

## Resolution-optimized NMR measurement of $^1D_{CH}$ , $^1D_{CC}$ and $^2D_{CH}$ residual dipolar couplings in nucleic acid bases

Jérôme Boisbouvier<sup>a,b</sup>, David L. Bryce<sup>a</sup>, Erin O'Neil-Cabello<sup>a,c</sup>, Edward P. Nikonowicz<sup>c</sup> & Ad Bax<sup>a,\*</sup>

<sup>a</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.; <sup>b</sup>Present address: Laboratoire de RMN, Institut de Biologie Structurale, Jean-Pierre Ebel, UMR 5075 CNRS-CEA-UJF, 41 rue Jules Horowitz, 38027 Grenoble, Cedex 1, France; <sup>c</sup>Department of Biochemistry and Cell Biology, Rice University, POB 1892, Houston, TX 77251, U.S.A.

Received 1 June 2004; Accepted 21 July 2004

**Key words:** carbon-13, cross-validation, DNA, heteronuclear NMR, residual dipolar coupling, RNA, resolution enhancement, TROSY

### Abstract

New methods are described for accurate measurement of multiple residual dipolar couplings in nucleic acid bases. The methods use TROSY-type pulse sequences for optimizing resolution and sensitivity, and rely on the E.COSY principle to measure the relatively small two-bond  $^2D_{CH}$  couplings at high precision. Measurements are demonstrated for a 24-nt stem-loop RNA sequence, uniformly enriched in  $^{13}C$ , and aligned in Pf1. The recently described pseudo-3D method is used to provide homonuclear  $^1H$ - $^1H$  decoupling, which minimizes cross-correlation effects and optimizes resolution. Up to seven  $^1H$ - $^{13}C$  and  $^{13}C$ - $^{13}C$  couplings are measured for pyrimidines (U and C), including  $^1D_{C5H5}$ ,  $^1D_{C6H6}$ ,  $^2D_{C5H6}$ ,  $^2D_{C6H5}$ ,  $^1D_{C5C4}$ ,  $^1D_{C5C6}$ , and  $^2D_{C4H5}$ . For adenine, four base couplings ( $^1D_{C2H2}$ ,  $^1D_{C8H8}$ ,  $^1D_{C4C5}$ , and  $^1D_{C5C6}$ ) are readily measured whereas for guanine only three couplings are accessible at high relative accuracy ( $^1D_{C8H8}$ ,  $^1D_{C4C5}$ , and  $^1D_{C5C6}$ ). Only three dipolar couplings are linearly independent in planar structures such as nucleic acid bases, permitting cross validation of the data and evaluation of their accuracies. For the vast majority of dipolar couplings, the error is found to be less than  $\pm 3\%$  of their possible range, indicating that the measurement accuracy is not limiting when using these couplings as restraints in structure calculations. Reported isotropic values of the one- and two-bond J couplings cluster very tightly for each type of nucleotide.

### Introduction

NMR spectroscopy provides the opportunity to study the structure and dynamics of oligonucleotides as large as 30 kD (Wu et al., 2001; Cabello-Villegas et al., 2002; Lawrence et al., 2003; Leeper et al., 2003; Lukavsky et al., 2003; D'Souza et al., 2004). Traditionally, many of the structural restraints have relied on quantitative interpretation of  $^1H$ - $^1H$  NOEs, supplemented by  $^3J_{HH}$  and  $^3J_{HP}$  couplings. However, with

the introduction of isotopic enrichment procedures for RNA and DNA (Batey et al., 1992; Nikonowicz et al., 1992; Ono et al., 1994; Farmer et al., 1995; Zimmer and Crothers, 1995; Masse et al., 1998) and methods for weakly aligning nucleic acids relative to the magnetic field (Kung et al., 1995; Tjandra and Bax, 1997; Clore et al., 1998; Hansen et al., 1998; Ruckert and Otting, 2000; Sass et al., 2000; Tycko et al., 2000; Chou et al., 2001; Ishii et al., 2001; Meier et al., 2002; Ulmer et al., 2003), residual dipolar coupling (RDC) restraints are now also available. RDCs are particularly useful in nucleic acids, where the number of NOE restraints is typically relatively low, especially

