Supplementary Material

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

Barbara Krähenbühl[1] · Peter Lukavsky [2] · Gerhard Wider[1,*]

[1]*Institute of Molecular Biology and Biophysics, ETH Zurich*
8093 Zurich, Switzerland

[2]*CEITEC – Central European Institute of Technology, Masaryk University,*
62500 Brno, Czech Republic

*corresponding author:
e-mail: gsw@mol.biol.ethz.ch
Phone: +41 44 633 34 55
Fax: +41 44 633 14 84

Table of contents:

1. Flowchart of RNA assignment strategy

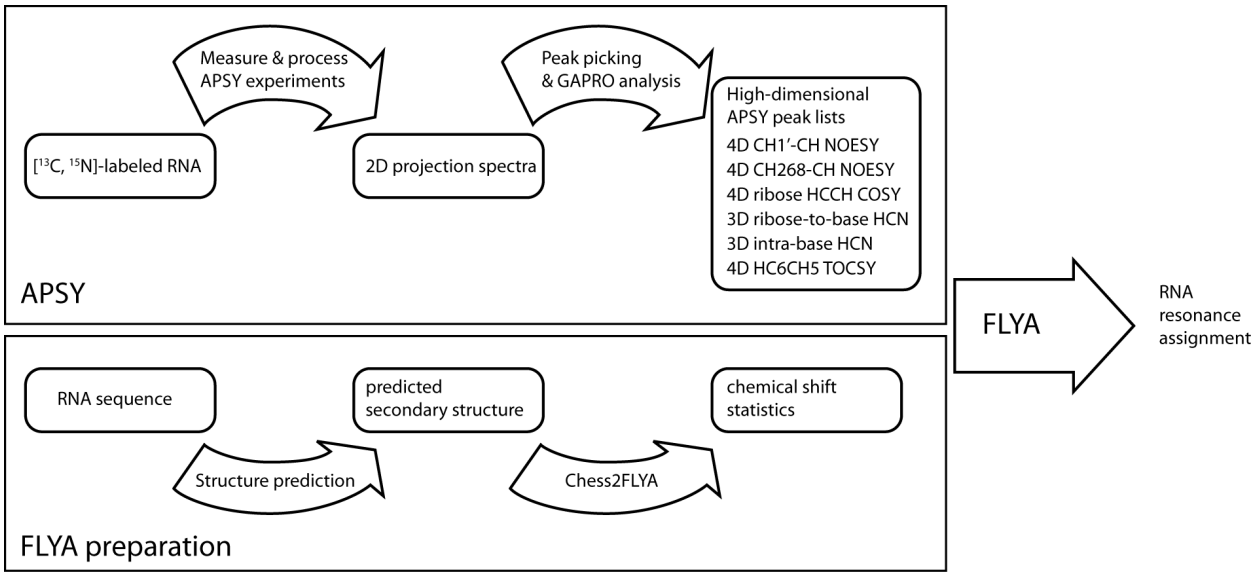

Figure S1: The strategy for automated NMR resonance assignment of RNA is based on APSY experiments and the algorithm FLYA for the analysis of the data. Using one sample of [^{13}C,^{15}N]-labeled RNA in D_2O solution six different APSY experiments are measured. The resulting peak lists are used as input for FLYA that calculates resonance assignments based on chemical shift statistics (1-3).

2. NMR spectroscopy - setup of APSY experiments

A very similar description of the setup of APSY experiments was provided in the Supplementary Material of the publication Krähenbühl et al. (4). We repeat this description for the convenience of the reader.

The APSY NMR experiments were performed on a 600, a 700 and a 900 MHz Bruker Avance III spectrometer with cryogenic probes. The 700 MHz and the 900 MHz spectrometers were equipped with triple-resonance probes (^{1}H, ^{13}C, ^{15}N), and the 600 MHz spectrometer with a quadruple-resonance probe (^{1}H, ^{13}C, ^{15}N, ^{31}P). They were operated with the software Topspin 3.1 (Bruker, Karlsruhe, Germany).

APSY experiments are started from a parent data set with the full dimensionality, which is set up like a conventional correlation experiment. All further steps are performed with the program *manageapsy*, which runs in Topspin (and is contained in its latest versions). It can first be used to create an angle set that is adjusted to the parameters in the parent data set, and it also automatically creates the 2D projection data sets in a different data directory. These projection spectra can either be run conventionally with the command ZG (as was done for the presented APSY measurements for manageability), or directly from *manageapsy* as a fully automated run. Both the processing and the GAPRO analysis are also operable from the user interface of *manageapsy*. The following explanations apply for the current version of *manageapsy*; the software is, however, currently updated, including a new manual. It is recommended to consult this manual for the updates.

Operating measurements of APSY experiments via the user interface of *manageapsy* does usually not require in-depth expertise of APSY. It is, however, advantageous to be familiar with some of the basic principles of APSY as described, e.g., in (5). We provide a short description of each setup and analysis step in the following, and also include some advice what to pay particular attention to. Topspin-specific names of parameters and functions are written in capital letters, and used without further explanations; they can be looked up in the Topspin manuals if required.

The specific parameters used for the setup, measurement and processing of the novel APSY experiment for K10 are explained later in this chapter. Furthermore is an alternative setup, processing and analysis approach briefly discussed at the end of this chapter. In the following chapters, each APSY experiment of the automated assignment approach for large RNA is presented with its pulse sequence (if not published before), and a set of tables with the corresponding parameters that were used for the measurement of the *K10* spectral data

Setup of APSY parent data set

The parameters for the setup of the *K10* measurements in the parent data set of the APSY experiments are provided in the ”Parameter” section for each experiment in the small table that includes the spectral width (SW), the maximal evolution time (AQ), the nucleus type (NUC1), and the carrier frequencies. The dimensions labeled ω_1-ω_4 in the parameter tables correspond to the frequency axes F1-F4 of a 4D experiment, which are ordered from right to left in Topspin (EDA). SW, AQ and NUC1 can directly be entered in the corresponding column; Topspin offsets for the same nucleus type need to be set in constants. The information which offset is defined in which CNST or O[1-4]P belongs to which dimension is contained in the header of the pulse programs.

An important parameter to set before *manageapsy* can be used with this file is USERA5, which has e.g. for the 4D APSY-CH268CH NOESY the form ’4D CHCH st disopt’ with the elements: [dimensionality] [nuclei types and order] [acquisition mode: st (States-TPPI) or ea (echo-antiecho)] [angle file and TD mode: ’resopt’ (resolution-optimized) or ’disopt’ (dispersion-optimized)]. USERA5, along with the experimental parameters in the parent data set and the angle file, fully define the setup of the 2D projection spectra. The large projection-spectra-specific parameter tables that are provided for the APSY experiments are thus just provided for the sake of completeness of information; but all these parameters are set up automatically by *manageapsy*.

