Supplementary Material

Automated NMR resonance assignment strategy for RNA via the phosphodiester backbone based on high-dimensional through-bond APSY experiments

Barbara Krähenbühl[1] · Issam El Bakkali[1] · Elena Schmidt[2] · Peter Güntert[2] · Gerhard Wider[1,*]

[1]*Institute of Molecular Biology and Biophysics, ETH Zurich*
8093 Zurich, Switzerland
[2]*Institute of Biophysical Chemistry, Goethe University Frankfurt am Main*
60438 Frankfurt am Main, Germany

*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. Setup of APSY experiments

The APSY NMR experiments were performed on a 500, and a 600 MHz Bruker Avance III spectrometer with cryogenic probes; the 500 MHz spectrometer was equipped with a triple-resonance probe (^{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 3.0 (Bruker, Karlsruhe, Germany). A detailed description of the setup of APSY experiments with the Topspin software is provided in the Appendix at the end of this Supplementary Material.

For all spectra, 1024 complex data points (TD = 2048) were measured in the acquisition dimension, multiplied with a 90°-shifted sine-square-bell function, and zero-filled to SI = 4096 complex data points. 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. The SI values in the projection spectra are usually set to 2-times TD in the acquisition dimension, and to ≥2-times the maximal TD in the indirect dimensions of the parent data set. The time domain data was processed using the software PROSA (Güntert et al., 1992) including the addition/subtraction operations to obtain individual 2D projections. The latter were analyzed by the software GAPRO (Hiller et al., 2005) resulting in high-dimensional peak lists which were further processed by the algorithm FLYA (Schmidt and Güntert, 2012). The specific parameters used for the setup, measurement and processing of the novel APSY experiment for SL14 are listed in the sections 4.1, 5.1, 6.2, 7.2 and 7.3 below. The use of MATLAB scripts for the analysis of the results, such as e.g. the comparison of the APSY resonance assignment with the previous manual assignment, has been described in the main text. The scripts are available upon request.

2. Use of FLYA

The FLYA algorithm (Schmidt and Güntert, 2012) 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++ program Chess2FLYA (Aeschbacher et al. (2013)). The FLYA algorithm applies the magnetization transfer pathways of the experiments to predict the expected cross peaks for the experiments and then maps these expected cross peaks to experimental ones, using the chemical shift statistics. The algorithm as implemented in CYANA (Güntert, 2003; Lopez-Mendez and Güntert, 2006) 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 new magnetization transfer pathways need to be defined. These definitions can be included either in the main CYANA library directly, or - as we recommend - in 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 HCCH 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 current FLYA/CYANA release as well as Chess2FLYA are available from Peter Güntert (University Frankfurt).

CYANA library appendix (or inclusion as a separate library file, e.g. 'APSY.lib', that can be read with the command 'read lib APSY.lib append') with magnetization transfer pathways of APSY experiments for 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 HCNCH_5D H1′ C1′ N1 C6 H6
0.980 H1′:H_ALI C1′:C_ALI N1:N_AMI C6:C_ARO H6:H_ARO

SPECTRUM HPCCH_5D H1 P C1 C2 H2
0.700 H2:H_ALI C2:C_ALI H1:H_ALI C1:C_ALI O_EST P:P_ALI
0.700 H2:H_ALI C2:C_ALI H1:H_ALI C1:C_ALI C_ALI O_EST P:P_ALI
0.700 H2:H_ALI C2:C_ALI C1:C_ALI H1:H_ALI C_ALI O_EST P:P_ALI
0.700 H2:H_ALI C2:C_ALI C1:C_ALI H1:H_ALI C_ALI C_ALI O_EST P:P_ALI
```

Selectivity definitions in the input file *init.cya*:

```
command HC1CHtocsy_4D expect peaks
spectrum HC1CHtocsy_4D C1="C1\′ " append
end

command HC5CHtocsy_4D expect peaks
spectrum HC1CHtocsy_4D C1="C5\′ " append
end

command HC6CH5_4D expect peaks
spectrum HC6CH5_4D C1="C6" C2="C5" append
end
```

3. Precision of APSY peak lists

To assess the precision of the APSY peak lists, we calculated the standard deviation of the chemical shift values over all peaks that are assigned to a given atom in any of the three APSY spectra that were used as input for the automated chemical shift assignment with the FLYA algorithm. The resulting standard deviations are 0.005 ppm for ^{1}H, 0.021 ppm for ^{13}C and 0.01 ppm for ^{31}P. This is significantly less than usually achieved with conventional 3D spectra and shows the high inherent chemical shift precision of the APSY approach. These chemical shift precisions are somewhat lower than those that have been reported e.g. for 7D APSY peak lists of soluble nonglobular proteins (Hiller et al., 2007) because the present APSY data sets were recorded with a considerably lower number of projections and shorter maximal evolution times.