\*To whom correspondence should be addressed. E-mail: bax@nih.gov

those involving long-range contacts. The fact that dipolar restraints provide information on internuclear bond vectors relative to a single axis system, that of the alignment tensor, allows for the study of global properties such as helix bending or relative helix orientations in multi-subunit molecules (Tjandra et al., 2000; Vermeulen et al., 2000; Al-Hashimi et al., 2002; Bondensgaard et al., 2002; MacDonald and Lu, 2002; Barbic et al., 2003; Lukavsky et al., 2003; Stefl et al., 2004). However, most applications have relied on a relatively small number of couplings per nucleotide, far fewer than the number of variable torsion angles.

We are currently engaged in an effort to evaluate which RDCs are most easily measured in DNA and RNA oligomers, and at what level of accuracy the different types of couplings can be measured. Here, we describe methods for measurement of couplings in nucleic acid bases. Measurements are demonstrated for a 24-nt stem-loop RNA sequence, uniformly enriched in  $^{13}\text{C}$ , and aligned in Pf1. For planar systems, at most three couplings carry linearly independent information (Zidek et al., 2003; Bryce and Bax, 2004). However, the measurement of a larger number allows for the direct evaluation of their accuracy. For uridine and cytidine we report methods for measurement of up to seven couplings; for adenine and guanine, the corresponding numbers (4 and 3, respectively) are lower, but nevertheless sufficient to define the base orientation. Availability of three or more dipolar couplings for a planar structure, such as a nucleic acid base, also has been shown useful for obtaining an improved estimate of the alignment tensor magnitude and rhombicity (Zidek et al., 2003; Bryce and Bax, 2004).

A number of alternate methods have been described previously for the measurement of base dipolar couplings, including schemes based on multiple quantum coherence (Yan et al., 2002) and spin-state selective excitation ( $\text{S}^3\text{E}$ ) (Zidek et al., 2001). Here, we focus on the types of couplings that provide the highest normalized accuracy. Optimal results are obtained with TROSY-based pulse schemes (Brutscher et al., 1998; Pervushin et al., 1998; Fiala et al., 2000). Not only do these methods yield RDCs which provide excellent cross-validation (rms errors less than 3% of their potential range), they also yield very narrow bands for the isotropic values of the various J couplings. When fitting the dipolar couplings observed in the stem region of the sequence to a canonical A-form helix, the residual in the fit considerably exceeds the measurement error, indicating that the minor,

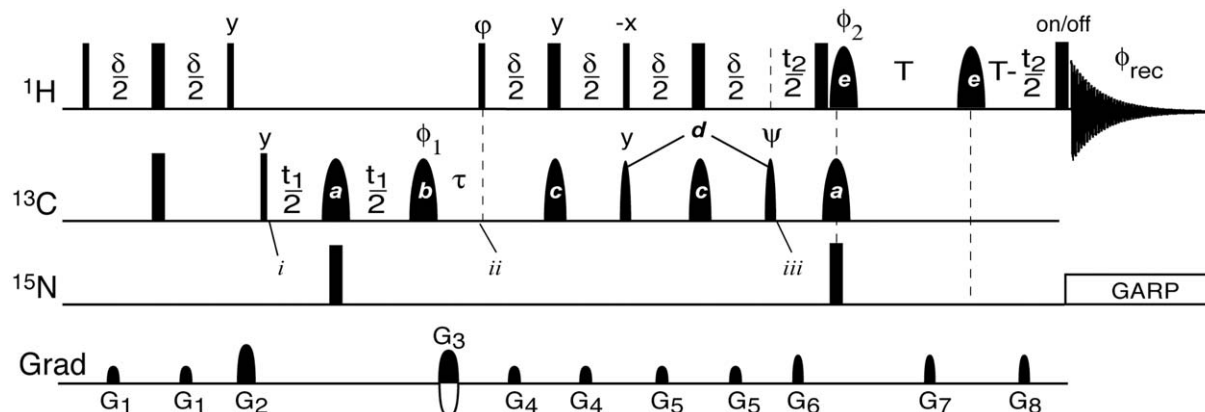
sequence-dependent deviations from the A-form helix indeed are observable by these dipolar couplings.