Angle file creation

A set of angles (created in the parent directory in the file ’angles.dat’) is proposed based on the dimensionality of the APSY experiment, the spectral width SW, and the TD/AQ value in the indirect dimensions. There are two versions of angle sets: dispersion-optimized angles (’disopt’ as USERA5 parameter), and resolution-optimized angles (’resopt’ as USERA5 parameter). Their concepts are explained, e.g., in (6). Usually, ’resopt’ is is used if most evolution periods are of constant-time type and ’disopt’ for mostly non-constant-time evolution periods. Since with ’resopt’ the evolution times in all projections are always maximal for at least one indirect dimension, the AQ values (maximal evolution time) should be set carefully to values where still reasonable signal intensity is detected for non-constant-time periods. Please note that each non-orthogonal angle α doubles the number of projections for which an angle set stands, since there are automatically the +α and −α projections measured. E.g. a (0°, 38°) angle set results in two projections for (0°, +38°) and (0°, −38°).

The measurement of the projection spectra can be started with QUMULTI and ZG, or with the automated mode provided in *manageapsy*.

Creation of 2D projection spectra; review orthogonal projections

The 2D projection data sets are created from the parent data set by selecting the 'Setup APSY series' function in *manageapsy*. They are created in a separate data directory with the form "<parent data set directory> <parent data set number>", numbered from 1 sequentially. The first 2D projection spectra correspond to normal 2D spectra, since the orthogonal angle sets (only 0° and/or 90° angles) involve only one active indirect dimension per spectrum. E.g. for the 4D APSY-HC6CH5 TOCSY experiment, the first three spectra are a ^{13}C5-^{1}H5 (0°, 0°), a ^{1}H6-^{1}H5 (0°, 90°) and a ^{13}C6-^{1}H5 (90°, 0°) correlation spectrum. The first spectrum (here the ^{13}C5-^{1}H5 spectrum) contains the correlation peaks of a conventional ^{13}C-^{1}H HSQC spectrum (this is true for all presented APSY experiments), and can be compared to such a spectrum to ensure completeness; the number of transient (NS) or the inter-scan delay (D1) in the parent data set can be modified accordingly, and the setup of the full data set be repeated.

Note: Based on the number of scans used in the direct projections (with one type of nucleus in the indirect dimension), the number of scans is usually doubled (unless there is excessive sensitivity) for each additional projected frequency axis, in order to compensate for the sensitivity losses of √2. Further criteria for adjustments of the number of scans are the maximal evolution times in the indirect dimension, and the relaxation in the included evolution periods.

Processing of measured 2D spectra

The processing of the raw data to 2D projection spectra with a hypercomplex Fourier transformation is performed automatically, and is started via *manageapsy*. The processing parameters are defined in the Topspin processing interface same as for any other experiment; the initial parameters of each 2D data set was copied upon setup from the two leftmost columns (frequency channels) of the parent data set processing parameters. Later serial parameter manipulation could be performed conveniently with the functions QUMULTI or MULTICMD. The procedure is thus as follows:

- Process the first 2D spectrum with XFB and adjust the phase in the direct dimension (PHC0, PHC1), the number of data points/zero-filling (SI; at least 2xTD in both dimensions), and the selective transformation of the stripe in the spectrum with the relevant signals (STSI/STSR) in the direct dimension.

- Transfer the values to all other 2D spectra directories (with QUMULTI or MULTICMD).

- Start the serial processing in *manageapsy* to derive 2D projection spectra.

GAPRO analysis

The input parameters for the GAPRO algorithm are stored in the file parameter.gap, which is created in the parent data set. The definition of these parameters is described in the basic APSY publication (7). It is usually not required to use exactly these or other precisely optimized values, since the GAPRO analysis is robust with respect to a range of values (approximately 20-30 % higher or lower values): but it is recommended to still test a range of values for each parameter to gain experience about their influence on the resulting APSY peak lists.

Processing parameters for APSY projection spectra

For all 2D projection spectra that were measured in this work, 1024 complex data points (TD 2048) were measured in the acquisition dimension, multiplied with a 75°-shifted sine-square-bell function, and zero-filled to 2048 or 4096 complex data points (SI). A sine-bell window function was applied in the indirect dimensions, which contained a variable number of data points; the data was zero-filled to 512 complex data points (SI). The SI values in the projection spectra are usually set to 2xTD in the acquisition dimension, and to ≥2 times the maximal TD in the indirect dimensions of the parent data set. For this work, a SI of 1xTD was used instead for the acquisition dimension of all spectra that included acquisition on ribose protons, since scalar coupling constants with respect to other protons or to phosphorus can take values in the range of 8 Hz depending on the conformation of the sugar ring; the slightly lower resolution was chosen to avoid signal splitting.

Alternative setup and processing strategy

We provide an alternative setup and processing strategy that is fast and reliable: The projection spectra can be created from the parent data set with the program *manageapsy*, same as for the main strategy. Input files for serial processing with the software PROSA (8) are created automatically upon setup in the child dataset directory: a parent processing file ’all apsy.pro’ which steers the full processing, and a sub-file for each set of projections that belong to the same angle set, ’apsy.pro’. PROSA can be started from the shell command line (Linux or Windows) with ’all apsy.pro’ as input. The resulting spectra can directly be used for the GAPRO analysis: GAPRO (the software that is named after the algorithm) can also be started in the shell command line of any Linux or Windows terminal, and takes ’spectra.gap’ and ’parameter.gap’ as inputs for peak picking, the subsequent geometric analysis, and the back-projection of the high- dimensional APSY peak list to the 2D projection spectra. The spectra and the picked and back-projected peak lists can be evaluated with the software XEASY as implemented in CARA (wiki.cara.nmr.ch).