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

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

Reference: Krähenbühl et al., submitted to J. Biomol. NMR

4.1. Parameters for SL14

Sample	SL14 (14 residues, cUUCGg tetraloop)
Spectrometer	Bruker Avance III 600 MHz with cryogenic probe
Temperature	25°C
Total experiment time	5 h 28 min
Nr. of projections	41
Interscan delay	1.0 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 12, R_{min} = 25$ Hz, $\Delta\nu = 5.0$ Hz, S/N = 5.0

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{1}H6	1.1	660	7.6	63.6
ω_2	^{13}C6	5.0	755	142.5	49.0
ω_3	^{13}C5	10.0	1509	101.0	49.7
ω_4	^{1}H5	12.9	7764	4.7	131.9

Dispersion-optimized angles, evolution times not maximized.

α [°]	β [°]	Time [min]	# of complex points in indirect dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	12	74	49.0	4	^{13}C5
0	90	7	42	63.6	4	^{1}H6
90	0	6	36	47.7	4	^{13}C6
90	±48.8	7	38	40.7/35.6	4	^{1}H6/^{13} C6
0	±66.4	8	48	51.4/22.5	4	^{1}H6/^{13} C5
±63.5	0	8	46	43.2/21.5	4	^{13}C6/^{13} C5
±63.5	±45.6	8	44	38.5/33.7/16.8	4	^{1}H6/^{13} C6/^{13} C5
90	±66.4	7	40	54.2/23.7	4	^{1}H6/^{13} C6
±45	0	10	58	34.4/34.4	4	^{13}C6/^{13} C5
0	±48.8	10	58	39.3/34.4	4	^{1}H6/^{13} C5
90	±29.7	7	38	25.7/45.1	4	^{1}H6/^{13} C6
±76	0	7	40	47.4/11.8	4	^{13}C6/^{13} C5
0	±77.7	8	44	59.7/13.0	4	^{1}H6/^{13} C5
±76	±65.7	7	40	52.9/23.2/5.8	4	^{1}H6/^{13} C6/^{13} C5
±45	±58.3	8	46	46.5/20.3/20.3	4	^{1}H6/^{13} C6/^{13} C5
±76	±29	7	40	24.7/43.3/10.8	4	^{1}H6/^{13} C6/^{13} C5
±45	±22	10	56	18.5/32.4/32.4	4	^{1}H6/^{13} C6/^{13} C5

5. The 5D APSY-HCNCH experiment

^{1}H1′-^{13}C1′-^{15}N1/9-^{13}C6/8-^{1}H6/8

Sugar-to-base correlations for small and intermediate RNA.

Reference: Krähenbühl et al. (2012)

5.1. Parameters for SL14

Sample	SL14 (14 residues, cUUCGg tetraloop)
Spectrometer	Bruker Avance III 500 MHz with cryogenic probe
Temperature	25°C
Total experiment time	1 h 33 min
Nr. of projections	16
Interscan delay	1 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 6$, $R_{min} = 30$ Hz, $\Delta\nu = 4.0$ Hz, S/N = 3.4

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{1}H1′	3.4	1700	4.7	34.1
ω_2	^{13}C1′	9.0	1132	90.5	33.6
ω_3	^{13}C6/8	35.0	1774	160	33.8
ω_4	^{13}C6/8	11.0	1384	139.5	16.6
ω_5	^{1}H6/8	17.0	8503	4.7	-

Resolution-optimized angles, evolution times maximized.

α [°]	β [°]	γ [°]	Time [min]	# of complex points in indirect dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	0	3	22	15.9	2	^{13}C6/8
0	0	90	6	58	34.1	2	^{1}H1′
0	90	0	4	38	33.6	2	^{13}C1′
90	0	0	6	60	33.8	2	^{15}N1/9
0	90	±45.4	7	68	33.4/33.0	2	^{1}H1′/^{13}C1′
±63.8	0	0	6	64	33.7/16.6	2	^{15}N1/9/^{13}C6/8
90	0	±45.2	8	82	33.5/33.3	2	^{1}H1′/^{15}N1/9
0	±63.7	0	4	44	33.3/16.4	2	^{13}C1′/^{13}C6/8
90	±44.8	0	7	70	33.1/33.3	2	^{13}C1′/^{15}N1/9
0	0	±64	6	62	33.9/16.5	2	^{1}H1′/^{13}C6/8

6. The 5D APSY-HCP-CCH COSY

$^{1}H_{i}^{ribose}$-$^{31}P_{i/i+1}$-$^{13}C_{i}^{ribose}$-$^{13}C_{i}^{ribose}$-$^{1}H_{i}^{ribose}$

Phosphodiester-backbone correlations for small RNA.