## Experimental section

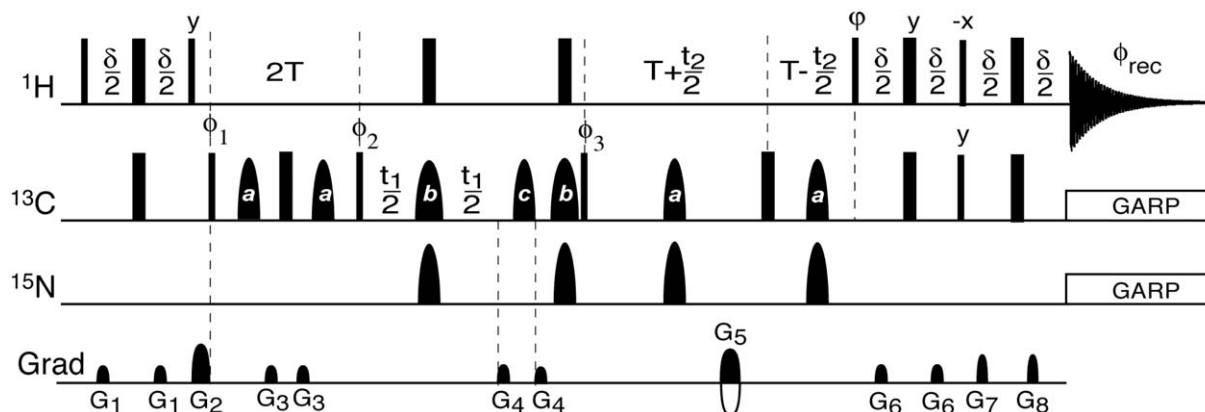
Two samples, each containing 1.9 mM of a uniformly  $^{13}\text{C}$ -enriched RNA oligomer derived from helix-35 of *E. coli* 23S ribosomal RNA, and modified to contain pseudouridine ( $\psi$ ) at position 746 were prepared in thin-wall Shigemi microcells (270  $\mu\text{L}$ ). The nucleotide sequence is: GGGCUAAUG $\psi$ UGAAAAAUAGCCC. In the numbering used in this paper, 700 is subtracted, and the numbering runs from G37 through C60. As a result of the procedure used to generate the oligomer (Nikonowicz et al., 1992), it lacked  $^{13}\text{C}$  enrichment for the 5'-terminal nucleotide, G37. Both samples were extensively dialyzed against the same buffer, which contained 10 mM NaCl, 10 mM potassium phosphate, and 0.02 mM EDTA in 99%  $\text{D}_2\text{O}$ , at pH 6.8 (direct meter reading in  $\text{D}_2\text{O}$ ). One of the two samples additionally contained 22 mg/mL Pf1, which serves as the liquid crystalline alignment medium (Hansen et al., 1998). The isotropic sample also was uniformly enriched in  $^{15}\text{N}$ . Using  $^{15}\text{N}$  decoupling, as indicated in the pulse schemes, all described methods work equally well in the presence or absence of  $^{15}\text{N}$  enrichment. The  $^2\text{H}$  lock solvent quadrupole splitting was  $19.24 \pm 0.04$  Hz for all experiments carried out in this study, with no discernable difference between measurements taken at the different magnetic field strengths (14.1 and 17.6 T) used. All the NMR data were collected at 25  $^\circ\text{C}$ .

Spectra related to measurement of purine  $^1\text{D}_{\text{C}_5\text{C}_6}$  and  $^1\text{D}_{\text{C}_5\text{C}_4}$  couplings were recorded at 600 MHz  $^1\text{H}$  frequency, where the reasonably favorable relaxation properties of the quaternary  $\text{C}_4$ ,  $\text{C}_5$ , and  $\text{C}_6$  resonances ( $T_2 \approx 60$  ms), dominated by their chemical shift anisotropy (CSA), were advantageous. All other measurements were carried out at 750 MHz, in order to optimize spectral resolution. Spectra at 600 MHz were recorded on a Bruker DRX600 spectrometer, equipped with a cryogenic probehead and a self-shielded z-gradient accessory. Spectra at 750 MHz were recorded on a Bruker DMX750 spectrometer, equipped with a 3-axis self-shielded gradient accessory.