3. The selective 4D APSY-CHCH NOESY experiment

The selectivity of the 4D APSY-CHCH NOESY experiment refers to the first ^{13}C and ^{1}H dimensions: the 180° pulse in the first transfer period selectively inverts the desired carbon coherences (Fig. S2). This leads also to a selection in the proton dimension, since the applied coherence transfer pathway selects protons that are covalently bound to the selected carbons. The shaped off-resonance pulses in the first ^{13}C evolution period serve for homonuclear decoupling of other carbons. In the pulse sequence description (Fig. S2), the two versions that were used for K10 assignment are described: one that starts on the base ^{13}C-^{1}H groups (^{1}H2/6/8-^{13}C2/6/8), and one that starts on the ^{13}C1′-^{1}H1′ group in the ribose of each nucleotide (Fig. S2). However, this pulse sequence can be used with other selectivities. It is important to adjust the length of the selective pulses carefully for the desired selectivity in order to avoid phase distortions in the spectra. The second ^{13}C and ^{1}H dimensions cover all ^{13}C-^{1}H groups in the RNA. If desired, selectivity could also be implemented after the NOE mixing time by using the same pulsing elements as before. This is not included in this APSY pulse sequence for two reasons: the first is, that a spectral region that could selectively be addressed with shaped pulses is anyway by definition separated from the rest, so that there is no reduction of overlap. The second is, that selectivity in both dimensions would often require more experiments due to their lower information content. The only advantage for an APSY experiment with selectivity in all dimensions would be that homonuclear decoupling could also be applied in the second ^{13}C evolution. For RNA with even more overlapping peaks than *K10*, that would be an option for a further resolution improvement of this APSY NOESY experiment. The situation with respect to selectivity is different before the NOE mixing, as is discussed in the main text: it was required to reduce the density of correlation peaks ending on the same proton after the mixing period.

It should also be considered that the large ^{13}C frequency range could, same as for any conventional experiment, lead to phase distortions for the ^{13}C5′ and the ^{13}C2 signals at the extremes of the spectral range due to the excitation profile of high-power pulses, particularly when measured at high magnetic field.

3.1. Pulse sequence

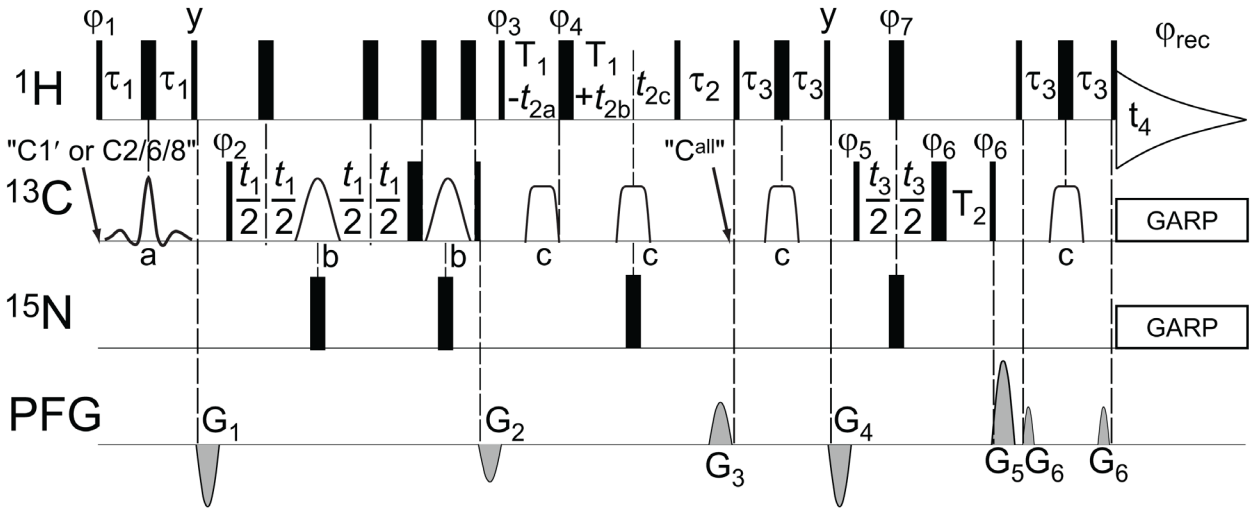

Figure S2: Pulse sequence of a 4D APSY-CHCH NOESY experiment for through-space correlations, with selectivity in the first $^{13}C/^{1}H$ dimensions. The experiment is proposed in two selectivity versions: either as a base version that selectively inverts $^{13}C2/6/8$ (in adenines/purines/pyrimidines) in the first ^{1}H-^{13}C transfer step, or as a ribose version that selectively inverts $^{13}C1'$ in this first transfer period. The carrier frequencies are set to 4.7 ppm for ^{1} H, and to 158 ppm for ^{15} N. All parameters apply for measurements at a field of 900 MHz. The ^{13}C carrier frequency is first set to 145 ppm on $^{13}C2/6/8$, or to 92.5 ppm on $^{13}C1'$, as is marked with an arrow that is labeled "C1′ or C2/6/8"; it is changed to the center of all ^{13}C frequencies of ^{13}C-^{1}H moieties at 115 ppm after the NOE mixing period, labeled "C^{all}". Thin and wide black bars represent nonselective 90° and 180° high-power pulses, respectively. The shaped pulses (a)-(c) are ^{13}C 180° pulses: (a) is an on-resonance Reburp (9) pulse with a duration of 667 µs for $^{13}C2/6/8$ selection, or 3000 µs for $^{13}C1'$ selection; (b) is an off-resonance Sinc (central lobe) pulse with a duration of 267 µs for $^{13}C5$ decoupling, or 800 µs for $^{13}C2'$ decoupling; and (c) is an adiabatic Chirp pulse with a duration of 500 µs. The time periods are $\tau_1 = T_1 = 1.25$ ms for the $^{1}H2/6/8$-$^{13}C2/6/8$ transfer, or 1.56 ms for the $^{1}H1'$-$^{13}C1'$ transfer; $\tau_2 = 150$ ms (NOE mixing period), and $\tau_3 = 1.38$ ms. GARP (10) decoupling sequences were applied during acquisition on ^{13}C with a field strength of 3.8 kHz, and on ^{15}N with 1.1 kHz. The phases of the pulses were set to x, unless indicated otherwise above the pulse symbol. The following phase cycle was applied: ϕ_1 = 4(x), 4(-x); ϕ_2 = 2(x), 2(-x); ϕ_3 = x, -x; ϕ_4 = y, -y; ϕ_5 = x; ϕ_6 = 8(x), 8(-x); ϕ_7 = 16(x), 16(-x); receiver phase ϕ_{rec} = a, 2(-a), a with a = x, 2(-x), x. Quadrature detection in the indirect dimensions was achieved by States-TPPI (11): ϕ_2 was incremented for the evolution period t_1 ($^{1}H2/6/8$ or $^{1}H1'$), ϕ_3 and ϕ_4 for t_2 ($^{13}C2/6/8$ or $^{13}C1'$), and ϕ_5 for t_3 (^{13} C). t_{2a}, t_{2b} and t_{2c} are incremented in a semi-constant time manner (12). The gray sinebell shaped gradient pulses (PFG) were applied along the z-axis with a duration of 1000 µs for G_1-G_5 , and 500 µs for G_6. Their strengths are: G_1: -22 G/cm; G_2: -15 G/cm; G_3: 5.5 G/cm; G_4: -22 G/cm; G_5:44 G/cm; G_6: 3 G/cm.