6.1. Pulse sequence

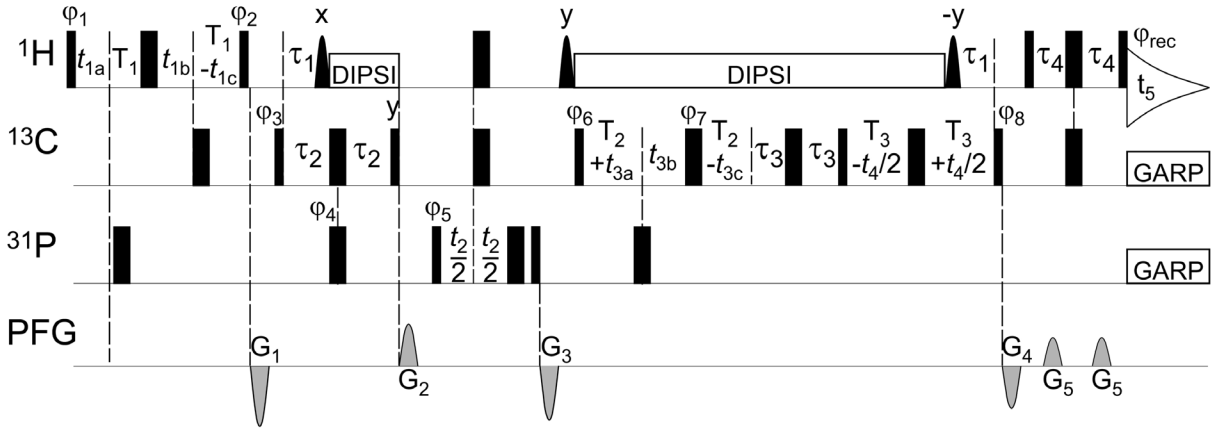

Figure S1: Pulse sequence of the 5D APSY- HCP-CCH COSY experiment. Black thin and wide rectangular bars represent 90° and 180° high-power radio-frequency pulses, respectively. The carrier frequencies are set to 4.7 ppm, 76.5 ppm, and -1.2 ppm on the ^{1}H, ^{13}C, and ^{31}P channel, respectively. All parameters apply for measurements at a field of 600 MHz. During the ^{31}P-^{13}C and the ^{13}C-^{13}C transfer, ^{1}H are decoupled with DIPSI-3 (Shaka et al., 1988); during acquisition, ^{13}C and ^{31}P are decoupled with GARP-4 (Shaka and Keeler, 1987). The time periods are set to τ_1 = 1/(3JCH) = 2.2 ms, $\tau_2 = T_2$ = 1/(4JCP) = 12.0 ms, $T_1 = \tau_4$ = 1/(4JCH) = 1.5 ms, and $\tau_3 = T_3$ = 1/(4JCC) = 3.3 ms. The shaped proton pulses represent rectangular 90° pulses at decoupling power. All pulses are applied along the x-axis unless indicated otherwise above the pulse symbol. The following phase cycles are applied: ϕ_1 = 4(x), 4(-x); ϕ_2 = y; ϕ_3 = 2(x), 2(-x); ϕ_4 = x; ϕ_5 = x, -x; ϕ_6 = x; ϕ_7 = x; ϕ_8 = 8(x), 8(-x); ϕ_{rec} = α, -α, -α, α with α = x, 2(-x), x (receiver phase). Quadrature detection for the indirect dimensions is achieved in States-TPPI manner by incrementing the following delays and phases: for the $^{1}H1$ evolution t_{1a} -t_{1c} and ϕ_2 , for ^{31}P t_2 and ϕ_4 /ϕ_5 , for $^{13}C_1$ t_{3a} -t_{3c} and ϕ_6 , and for $^{13}C_2$ t_4 and ϕ_8 . For $^{1}H_1$ and $^{13}C_1$ were semi-constant time evolutions applied, and for $^{13}C_2$ a constant time evolution. The trigonometric addition theorem was used to obtain pure cosine and sine terms for a subsequent hypercomplex Fourier transformation (Brutscher et al., 1995; Kupce and Freeman, 2004). The gray sine-bell shaped pulsed field gradient (PFG) pulses were applied with a length of 1000 µs and the following strengths: G_1 : -27.5 G/cm; G_2 : 16.5 G/cm; G_3 : -24.75 G/cm; G_4 : -16.5 G/cm; G_5 : 9.35 G/cm

6.2. Parameters

Sample	SL14 (14 residues, cUUCGg tetraloop)
Spectrometer	Bruker Avance III 600 MHz with cryogenic quadruple probe
Temperature	25°C
Total experiment time	13h 55 min
Nr. of projections	28
Interscan delay	1.0 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 7$, $R_{min} = 18$ Hz, $\Delta\nu = 4.0$ Hz, S/N = 3.8

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	$^{1}H_1$	2.8	1680	4.7	8.0
ω_2	^{31}P	2.0	486	-1.2	11.2
ω_3	$^{13}C_1$	28.0	4226	76.5	19.7
ω_4	$^{13}C_2$	28.0	4226	76.5	6.0
ω_5	$^{1}H_2$	16.0	9615	4.7	102.5

Resolution-optimized angles, evolution times maximized.

