotides with conventional 2D experiments, which is remarkable. The basis is a detailed chemical shift statistics (Aeschbacher et al. 2012) and the FLYA algorithm. However, this strategy is only reliable for the central nucleotides of Watson–Crick base-pair triplets in A-form RNA and termini. The low dimensionality of the spectra may further limit the general applicability for larger RNAs or RNA including repetitive elements. For K10, only a third of the nucleotides form Watson–Crick base-pairs which fulfill the conditions for the detailed chemical shift statistics. Further, *K10* consists mostly of A/U base pairs reducing the dispersion of resonances substantially.

In *K10* the five base pairs U14-17/A31-34 and U5/A44 (see Fig. 3) have very similar chemical shifts in all spectral regions. Even APSY through-bond correlation experiments could not completely resolve these resonances, but the through-space APSY experiments did. The two NOESY APSY experiments allowed assigning also these challenging regions correctly: 30 out of 35 CH5/6/8/1′/2′ groups were correctly assigned. The chemical shifts of the remaining 5 CH groups were also quite close to the correct positions. The correctness of the assignments was confirmed by checking the correlations via MATLAB plots, since FLYA did not recognize them as reliable due to their overlapping correlation peaks in the through-bond experiments.

The performance of an automated assignment strategy with *K10* based on 2D peak lists only was evaluated using ^{1}H–^{1}H peak lists obtained from reduced APSY peak lists—not considering HCN peak lists. In addition, a peak lists from a ^{1}H-^{13}C HSQC spectrum was added. Using these peak lists as input for the automated FLYA assignment resulted in a completeness of less than 30 %—even though the homonuclear proton peak lists were derived from precise 4D APSY peak lists, and also contained peaks that would not have been resolved in the corresponding 2D spectra. Thus, adequate assignment completeness for the *K10* sample required isotope labeling.

The APSY peak lists contained up to three times the number of peaks expected for the corresponding magnetization transfer pathways in *K10*. FLYA was able to distinguish these peaks for the CH5/6/8/1′/2′ resonances since the four-dimensional peaks of many impurity correlations included exceptional chemical shifts, such as, e.g., H2′ resonances above 5 ppm, which reduce their probability to belong to a well-structured RNA substantially. The reason for these exceptional chemical shifts is that the peaks arise from hydrolyzed degradation products of *K10*. Further, since the degradation proceeded during the measurements, a large part of the impurity peaks was not contained in the initial spectra, which worked like an inherent filter, since only peaks contained in most peak lists could reliably be

connected by FLYA. For the ribose signals of the 3′, 4′ and 5′ moieties this filtering did not work due to too severe overlap. That may have been the main reason why the 4D APSY-HCCH COSY experiment was not sufficient to unambiguously connect the ribose resonances with each other.

The assignment of *K10* resonances with conventional methods required extensive measurements (3D ^{13}C-edited NOESY spectra for ribose and aromatic region and numerous 2D NOESY experiments both in D_2O and H_2O and at different temperatures) and labour-intensive manual assignments. Furthermore, persisting overlap unresolvable in ^{13}C-edited 3D NOESY spectra required additional preparation of different *K10* RNA samples with site-specific deuteration of pyrimidine bases to complete sequential assignments with thus simplified 2D NOESY spectra (Bullock et al. 2010). In contrast, the APSY strategy worked with only one labeled sample (0.3 mM) in D_2O, took less than 11 days of spectrometer time for the measurements, and allowed to conduct the full processing, analysis and resonance assignment procedure in an automated fashion within a few hours. However, there is a price to pay: less complete assignments and some wrong assignments. The assignments for CH2/5/6/8/1′/2′ resonances, which usually display the best chemical shift dispersion in contrast to the remaining ribose CH groups, are essential to derive secondary structure information, to identify A-form and non A-form features of the structure and to aid interpretation of interaction site mappings. When complemented with RDCs this assignment is sufficient to validate the RNA fold as shown by Mollova et al. (2000). For a full high resolution structure calculation these assignments have to be supplemented by assignments of imino/amino protons and possibly others depending on the quality of the sample and the spectral properties.