3.2. The 4D APSY-CH268CH NOESY parameters for *K10*

Sample	*K10* (stem-loop RNA with 48 nucleotides)
Spectrometer	Bruker Avance III 900 MHz with cryogenic probe
Temperature	25°C
Total experiment time	45h
Nr. of projections	57
Interscan delay	1.5 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 10$, $R_{min} = 20$ Hz, $\Delta\nu = 9.0$ Hz, S/N = 3.2

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{13}C2/6/8	21.0	4749	145.0	24.0
ω_2	^{1}H2/6/8	8.4	7554	4.7	15.0
ω_3	^{13}C	111.0	25102	115.0	6.0
ω_4	^{1}H	20.0	18029	4.7	56.8

Dispersion-optimized angles, evolution times not maximized.

α [°]	β [°]	Time [min]	# of complex points in ind. dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	35	150	6.0	4	^{13}C
0	90	18	80	16.8	4	^{13}C2/6/8
90	0	26	112	14.8	4	^{1}H2/6/8
±73.3	0	53	116	10.9,3.3	8	^{1}H2/6/8,^{13}C
0	±79.3	37	82	12.2,2.3	8	^{13}C2/6/8,^{13}C
90	±57.8	41	90	13.4,8.4	8	^{13}C2/6/8,^{1}H2/6/8
±73.3	±56.7	42	92	11.2,7.0,2.1	8	^{13}C2/6/8,^{1}H2/6/8,^{13}C
±59	0	56	124	7.4,4.4	8	^{1}H2/6/8,^{13}C
0	±69.3	42	92	8.7,3.3	8	^{13}C2/6/8,^{13}C
90	±38.5	46	100	9.4,11.8	8	^{13}C2/6/8,^{1}H2/6/8
±81.4	0	52	114	13.5,2.0	8	^{1}H2/6/8,^{13}C
0	±84.6	37	80	15.1,1.4	8	^{13}C2/6/8,^{13}C
90	±72.5	38	84	15.8,5.0	8	^{13}C2/6/8,^{1}H2/6/8
±81.4	±72.4	38	84	15.4,4.8,0.7	8	^{13}C2/6/8,^{1}H2/6/8,^{13}C
±59	±69.9	39	86	12.1,3.8,2.3	8	^{13}C2/6/8,^{1}H2/6/8,^{13}C
±81.4	±38.2	47	102	8.8,11.0,1.7	8	^{13}C2/6/8,^{1}H2/6/8,^{13}C
±59	±34.3	51	112	5.2,6.5,3.9	8	^{13}C2/6/8,^{1}H2/6/8,^{13}C
±39.7	0	62	136	4.4,5.3	8	^{1}H2/6/8,^{13}C
0	±52.9	50	110	5.6,4.3	8	^{13}C2/6/8,^{13}C
90	±21.7	49	108	5.5,13.9	8	^{13}C2/6/8,^{1}H2/6/8
±68.1	0	54	118	9.4,3.8	8	^{1}H2/6/8,^{13}C
0	±75.8	39	86	10.8,2.7	8	^{13}C2/6/8,^{13}C
90	±50	43	94	11.9,10.0	8	^{13}C2/6/8,^{1}H2/6/8
±22.6	0	65	144	2.4,5.7	8	^{1}H2/6/8,^{13}C
0	±33.5	60	132	3.5,5.2	8	^{13}C2/6/8,^{13}C
90	±11.2	51	112	2.9,14.7	8	^{13}C2/6/8,^{1}H2/6/8

3.3. The 4D APSY-CH1′CH NOESY

3.4 Parameters for K10

Sample	*K10* (stem-loop RNA with 48 nucleotides)
Spectrometer	Bruker Avance III 900 MHz with cryogenic probe
Temperature	25°C
Total experiment time	66h
Nr. of projections	59
Interscan delay	1.3 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 10$, $R_{min} = 20$ Hz, $\Delta\nu = 9.0$ Hz, S/N = 3.2

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{13}C1′	7.0	1583	92.5	24.0
ω_2	^{1}H1′	8.4	7554	4.7	15.0
ω_3	^{13}C	111.0	25102	115.0	6.0
ω_4	^{1}H	20.0	18029	4.7	56.8

Dispersion-optimized angles, evolution times not maximized.

α [°]	β [°]	Time [min]	# of complex points in ind. dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	65	150	6.0	8	^{13}C
0	90	17	38	24.0	8	^{13}C1′
90	0	48	112	14.8	8	^{1}H1′
±73.3	0	99	116	10.9,3.3	16	^{1}H1′,^{13}C
0	±86.4	33	116	51.9,3.3	16	^{13}C1′,^{13}C
90	±78.2	36	38	17.0,3.6	16	^{13}C1′,^{1}H1′
±73.3	±77.7	75	42	15.4,3.2,1.0	32	^{13}C1′,^{1}H1′,^{13}C
±59	0	106	44	2.6,1.6	16	^{1}H1′,^{13}C
0	±82.8	36	124	35.0,4.4	16	^{13}C1′,^{13}C
90	±67.3	48	42	11.9,5.0	16	^{13}C1′,^{1}H1′
±81.4	0	97	56	6.6,1.0	16	^{1}H1′,^{13}C
0	±88.2	33	114	64.5,2.0	16	^{13}C1′,^{13}C
90	±84	33	38	21.5,2.3	16	^{13}C1′,^{1}H1′
±81.4	±84	65	38	21.0,2.2,0.3	32	^{13}C1′,^{1}H1′,^{13}C
±59	±83	69	38	16.0,1.7,1.0	32	^{13}C1′,^{1}H1′,^{13}C
±81.4	±67	72	40	10.3,4.3,0.7	24	^{13}C1′,^{1}H1′,^{13}C
±59	±63.9	82	56	7.7,3.2,1.9	24	^{13}C1′,^{1}H1′,^{13}C
±39.7	0	116	64	2.1,2.5	16	^{1}H1′,^{13}C
0	±75.8	45	136	20.8,5.3	16	^{13}C1′,^{13}C
90	±50	67	52	8.0,6.7	16	^{13}C1′,^{1}H1′
±68.1	0	101	118	9.4,3.8	16	^{1}H1′,^{13}C
0	±85.2	33	38	14.4,1.2	16	^{13}C1′,^{13}C
90	±74.4	40	46	17.4,4.9	16	^{13}C1′,^{1}H1′
±22.6	0	123	144	2.4,5.7	16	^{1}H1′,^{13}C
0	±63.2	63	74	5.8,2.9	16	^{13}C1′,^{13}C
90	±30.8	84	98	7.7,12.9	16	^{13}C1′,^{1}H1′

4. The fully selective 4D APSY-HCCH TOCSY experiment

^{1}H6-^{13}C6-^{13}C5-^{1}H5

Intra-base experiment for pyrimidines in RNA.