α [°]	β [°]	γ [°]	Time [min]	# of complex points in indirect dim.	Evolution time [ms]	Nr. of scans	Ind. dim. frequencies
0	0	0	8	26	6.2	8	$^{13}C_2$
0	0	90	11	34	20.2	8	$^{1}H_1$
0	90	0	6	18	37.0	8	^{31}P
90	0	0	32	100	23.7	8	$^{13}C_1$
0	90	±16.1	17	26	10.9/37.9	16	$^{1}H_1/^{31}P$
±45	0	0	24	38	6.4/6.4	16	$^{13}C_1/^{13}C_2$
90	0	±68.3	31	48	20.2/8.0	16	$^{1}H_1/^{13}C_1$
0	±83.5	0	17	26	38.0/4.3	16	$^{31}P/^{13}C_2$
90	±83.5	0	17	26	38.0/4.3	16	$^{31}P/^{13}C_1$
0	0	±68.3	24	38	16.0/6.4	16	$^{1}H_1/^{13}C_2$
0	90	±27.3	25	38	19.7/38.2	16	$^{1}H_1/^{31}P$
±75	0	0	66	104	23.8/6.4	16	$^{13}C_1/^{13}C_2$
90	0	±40.3	68	106	20.2/23.8	16	$^{1}H_1/^{13}C_1$
0	±60	0	17	26	10.5/6.0	16	$^{31}P/^{13}C_2$
90	±58.6	0	66	102	38.9/23.7	16	$^{31}P/^{13}C_1$
0	0	±50	18	28	7.1/6.0	16	$^{1}H_1/^{13}C_2$

7. The 4D APSY-HCCH TOCSY

(selective in first and constant- time in second ^{13}C evolution period; or adenine ^{13}C2-selective in second ^{13}C evolution period / no constant-time)

Intra-ribose correlations for small and intermediate size RNA:

^{1}H1′-^{13}C1′-^{13}C^{ribose}-^{1}H^{ribose}

^{1}H5′/″-^{13}C5′-^{13}C^{ribose}-^{1}H^{ribose}

Complementary experiment for adenine:

^{1}H8-^{13}C8-^{13}C2-^{1}H2

7.1. Pulse sequence

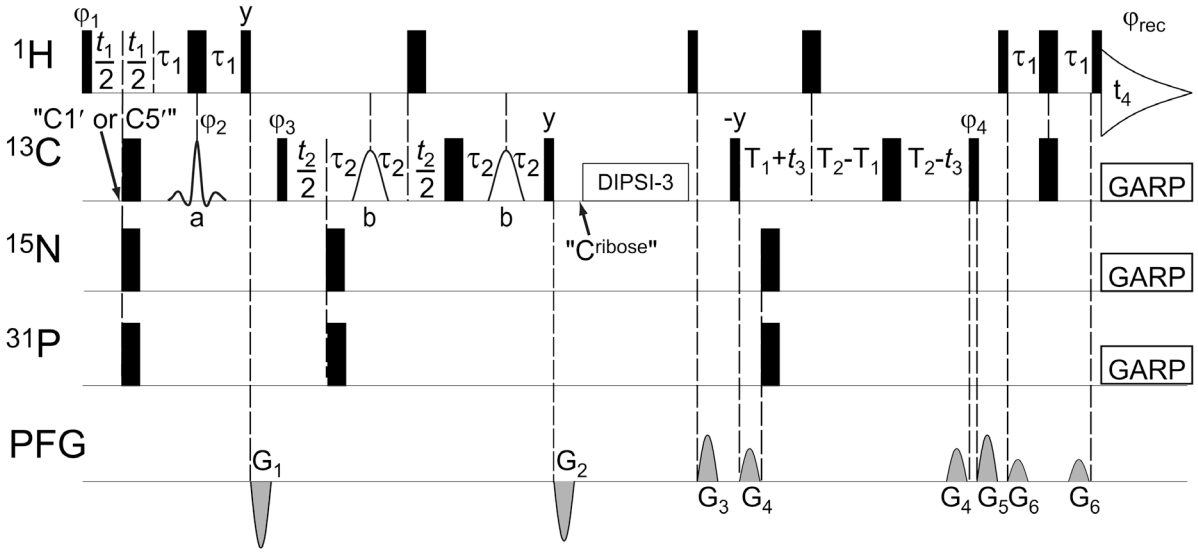

Figure S2: Pulse sequence of the 4D APSY-HCCH TOCSY experiment with selectivity in the first ^{13}C dimension. The offsets and the lengths of the selective pulses in this description are adjusted for the coherence transfer pathway ^{1}H1′-^{13}C1′-^{13}C^{ribose} -^{1}H^{ribose} or alternatively ^{1}H5′/″-^{13}C5′-^{13}C^{ribose} -^{1}H^{ribose} . The carrier frequency of the ^{1}H channel is set to 4.7 ppm; the ^{13}C carrier frequency is first set to 91.5 ppm (^{13}C1′) or 67.0 ppm (^{13}C5′) as labeled with an arrow and " C1′ or C5′ ", and before the TOCSY transfer to 80 ppm ("C^{ribose}"); the ^{15}N carrier frequency is set to 160 ppm. The ^{31}P channel is included only for probes with the corresponding equipment, for decoupling of the small ^{13}C-^{31}P or ^{1}H-^{31}P scalar couplings; its carrier frequency is set to -1.2 ppm. All parameters apply for measurements at a field of 500 MHz. Thin and wide black bars represent non-selective 90° and 180° high-power pulses, respectively. The shaped pulse (a) is an on-resonance Reburp (Geen and Freeman, 1991) 180° pulse with a duration of 2800 µs (^{13}C1′) or 3000 µs (^{13}C5′); (b) is an off-resonance Sinc pulse (central lobe) with a duration of 1400 µs (^{13}C2′ at 76 ppm) or 760 µs (^{13}C4′ at 83 ppm). ^{13}C TOCSY mixing is performed during 26 ms with four DIPSI-3 (Shaka et al., 1988) cycles with a field strength of 8.33 kHz. GARP (Shaka and Keeler, 1987) decoupling sequences are applied during acquisition on ^{13}C with a field strength of 2.9 kHz, on ^{15}N with 1.56 kHz, and on ^{31}P with 1.13 kHz. The time periods are set to τ_1 = 1/(4JCH) = 2 ms, τ_2 = 1.6 ms, T_1 = 1/(6JCH) = 1.1 ms, and T_2 = 1/(2JCC)= 12 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 (Marion et al., 1989): ϕ_1 was incremented for the evolution period t_1 (^{1}H1′ or ^{1}H5′/″), ϕ_3 for t_2 (^{13}C1′ or ^{13}C5′), and ϕ_4 for the constant-time evolution period t_3 ($^{13}C^{ribose}$); t_4 denotes the acquisition time. The trigonometric addition theorem was used to obtain pure cosine and sine terms for a subsequent hypercomplex Fourier transformation (Brutscher et al., 1995; Kupce and Freeman, 2004). The gray sinebell shaped gradient pulses (PFG) were applied along the z-axis. Their 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 : 700 µs, 16.5 G/cm; G_5 : 1000 µs, 44 G/cm; G_6 : 800 µs, 9.9 G/cm.

The pulse sequence of the 4D APSY-H8C8(CC-TOCSY)C2H2 experiment for adenine is similar to the one presented in this figure for ribose. The experiment deviates in the following parameters: The ^{31}P channel elements are discarded. Pulses (b) are not applied, (a) is a conventional hard pulse, and the last 180° pulse on ^{13}C before acquisition is an on-resonance Reburp pulse on ^{13}C2 with a duration of 2480 µs. The ^{13}C carrier frequency is first set to 145 ppm (^{13}C8), during the TOCSY mixing period to 150 ppm, and afterwards to 153 ppm (^{13}C2). ^{13}C TOCSY mixing is performed during 106 ms with seven DIPSI3 cycles with a field strength of 3.6 kHz. The time periods are set to τ_1 =1/(4JCH)=2 ms, τ_2 = T_1 = T_2= 1.25 ms. The ^{13}C evolution time is thus not a constant-time evolution period: all three indirect evolution periods are performed in semi-constant time manner.

7.2. 4D APSY-H1′C1′CH TOCSY parameters for SL14

Sample	SL14 (14 residues, cUUCGg tetraloop)
Spectrometer	Bruker Avance III 500 MHz with cryogenic probe
Temperature	25℃
Total experiment time	12 h 1 min
Nr. of projections	41
Interscan delay	1.0 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 7$, $R_{min} = 20$ Hz, $\Delta\nu = 5.0$ Hz, S/N = 5.0

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{1}H1′	3.6	1800	4.7	35.0
ω_2	^{13}C1′	8.6	1082	91.5	50.0
ω_3	$^{13}C^{ribose}$	42.0	5282	80.0	22.1