In summary, we present a suite of through-space and through-bond high-dimensional APSY experiments that is able to reliably correlate the ^{13}C–^{1}H groups of large RNA. The resulting precise high-dimensional peak lists provide an ideal basis for an automated assignment of NMR resonances with the algorithm FLYA. The application with one 0.3 mM sample of the 48-nucleotide RNA *K10* led to an assignment of more than 90 % of the CH5/6/8 resonances, more than 70 % of the CH1′/2′ resonances, and 61 % of the CH2. The method for the automated assignment of RNA resonances proved to be very robust and applicable with a low-concentrated and heterogeneous sample of a large RNA with many structural irregularities. It proved particularly useful not only for the NMR-wise challenging, but biologically often highly relevant RNA loop region, but also for the biologically relevant helical regions in *K10* (Bullock et al. 2010), which contain significant signal overlap.

Acknowledgments We would like to gratefully acknowledge Prof. Peter Güntert and Elena Schmidt (both University Frankfurt) for support with the use of FLYA. We thank Prof. Frédéric Allain (ETH Zurich) for his manifold support. This work was supported by the Swiss National Science Foundation [Grant Numbers 120048, 140559].

References

Aeschbacher T, Schubert M, Allain FHT (2012) A procedure to validate and correct the C-13 chemical shift calibration of RNA datasets. J Biomol NMR 52:179–190

Aeschbacher T, Schmidt E, Blatter M, Maris C, Duss O, Allain FHT, Güntert P, Schubert M (2013) Automated and assisted RNA resonance assignment using NMR chemical shift statistics. Nucleic Acids Res 41:e178

Atreya HS, Sathyamoorthy B, Jaipuria G, Beaumont V, Varani G, Szyperski T (2012) GFT projection NMR for efficient H-1/C-13 sugar spin system identification in nucleic acids. J Biomol NMR 54:337–342

Billeter M, Orekhov V (eds) (2012) Novel sampling approaches in higher dimensional NMR. Topics in current chemistry, Springer, Berlin Heidelberg, 316:1–148

Bullock SL, Ringel I, Ish-Horowicz D, Lukavsky PJ (2010) A'-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. Nat Struct Mol Biol 17:703–709

Cléry A, Schubert M, Allain FHT (2012) NMR spectroscopy: an excellent tool to understand RNA and carbohydrate recognition by proteins. Chimia 66:741–746

Clore GM, Kay LE, Bax A, Gronenborn AM (1991) 4-Dimensional C-13/C-13-edited nuclear overhauser enahncement spectroscopy of a protein in solution—application to interleukin 1-beta. Biochemistry 30:12–18

Douglas JT, Latham MP, Armstrong GS, Bendiak B, Pardi A (2008) High-resolution pyrimidine- and ribose-specific 4D HCCH-COSY spectra of RNA using the filter diagonalization method. J Biomol NMR 41:209–219

Easton LE, Shibata Y, Lukavsky PJ (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. RNA Publ RNA Soc 16:647–653

Fiala R, Jiang F, Sklenar V (1998) Sensitivity optimized HCN and HCNCH experiments for C-13/N-15 labeled oligonucleotides. J Biomol NMR 12:373–383

Fiala R, Czernek J, Sklenar V (2000) Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. J Biomol NMR 16:291–302

Fiorito F, Hiller S, Wider G, Wüthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. J Biomol NMR 35:27–37

Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

Hiller S, Wider G (2012) Automated projection spectroscopy and its applications. Top Curr Chem 316:21–47

Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102:10876–10881

Hiller S, Wasmer C, Wider G, Wüthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR Spectroscopy. J Am Chem Soc 129:10823–10828

Hiller S, Joss R, Wider G (2008a) Automated NMR assignment of protein side chain resonances using automated projection spectroscopy (APSY). J Am Chem Soc 130:12073–12079