4.1 Pulse sequence

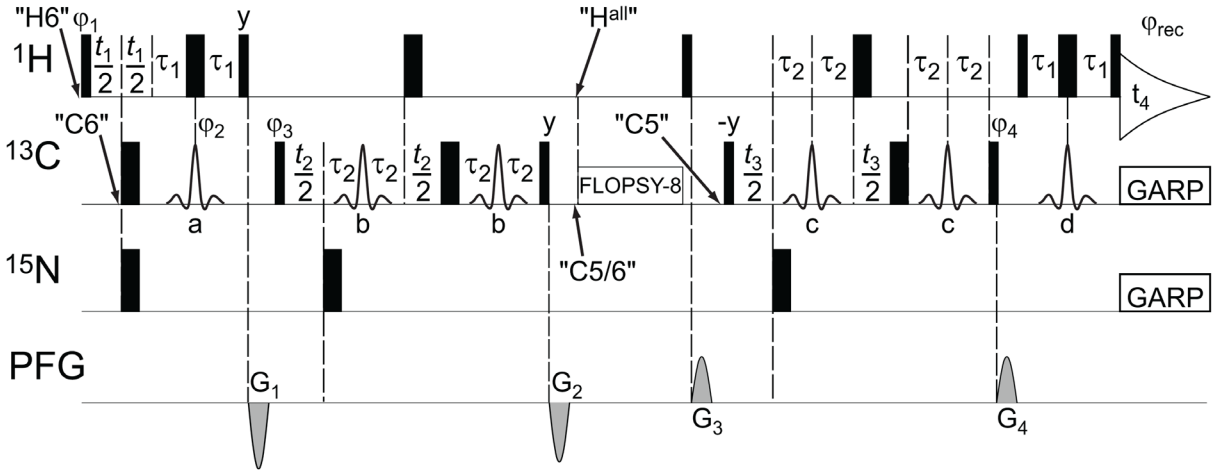

Figure S3: Pulse sequence of the 4D APSY-HCCH TOCSY experiment with selectivity in both ^{13}C dimensions. The offsets and the length of the selective pulses in this description are adjusted for the coherence transfer pathway ^{1}H6-^{13}C6-^{13} C5-^{1}H5 in pyrimidines. The carrier frequency of the ^{1}H channel is set to 7.6 ppm on ^{1}H6 in the beginning (labeled with an arrow and "H6"), and to 4.7 ppm before the TOCSY transfer ("H^{all} "); the ^{13} C carrier frequency is first set to 145 ppm on ^{13}C6 ("C6"), before the TOCSY mixing period to 122 ppm between ^{13}C5 and ^{13}C6 ("C5/6"), and after the TOCSY mixing period to 101 ppm on ^{13}C5 ("C5"); the ^{15}N carrier frequency is set to 160 ppm. All parameters apply for measurements at a field of 600 MHz. Thin and wide black bars represent non-selective 90° and 180° high-power pulses, respectively. The shaped pulses (a)-(d) are ^{13}C Reburp (9)180° pulses: (a) is an on-resonance pulse on ^{13}C6 with a duration of 1300 µs; (b) is an off-resonance pulse on ^{13}C5 at 101 ppm with a duration of 1300 µs; (c) is an off-resonance pulse on ^{13}C6 and ^{13}C4 at 154 ppm with a duration of 800 µs; (d) is an on-resonance pulse on ^{13}C5 with a duration of 1300 µs. ^{13}C TOCSY mixing was performed during 5.6 ms with two FLOPSY-8 (13) cycles with a field strength of 8.33 kHz. GARP (10) decoupling sequences were applied during acquisition on ^{13}C with a field strength of 2.5 kHz, and on ^{15}N with 1.56 kHz. The time periods are set to $1/(4JCH) = \tau_1 = \tau_2 = 1.5$ ms. The phases of the pulses were set to x, unless indicated otherwise above the pulse symbol. The following phase cycle was applied: ϕ_1 = 2(x), 2(-x); ϕ_2 = 4(x), 4(-x); ϕ_3 = x, -x; ϕ_4 = -x; receiver phase ϕ_{rec} = x, 2(-x), x. Quadrature detection in the indirect dimensions was achieved by States-TPPI (11): ϕ_1 was incremented for the evolution period t_1 (^{1}H6), ϕ_3 for t_2 (^{13}C6), and ϕ_4 for t_3 (^{13}C5). The gray sinebell shaped gradient pulses (PFG) were applied along the z-axis. The lengths and strengths are: G_1: 800 µs, 30 G/cm; G_2: 1000 µs, -33 G/cm; G_3: 1000 µs, -26 G/cm; G_4: 1000 µs, 44 G/cm.

4.2 Parameters for K10

Sample	*K10* (stem-loop RNA with 48 nucleotides)
Spectrometer	Bruker Avance III 600 MHz with cryogenic probe
Temperature	25°C
Total experiment time	18h 22min
Nr. of projections	41
Interscan delay	1.0 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 10$, $R_{min} = 20$ Hz, $\Delta\nu = 9.0$ Hz, S/N = 3.2

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{1}H6	1.2	720	7.6	29.8
ω_2	^{13}C6	7.0	1056	142.5	28.4
ω_3	^{13}C5	10.0	1509	101.0	23.7
ω_4	^{1}H5	15.0	9014	4.7	56.8

Dispersion-optimized angles, evolution times not maximized.

α [°]	β [°]	Time [min]	# of complex points in ind. dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	13	36	23.7	8	^{13}C5
0	90	7	20	27.8	8	^{1}H6
90	0	10	30	28.4	8	^{13}C6
±55	0	23	36	24.1,16.9	16	^{13}C6,^{13}C5
0	±64.5	17	26	25.5,12.2	16	^{1}H6,^{13}C5
90	±55.7	16	24	23.6,16.1	16	^{1}H6,^{13}C6
±55	±50.2	36	28	22.4,15.3,10.7	32	^{1}H6,^{13}C6,^{13}C5
±35.5	0	26	40	16.9,23.7	16	^{13}C6,^{13}C5
0	±46.3	22	34	21.1,20.2	16	^{1}H6,^{13}C5
90	±36.3	18	28	17.4,23.7	16	^{1}H6,^{13}C6
±70.7	0	21	32	27.1,9.5	16	^{13}C6,^{13}C5
0	±76.6	14	22	27.3,6.5	16	^{1}H6,^{13}C5
90	±71.2	14	22	27.3,9.3	16	^{1}H6,^{13}C6
±70.7	±70.1	28	22	26.6,9.1,3.2	32	^{1}H6,^{13}C6,^{13}C5
±35.5	±59.6	33	26	24.1,8.2,11.5	32	^{1}H6,^{13}C6,^{13}C5
±70.7	±34.7	36	28	15.9,21.6,7.6	32	^{1}H6,^{13}C6,^{13}C5
±35.5	±23.1	48	38	11.5,15.7,22.0	32	^{1}H6,^{13}C6,^{13}C5

5. The 3D MQ APSY-s-NCH

The pulse sequence of this experiment is published (14).