ω_4	$^{1}H^{ribose}$	15.5	7764	4.7	131.9

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	19	116	22.0	4	$^{13}C^{ribose}$
0	90	11	62	34.4	4	^{1}H1′
90	0	10	54	49.9	4	^{13}C1′
90	±35	14	82	34.6/49.4	4	^{1}H1′/^{13}C1′
0	±57.7	22	132	34.8/22.0	4	^{1}H1′/$^{13}C^{ribose}$
±66.1	0	22	128	49.6/22.0	4	^{13}C1′/$^{13}C^{ribose}$
±66.1	±32.6	24	142	34.6/49.5/21.9	4	^{1}H1/^{13}C1/$^{13}C^{ribose}$
90	±54.5	12	68	34.7/24.8	4	^{1}H1′/^{13}C1′
±48.4	0	20	120	24.9/22.1	4	^{13}C1 /$^{13}C^{ribose}$
0	±38.3	20	120	17.3/21.9	4	^{1}H1′/$^{13}C^{ribose}$
90	±19.3	11	62	17.3/49.5	4	^{1}H1′/^{13}C1′
±77.5	0	14	78	48.9/10.8	4	^{13}C1′/$^{13}C^{ribose}$
0	±72.4	15	86	35.0/11.1	4	^{1}H1′/$^{13}C^{ribose}$
±77.5	±53.8	13	74	34.7/24.8/5.5	4	^{1}H1′/^{13}C1′/$^{13}C^{ribose}$
±48.4	±46.4	23	134	34.6/24.6/21.9	4	^{1}H1′/^{13}C1′/$^{13}C^{ribose}$
±77.5	±18.9	15	84	17.2/49.0/10.9	4	^{1}H1′/^{13}C1′/$^{13}C^{ribose}$
±48.4	±14.7	20	120	8.7/24.7/21.9	4	^{1} H1′/^{13}C1′/$^{13}C^{ribose}$

7.3. 4D APSY-H5′/′′C5′CH TOCSY parameters for SL14

Sample	SL14 (14 residues, cUUCGg tetraloop)
Spectrometer	Bruker Avance III 500 MHz with cryogenic probe
Temperature	25°C
Total experiment time	6h 51 min
Nr. of projections	25
Interscan delay	1.0 s

GAPRO parameters

$S_{min,1} = S_{min,2} = 7$, $R_{min} = 15$ Hz, $\Delta\nu = 5.0$ Hz, S/N = 5.0

Dimension	Nucleus	Sweep width [ppm]	Sweep width [Hz]	Carrier frequency [ppm]	Maximal evolution time [ms]
ω_1	^{1}H5′	4.4	2201	4.7	25.0
ω_2	^{13}C5′	10.0	1258	67.0	35.0
ω_3	$^{13}C^{ribose}$	42.0	5282	80.0	22.1
ω_4	$^{1}H^{ribose}$	15.5	7764	4.7	131.9

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	21	116	22.0	4	$^{13}C^{ribose}$
0	90	9	54	24.5	4	^{1}H5′′′
90	0	8	44	35.0	4	^{13}C5′
±57.7	0	21	124	34.8/22.0	4	^{13}C5 / $^{13}C^{ribose}$
0	±48.5	22	128	24.8/21.9	4	^{1}H5′′′/$^{13}C^{ribose}$
90	±35.5	12	70	24.8/34.8	4	^{1}H5′′/^{13}C5′
±57.7	±31.1	23	136	24.9/34.9/22.1	4	^{1}H5′′′/^{13}C5′/$^{13}C^{ribose}$
±38.3	0	20	118	17.3/22.0	4	^{13}C5′/$^{13}C^{ribose}$
0	±29.4	20	120	12.5/22.1	4	^{1}H5′′′/$^{13}C^{ribose}$
90	±19.7	9	50	12.1/33.7	4	^{1}H5 ′′/^{13}C5′
±72.4	0	12	72	34.4/10.9	4	^{13}C5′/$^{13}C^{ribose}$
0	±66.1	14	80	24.9/11.0	4	^{1}H5′′′/$^{13}C^{ribose}$
90	±55	10	58	24.5/17.1	4	^{1}H5′′′/^{13}C5′
±72.4	±53.7	11	66	25.0/17.5/5.5	4	^{1}H5′′′/^{13}C5′/$^{13}C^{ribose}$
±38.3	±41.5	22	130	24.8/17.3/22.0	4	^{1}H5′′′/^{13}C5′/$^{13}C^{ribose}$
±72.4	±18.8	13	78	12.4/34.9/11.1	4	^{1}H5′′′/^{13}C5′/$^{13}C^{ribose}$
±38.3	±12.5	19	118	6.2/17.2/21.8	4	^{1}H5′′′/^{13}C5′/$^{13}C^{ribose}$