Hiller S, Wider G, Wüthrich K (2008b) APSY-NMR with proteins: practical aspects and backbone assignment. J Biomol NMR 42:179–195

Hosaka H, Hosono K, Kawai G, Takai K, Takaku H (2000) Self-cleavage of p2Sp1 RNA with Mg2 + and non-ionic detergent (Brij 58). J Inorg Biochem 82:215–219

IUPAC-IUB (1983) Abbreviations and symbols for the description of conformations of polynucleotide chains. Pure Appl Chem 55:1273–1280

Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. J Am Chem Soc 125:1385–1393

Krähenbühl B, Wider G (2012) Automated projection spectroscopy (APSY) for the assignment of NMR resonances of biological macromolecules. Chimia 66:770–774

Krähenbühl B, Hofmann D, Maris C, Wider G (2012) Sugar-to-base correlation in nucleic acids with a 5D APSY-HCNCH or two 3D APSY-HCN experiments. J Biomol NMR 52:141–150

Krähenbühl B, Boudet J, Wider G (2013) 4D experiments measured with APSY for automated backbone resonance assignments of large proteins. J Biomol NMR 56:149–154

Krähenbühl B, El Bakkali I, Schmidt E, Güntert P, Wider G (2014) Automated NMR resonance assignment strategy for RNA via the phosphodiester backbone based on high-dimensional through-bond APSY experiments. J Biomol NMR 59:87–93

Kupce E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. J Am Chem Soc 125:13958–13959

Lopez-Mendez B, Güntert P (2006) Automated protein structure determination from NMR spectra. J Am Chem Soc 128: 13112–13122

Marino JP, Prestegard JH, Crothers DM (1994) Correlation of adenine H2/H8 resonances in uniformly C-13 labeled RNAs by 2D HCCH-TOCSY—a new tool for H-1 assignment. J Am Chem Soc 116:2205–2206

Mollova ET, Hansen MR, Pardi A (2000) Global structure of RNA determined with residual dipolar couplings. J Am Chem Soc 122:11561–11562

Narayanan RL, Durr UHN, Bibow S, Biernat J, Mandelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein tau with 441-residues. J Am Chem Soc 132:11906–11907

Nikonowicz EP, Pardi A (1992) Application of 4-dimensional heteronuclear NMR to the structure determination of a uniformly C-13 labeled RNA. J Am Chem Soc 114:1082–1083

Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. J Biomol NMR 27:165–173

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T-2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371

Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW (2012) Automated 3D structure composition for large RNAs. Nucleic Acids Res 40:e112

Ranjan N, Damberger FF, Sutter M, Allain FHT, Weber-Ban E (2011) Solution structure and activation mechanism of ubiquitin-like small archaeal modifier proteins. J Mol Biol 405:1040–1055

Schmidt E, Güntert P (2012) A new algorithm for reliable and general NMR resonance assignment. J Am Chem Soc 134:12817–12829

Serano T, Cohen RS (1995) A small predicted stem-loop structure mediates oocyte localization of Drosophila K10 messenger-RNA. Development 121:3809–3818

Simon B, Zanier K, Sattler M (2001) A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly C-13-labeled RNA molecules. J Biomol NMR 20:173–176

Vuister GW, Clore GM, Gronenborn AM, Powers R, Garrett DS, Tschudin R, Bax A (1993) Increased resolution and improved spectral quality in 4-dimensional C-13/C-13-separated HMQC-NOESY-HMQC spectra using pulsed-field gradients. J Magn Reson, Ser B 101:210–213

Wider G, Dreier L (2006) Measuring protein concentrations by NMR spectroscopy. J Am Chem Soc 128:2571–2576

Wüthrich K, Spitzfaden C, Memmert K, Widmer H, Wider G (1991) Protein secondary structure determination by NMR—application with recombinant human cyclophilin. FEBS Lett 285:237–247