5.1 Parameters for K10

Sample	*K10* (stem-loop RNA with 48 nucleotides)
Spectrometer	Bruker Avance III 700 MHz with cryogenic probe
Temperature	25°C
Total experiment time	39h
Nr. of projections	60
Interscan delay	0.8 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 11$, $R_{min} = 7$ Hz, $\Delta\nu = 2.1$ Hz, S/N = 3.6

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{15}N1/9	33.0	2341	164.0	35.0
ω_2	^{13}C1′	11.0	1937	89.0	22.9
ω_3	^{1}H1′	12.1	8503	4.7	120.4

Resolution-optimized angles, evolution times maximized.

α [°]	Time [min]	# of complex points in ind. dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	23	44	22.7	8	^{15}N1/9
90	44	82	35.0	8	^{13}C1′
±71.8	45	84	34.6,11.4	16	^{13}C1′,^{15}N1/9
±80.7	44	82	34.7,5.7	16	^{13}C1′,^{15}N1/9
±56.7	49	92	34.5,22.7	16	^{13}C1′,^{15}N1/9
±85.3	44	82	34.9,2.9	16	^{13}C1′,^{15}N1/9
±37.2	32	60	17.3,22.8	16	^{13}C1′,^{15}N1/9
±87.6	44	82	35.0,1.5	16	^{13}C1′,^{15}N1/9
±20.8	25	48	8.6,22.5	16	^{13}C1′,^{15}N1/9
±77.6	44	82	34.5,7.6	16	^{13}C1′,^{15}N1/9
±63.7	47	88	34.8,17.2	16	^{13}C1′,^{15}N1/9
±12	23	44	4.7,22.0	16	^{13}C1′,^{15}N1/9
±28	28	52	12.0,22.6	16	^{13}C1′,^{15}N1/9
±47	38	72	24.3,22.7	16	^{13}C1′,^{15}N1/9
±60.2	48	90	34.7,19.9	16	^{13}C1′,^{15}N1/9
±67.7	46	86	34.8,14.3	16	^{13}C1′,^{15}N1/9
±74.7	45	84	35.0,9.6	16	^{13}C1′,^{15}N1/9
±83	44	82	34.8,4.3	16	^{13}C1′,^{15}N1/9
±51.8	43	80	28.6,22.5	16	^{13}C1′,^{15}N1/9
±42.1	34	64	20.2,22.3	16	^{13}C1′,^{15}N1/9
±32.6	30	56	14.6,22.9	16	^{13}C1′,^{15}N1/9
±24.4	26	50	10.3,22.6	16	^{13}C1′,^{15}N1/9
±16.4	24	46	6.6,22.4	16	^{13}C1′,^{15}N1/9
±8	23	44	3.1,22.4	16	^{13}C1′,^{15}N1/9
±79.1	44	82	34.6,6.7	16	^{13}C1′,^{15}N1/9
±76.1	44	82	34.3,8.5	16	^{13}C1′,^{15}N1/9
±73.3	45	84	34.8,10.4	16	^{13}C1′,^{15}N1/9
±69.7	45	84	34.3,12.7	16	^{13}C1′,^{15}N1/9
±65.7	46	86	34.4,15.5	16	^{13}C1′,^{15}N1/9
±62	47	88	34.4,18.3	16	^{13}C1′,^{15}N1/9
±58.4	49	92	35.0,21.5	16	^{13}C1′,^{15}N1/9

6. The 3D APSY-b-NCH (TROSY or MQ)

The pulse sequence of this experiment is published (14). As is suggested for large RNA in that article, the pulse sequence that includes TROSY elements was used for application with the K10.

6.1 Parameters for K10

Sample	*K10* (stem-loop RNA with 48 nucleotides)
Spectrometer	Bruker Avance III 700 MHz with cryogenic probe
Temperature	25°C
Total experiment time	26h
Nr. of projections	60
Interscan delay	0.8 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 11$, $R_{min} = 7$ Hz, $\Delta\nu = 2.1$ Hz, S/N = 3.6

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{13}C6/8	21.0	3697	158.0	23.1
ω_2	^{15}N1/9	35.0	2483	142.0	35.0
ω_3	^{1}H6/8	12.1	8503	4.7	120.4

Resolution-optimized angles, evolution times maximized.

α [°]	Time [min]	# of complex points in ind. dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	23	86	34.6	8	^{15}N1/9
90	22	84	22.7	8	^{13}C6/8
±31.2	31	116	21.0,34.7	8	^{13}C6/8,^{15}N1/9
±16.9	25	94	10.5,34.5	8	^{13}C6/8,^{15}N1/9
±50.5	25	96	22.7,18.7	8	^{13}C6/8,^{15}N1/9
±8.6	24	88	5.2,34.6	8	^{13}C6/8,^{15}N1/9
±67.6	23	88	22.9,9.5	8	^{13}C6/8,^{15}N1/9
±24.5	28	104	15.8,34.7	8	^{13}C6/8,^{15}N1/9
±38.9	29	110	22.9,28.3	8	^{13}C6/8,^{15}N1/9
±4.3	23	86	2.6,34.4	8	^{13}C6/8,^{15}N1/9
±78.4	23	86	23.0,4.7	8	^{13}C6/8,^{15}N1/9
±12.8	24	90	7.8,34.3	8	^{13}C6/8,^{15}N1/9
±58.2	24	92	23.0,14.2	8	^{13}C6/8,^{15}N1/9
±20.8	26	98	13.0,34.3	8	^{13}C6/8,^{15}N1/9
±44.1	28	104	23.1,23.9	8	^{13}C6/8,^{15}N1/9
±28	29	110	18.5,34.7	8	^{13}C6/8,^{15}N1/9
±34.7	31	118	22.9,33.1	8	^{13}C6/8,^{15}N1/9
±54.1	25	94	22.9,16.6	8	^{13}C6/8,^{15}N1/9
±47	27	100	22.9,21.4	8	^{13}C6/8,^{15}N1/9
±41.5	28	106	22.8,25.8	8	^{13}C6/8,^{15}N1/9
±35	31	118	23.0,32.9	8	^{13}C6/8,^{15}N1/9
±27	29	108	17.7,34.6	8	^{13}C6/8,^{15}N1/9
±6.4	24	88	3.9,35.0	8	^{13}C6/8,^{15}N1/9
±72	23	86	22.7,7.4	8	^{13}C6/8,^{15}N1/9
±14.8	25	92	9.1,34.5	8	^{13}C6/8,^{15}N1/9
±62	24	90	22.9,12.2	8	^{13}C6/8,^{15}N1/9
±22.6	27	102	14.5,34.9	8	^{13}C6/8,^{15}N1/9
±83	22	84	22.6,2.8	8	^{13}C6/8,^{15}N1/9
±10.6	24	90	6.5,34.9	8	^{13}C6/8,^{15}N1/9
±18.8	26	96	11.7,34.5	8	^{13}C6/8,^{15}N1/9
±2.4	23	86	1.4,34.6	8	^{13}C6/8,^{15}N1/9