8. Example spectra

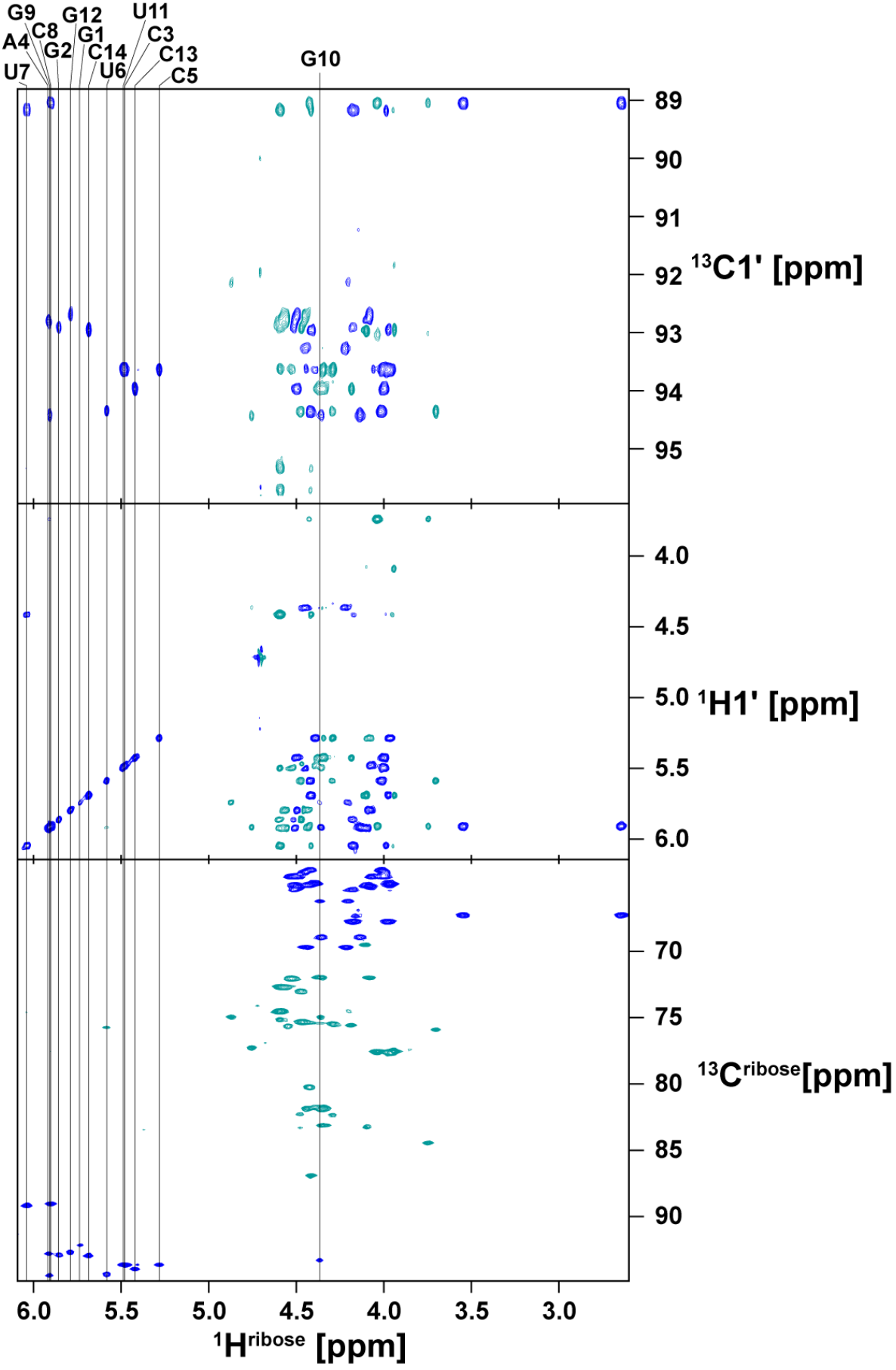

Figure S3: 2D projection spectra of the selective constant-time 4D APSY-H1′C1′CH TOCSY experiment measured with the orthogonal angles 90°/0° (^{13}C1′-^{1}H^{ribose}), 0°/90° (^{1}H1′-^{1}H^{ribose}), and 0°/0° (^{13}C^{ribose}-^{1}H^{ribose}). The parameters of this experiment are listed in the tables in the section 7.2 of this Supplementary Material: note the long evolution times of 35 ms, 50 ms and 22.1 ms for the ^{1}H1′, the ^{13}C1′ and the ^{13}C^{ribose} experiment, and the resulting high precision of the projection spectra in this APSY experiment. The negative peaks (green) are projections of 4D peaks that end on ^{13}C-^{1}H groups with two ^{13}C neighbors (^{13}C2′/^{13}C3′/^{13}C4′), whereas the positive peaks (blue) represent those that end on ^{13}C-^{1}H groups with one ^{13}C neighbors (^{13}C1′/^{13}C5′). Each vertical line that crosses all three projections marks a set of projection peaks that belongs to the same 4D ^{1}H1′-^{13}C1′-^{13}C1′-^{1}H1′ peak of the residue that is indicated above the spectra. Orthogonal projections represent the 2D projection spectra with only one active evolution period in the indirect dimension; this allows the comparison of completeness and resolution with the corresponding 2D spectra of other APSY or non-APSY experiments. The ^{13}C^{ribose} -^{1}H^{ribose} spectrum at the bottom corresponds to a [^{13}C]-HSQC spectrum. A corresponding HSQC fingerprint spectrum with assignments was published by Fürtig et al (2003).

9. Appendix: running APSY experiments with Topspin

The APSY experiments are started from a parent data set with the full dimensionality, which is set up like a conventional correlation experiment with the spectrometer operation software Topspin (Bruker, Karlsruhe, Germany). All further steps are performed with the program *manageapsy* in Topspin. First an angle set is created that is adjusted to the parameters in the parent data set. In addition the corresponding 2D projection data sets are created in a different directory. These projection experiments can either be run conventionally with the command ZG, 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 (Hiller et al., 2008). We provide a short description of the setup and analysis steps in the following, and also include some advice what to pay particular attention to. Topspin-specific names of parameters and functions are used without further explanations; they can be looked up in the Topspin manuals if required.