7. The 4D APSY-HCCH COSY experiment

$^{1}H^{ribose}$-$^{13}C^{ribose}$-$^{13}C^{ribose}$-$^{1}H^{ribose}$

This experiment is based on the G-matrix Fourier transform (GFT) (4,3)D HCCH experiment for labeled nucleic acids (15).

7.1 Parameters for K10

Sample	*K10* (stem-loop RNA with 48 nucleotides)
Spectrometer	Bruker Avance III 900 MHz with cryogenic probe
Temperature	25°C
Total experiment time	58h
Nr. of projections	59
Interscan delay	1.0 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 12$, $R_{min} = 18$ Hz, $\Delta\nu = 4.0$ Hz, S/N = 5.0

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	$^{1}H_1$	4.4	3957	4.7	15.2
ω_2	$^{13}C_1$	34.0	7689	80.0	8.0
ω_3	$^{13}C_2$	34.0	7689	80.0	8.0
ω_4	$^{1}H_2$	20.0	18029	4.7	56.8

Resolution-optimized angles, evolution times maximized.

α [°]	β [°]	Time [min]	# of complex points in ind. dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	18	60	7.8	8	$^{13}C_2$
0	90	18	60	15.2	8	$^{13}C_1$
90	0	18	60	7.8	8	$^{1}H_1$
±45	0	51	86	7.9/7.9	16	$^{1}H_1/^{13}C_2$
0	±62.2	50	84	14.8/7.8	16	$^{13}C_1/^{13}C_2$
90	±62.2	50	84	14.8/7.8	16	$^{13}C_1/^{1}H_1$
±45	±53.3	124	104	14.9/7.9/7.9	32	$^{13}C_1/^{1}H_1/^{13}C_2$
±26.6	0	40	68	4.0/7.9	16	$^{1}H_1/^{13}C_2$
0	±43.5	40	68	7.5/7.9	16	$^{13}C_1/^{13}C_2$
90	±43.5	40	68	7.5/7.9	16	$^{13}C_1/^{1}H_1$
±63.4	0	40	68	7.9/4.0	16	$^{1}H_1/^{13}C_2$
0	±75.2	39	66	14.8/3.9	16	$^{13}C_1/^{13}C_2$
90	±75.2	39	66	14.8/3.9	16	$^{13}C_1/^{1}H_1$
±63.4	±73.6	81	68	14.9/3.9/2.0	32	$^{13}C_1/^{1}H_1/^{13}C_2$
±26.6	±59.5	107	90	15.0/3.9/7.9	32	$^{13}C_1/^{1}H_1/^{13}C_2$
±63.4	±40.3	88	74	7.5/7.9/3.9	32	$^{13}C_1/^{1}H_1/^{13}C_2$
±26.6	±23	83	70	3.8/4.0/8.0	32	$^{13}C_1/^{1}H_1/^{13}C_2$
±14	0	37	62	2.0/7.8	16	$^{1}H_1/^{13}C_2$
0	±25.4	37	62	3.7/7.8	16	$^{13}C_1/^{13}C_2$
90	±25.4	37	62	3.7/7.8	16	$^{13}C_1/^{1}H_1$
±36.9	0	45	76	5.9/7.9	16	$^{1}H_1/^{13}C_2$
0	±54.9	45	76	11.3/8.0	16	$^{13}C_1/^{13}C_2$
90	±54.9	45	76	11.3/8.0	16	$^{13}C_1/^{1}H_1$
±7.1	0	36	60	1.0/7.7	16	$^{1}H_1/^{13}C_2$
0	±13.3	36	60	1.8/7.7	16	$^{13}C_1/^{13}C_2$
90	±13.3	36	60	1.8/7.7	16	$^{13}C_1/^{1}H_1$

8. FLYA: Magnetization transfer pathways used for the generation of expected peaks

The FLYA algorithm (3) relies on the mapping of the chemical shifts in the input peak lists to the chemical shift that is predicted for each potentially measured nuclear resonance frequency. The chemical shift prediction for each resonance of the RNA based on the secondary structure is a precursor step that is done with the C++ script Chess2FLYA (1). The FLYA algorithm makes this chemical shift statistics, along with the magnetization transfer pathways of the experiments, to predict the expected cross peaks for the experiments. The algorithm as implemented in CYANA (16,17) provides a high degree of flexibility: it is able to perform automated assignment calculations also with peak lists that include correlations from magnetization pathways that have previously been unknown to CYANA. To run such a calculation, e.g. with peak lists from APSY experiments, the novel magnetization transfer pathways need to be defined. These definitions can be included either in the main CYANA library directly, or - as we recommend - as a separate library. The experiments could already be defined with their specific selectivity in the library; it is, however, more convenient to keep the experiment descriptions in the library general, and to define the specific selectivities in the input files. The 4D APSY-HC6CH5 COSY experiment, e.g., is first defined as the general HCCHcosy_4D experiment that is provided in the library, and its selectivity is complemented in the CYANA input file 'init.cya'. Details about these files, as well as about the general procedure, can be found in the CYANA WIKI (www.cyana.org), where also a FLYA tutorial is provided (www.cyana.org/wiki/index.php/Automated_resonance_assignment_with_FLYA). The input file used for the automated assignment of *K10* resonances is provided upon request by the authors. The current FLYA/CYANA release as well as Chess2FLYA are available from Peter Güntert (University Frankfurt) upon request.