9.1. Setup of APSY parent data set

The spectral width (SW), the maximal evolution time (AQ), the nucleus type (NUC1), and the carrier frequencies are provided in the experiment tables in sections 4.1, 5.1, 6.2, 7.2 and 7.3; they are also usually listed in a table in the header of the APSY pulse program (e.g. provided with Topspin). 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; in 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 is contained in the header of the pulse programs.

An important parameter to set before *manageapsy* can be used is USERA5, which has e.g. for the 4D APSY-HC6CH5 TOCSY the form ’4D HCCH 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 (evolution-optimized) or disopt (dispersion-optimized)]. This parameter, along with the experimental parameters in the parent data set and the angle file, fully define the setup of the 2D projection spectra. The additional parameters (e.g. angles) are set up automatically by *manageapsy,* but can also be set manually.

9.2. 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 evolution-optimized angles ('resopt' as USERA5 parameter). Their concepts are explained, e.g., in (Krähenbühl and Wider, 2012). Usually, 'disopt' is used if most evolution periods have a decaying envelope, and 'resopt' for mostly 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*.

9.3. 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) which involve only one active indirect dimension per spectrum are measured first. 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 the case 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.

9.4. 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 as for any standard experiment; the initial parameters of each 2D data set are 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 for reprocessing is thus as follows:

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

9.5. 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 (Hiller et al., 2005). It is usually not required to use 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.

9.6. 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 (Güntert et al., 1992) 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).

10. References

Aeschbacher, T., Schmidt, E., Blatter, M., Maris, C., Duss, O., Allain, F.H., Güntert, P., Schubert, M.: Automated and assisted RNA resonance assignment using NMR chemical shift statistics. Nucleic Acids Res 40(18), e128 (2013)

Brutscher, B., Morelle, N., Cordier, F., Marion, D.: Determination of an initial set of NOE-derived distance constraints for the structure determination of $^{15}N/^{13}C$-labeled proteins. Journal of Magnetic Resonance Series B 109(2), 238–242 (1995)

Fürtig, B., Richter, C., Wöhnert, J., Schwalbe, H.: NMR spectroscopy of RNA. Chembiochem 4(10), 936-962 (2003)

Geen, H., Freeman, R.: Band-selective radiofrequency pulses. J Magn Reson 93(1), 93–141 (1991)

Güntert, P.: Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43(3-4), 105–125 (2003)

Güntert, P., Dötsch, V., Wider, G., Wüthrich, K.: Processing of multidimensional NMR data with the new software PROSA. J Biomol NMR 2(6), 619–629 (1992)

Hiller, S., Fiorito, F., Wüthrich, K., Wider, G.: Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102(31), 10876–10881 (2005)

Hiller, S., Fiorito, F., Wüthrich, K., Wider, G.: Sequence-Specific Resonance Assignment of Soluble Nonglobular Proteins by 7D APSY-NMR Spectroscopy. J. Am. Chem. Soc. 129, 10823–10828 (2007)

Hiller, S., Wasmer, C., Wider, G., Wüthrich, K.: APSY-NMR with proteins: practical aspects and backbone assignment. J Biomol NMR 42(3), 179–195 (2008)

Krähenbühl, B., Hofmann, D., Maris, C., Wider, G.: Sugar-to-base correlation in nucleic acids with a 5D APSY-HCNCH or two 3D APSY-HCN experiments. J Biomol NMR 52(2), 141–150 (2012)

Krähenbühl, B., Wider, G.: Automated projection spectroscopy (APSY) for the as- signment of NMR resonances of biological macromolecules. Chimia 66(10), 770–774 (2012)

Kupce, E., Freeman, R.: Projection-reconstruction technique for speeding up multidi-mensional NMR spectroscopy. J Am Chem Soc 126(20), 6429–6440 (2004)

Lopez-Mendez, B., Güntert, P.: Automated protein structure determination from NMR spectra. J Am Chem Soc 128(40), 13112–13122 (2006)

Marion, D., Ikura, M., Tschudin, R., Bax, A.: Rapid recording of 2D NMR spectra withoug phase cycling - application to the study of hydrogen-exchange in proteins. J Magn Reson 85(2), 393–399 (1989)

Schmidt, E., Güntert, P.: A new algorithm for reliable and general NMR resonance assignment. J Am Chem Soc 134(30), 12817–12829 (2012)

Shaka, A.J., Keeler, J.: Broadband spin decoupling in isotropic liquids. Prog Nucl Magn Reson Spectrosc 19, 47–129 (1987)

Shaka, A.J., Lee, C.J., Pines, A.: Iterative schemes for bilinear operators - application to spin decoupling. J Magn Reson 77(2), 274–293 (1988)