CYANA library appendix with magnetization transfer pathways of APSY experiments for large RNA:

```
SPECTRUM  HCCH 4D H1 C1 C2 H2
0.980 H1:H_A*  C1:C_A*  C2:C_A*  H2:H_A*
0.980 H1:H_A*  C1:C_A*  H_A*  C2:C_A*  H2:H_A*
0.980 H1:H_A*  C1:C_A*  C_A*  C2:C_A*  H2:H_A*
0.800 H1:H_A*  C1:C_A*  C_A*  C_A*  C2:C_A*  H2:H_A*
0.800 H1:H_A*  C1:C_A*  C_A*  C_A*  C_A*  C2:C_A*  H2:H_A*

SPECTRUM  HCN 3D H C N
0.980 H:H_A*  C:C_A*  N:N_AMI

SPECTRUM  CHCHNO 4D H1 H2 C2 C1
0.980 H1:H_A*  C1:C_A*  H_A*  C2:C_A*  H2:H_A*
0.980 H1:H_ALI  C1:C_ALI  C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
0.784 H1:H_ARO C1:C_ARO C_ARO H_ARO C2:C_ARO H2:H_ARO
0.784 H1:H_ARO C1:C_ARO C_ARO C_ARO H_ARO C2:C_ARO H2:H_ARO
0.735 H1:H_ARO C1:C_ARO C_VIN  C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
0.735 H1:H_ALI  C1:C_ALI  C_VIN  C_ARO H_ARO C2:C_ARO H2:H_ARO
0.588 H1:H_ALI  C1:C_ALI  C_ALI  C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
0.294 H1:H_ALI  C1:C_ALI  N_AMI  C_BYL  N_AMI  C_ALI  C_ALI  C_ALI  C_ALI  N_AMI  C_BYL
C_ALI H_ALI  C2:C_ALI  H2:H_ALI
0.294 H1:H_ALI  C1:C_ALI  C_BYL  N_AMI  C_ALI  C_ALI  C_ALI  C_ALI  N_AMI  C_BYL  N_AMI
C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
0.980 H1:H_ALI  C1:C_ALI  S_RED C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
0.735 H1:H_ALI  C1:C_ALI  S_RED C_ALI  C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
```

```
0.735 H1:H_ALI  C1:C_ALI  C_ALI  S_RED C_ALI  H_ALI  C2:C_ALI  H2:H_ALI
```

Selectivity and NOE peak list definitions in input file init.cya :

```
arocarbon:=”C8  @RADE + C8 @RGUA + C6 @RCYT + C6 @URA + C2 @RADE”
alicarbon:=”C1\′  + C2\′  + C3\′  + C4\′  + C5\′”

command CH268CHnoesy_4D_expect peaks
spectrum CH268CHnoesy_4D C1=”$arocarbon” append distance=4.0 structures=12 probability=0.9
spectrum CH268CHnoesy_4D C1=”$arocarbon” append distance=5.0 structures=12 probability=0.8
spectrum CH268CHnoesy_4D C1=”$arocarbon” append distance=7.0 structures=12 probability=0.3
write
end


command CH1CHnoesy_4D expect peaks
spectrum CH1CHnoesy_4D C1=”C1\′” append distance=4.0 structures=12 probability=0.9
spectrum CH1CHnoesy_4D C1=”C1\′” append distance=5.0 structures=12 probability=0.8
spectrum CH1CHnoesy_4D C1=”C1\′” append distance=7.0 structures=12 probability=0.3
write
end

command HCCHcosy_4D_expect peaks
spectrum HCCHcosy_4D C1=”$alicarbon” append
end

command HC6CH5_4D_expect peaks
spectrum HC6CH5 4D C1=”C6” C2=”C5” append
end

command HCNsugar_3D expect peaks
spectrum HCNsugar 3D H=”H1\′”
end

command HCNbase_3D_expect peaks
spectrum HCNbase_3D C=”$arocarbon” N=”N9 @RADE + N9 @RGUA + N1 @RCYT + N1 @URA”
append
end
```

9. References

1. Aeschbacher, T., Schmidt, E., Blatter, M., Maris, C., Duss, O., Allain, F. H. T., Güntert, P., and Schubert, M. (2013) Automated and assisted RNA resonance assignment using NMR chemical shift statistics. *Nucleic Acids Res.* **41**, e178
2. Aeschbacher, T., Schubert, M., and Allain, F. H. T. (2012) A procedure to validate and correct the C-13 chemical shift calibration of RNA datasets. *J. Biomol. NMR* **52**, 179-190
3. Schmidt, E., and Güntert, P. (2012) A New Algorithm for Reliable and General NMR Resonance Assignment. *J. Am. Chem. Soc.* **134**, 12817-12829
4. Krähenbühl, B., El Bakkali, I., Schmidt, E., Güntert, P., and 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
5. Hiller, S., Wider, G., and Wuthrich, K. (2008) APSY-NMR with proteins: practical aspects and backbone assignment. *J. Biomol. NMR* **42**, 179-195
6. Krähenbühl, B., and Wider, G. (2012) Automated Projection Spectroscopy (APSY) for the Assignment of NMR Resonances of Biological Macromolecules. *Chimia* **66**, 770-774
7. Hiller, S., Fiorito, F., Wuthrich, K., and Wider, G. (2005) Automated projection spectroscopy (APSY). *Proc. Natl. Acad. Sci. USA* **102**, 10876-10881
8. Güntert, P., Dötsch, V., Wider, G., and Wuthrich, K. (1992) Processing of multidimensional NMR data with the new software PROSA. *J. Biomol. NMR* **2**, 619-629
9. Geen, H., and Freeman, R. (1991) Band-selective radiofrequency pulses. *J. Magn. Reson.* **93**, 93-141
10. Shaka, A. J., and Keeler, J. (1987) Broadband spin decoupling in isotropic liquids. *Prog. Nucl. Magn. Reson. Spectrosc.* **19**, 47-129
11. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) Rapid recording of 2D NMR-spectra without phasecycling - Application to the study of hydrogen-exchange in proteins. *J. Magn. Reson.* **85**, 393-399
12. Bax, A., and Freeman, R. (1981) Investigation of complex networks of spin-spin coupling by two-dimensional NMR. *J. Magn. Reson.* **44**, 542-561
13. Kadkhodaie, M., Rivas, O., Tan, M., Mohebbi, A., and Shaka, A. J. (1991) Broad-band homonuclear cross polarization using flip-flop spectroscopy. *J. Magn. Reson.* **91**, 437-443
14. Krähenbühl, B., Hofmann, D., Maris, C., and 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
15. Atreya, H. S., Sathyamoorthy, B., Jaipuria, G., Beaumont, V., Varani, G., and 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
16. Güntert, P. (2003) Automated NMR protein structure calculation. *Prog. Nucl. Magn. Reson. Spectrosc.* **43**, 105-125
17. Lopez-Mendez, B., and Güntert, P. (2006) Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc.* **128**, 13112-13122