NMR data were processed with nmrPipe (Delaglio et al., 1995). For all spectra acquired with the pulse schemes of Figures 1 and 2, the data in the indirect  $^{13}\text{C}$   $t_1$  dimension  $F_1$  were apodized by a sine-bell



**Figure 1.** Pulse scheme for homonuclear-decoupled HC(C) hd-TROSY-ECOSY experiments for  $^1\text{D}_{\text{CH}}$   $^1\text{D}_{\text{CC}}$  and  $^2\text{D}_{\text{CH}}$  RDC measurements in pyrimidine bases. Narrow and wide bars indicate non-selective  $90^\circ$  and  $180^\circ$  pulses, respectively. Unless specified, pulse phases are  $x$ . Delay durations:  $\delta = 1/(2J_{\text{CH}})$ ;  $T=18$  ms, the delay  $\tau$  is equal to the duration of shaped pulse 'a' and includes the duration of  $G_3$  and its recovery period. For selection of the TROSY ( $\text{C}^{(\beta)}$ ,  $\text{H}^{(\alpha)}$ )-component ( $^{13}\text{C}$  downfield,  $^1\text{H}$  upfield):  $\phi = y$ ;  $\psi = x$ ;  $\phi_1 = x, y$ ;  $\phi_2 = x, x, y, y$ ;  $\phi_{\text{rec}} = (x, -x, -x, x)$ . Pulse phases apply for Bruker spectrometers; for Varian spectrometers  $y$  and  $-y$  should be interchanged. The experiment is recorded in a double echo anti-echo manner: for each  $t_1$  increment, two FIDs are acquired, one with  $G_3$ ,  $\phi$  and  $\psi$  inverted, and stored separately (Weigelt, 1998); for each  $t_2$  increment two FIDs are acquired with the insertion on alternate scans of the extra  $^1\text{H}$   $180^\circ$  pulse just before acquisition (Boisbouvier et al., 2003). Data are processed in both dimensions using a standard 'echo-antiecho' Fourier transform processing method (Bachmann et al., 1977; Palmer et al., 1991; Kay et al., 1992). Field gradients are sine-bell shaped with durations  $G_{1,\dots,8}$  of 0.3, 1, 1, 0.2, 0.2, 0.125, 0.125, 0.125 ms, and amplitudes of 8, 6, 30, 10, 14, 9, 27, 24 G/cm, and directions  $(x, y, z)$ ,  $(x, -y, z)$ ,  $z$ ,  $(x, -y, -z)$ ,  $(-x, -y, z)$ ,  $z$ ,  $z$ ,  $-z$ .  $^{15}\text{N}$  decoupling during acquisition is applied using a GARP decoupling sequence with  $\gamma B_1/2\pi = 1.1$  kHz. Pulse shapes, frequencies, and durations are given in Table 2. For measurement of  $\text{C}_5\text{H}_5$  or  $\text{C}_6\text{H}_6$  couplings in the  $^1\text{H}$  dimension ( $F_2$ ), two spectra are acquired in interleaved mode, one with  $\psi$  inverted. Corresponding Bruker pulse programs and NMRpipe data processing macros are available at <http://spin.niddk.nih.gov/bax/>



**Figure 2.** Pulse scheme of the HCC(C)-TROSY experiments for  $^1\text{D}_{\text{CC}}$  RDC measurements in purine bases. Narrow and wide bars indicate non-selective  $90^\circ$  and  $180^\circ$  pulses, respectively. Unless specified, pulse phases are  $x$ . Delay duration:  $\delta = 1/(2J_{\text{CH}})$ . For selection of the TROSY (downfield)  $^{13}\text{C}^{(\beta)}$  doublet component:  $\phi = y$ ;  $\phi_1 = x, -x$ ;  $\phi_2 = y, y, -y, -y$ ;  $\phi_3 = y, y, y, y, -y, -y, -y, -y$ ;  $\phi_{\text{rec}} = x, -x$ . Pulse phases apply for Bruker spectrometers; for Varian spectrometers  $y$  and  $-y$  should be interchanged. Quadrature detection in  $t_1$  is achieved by phase incrementation of  $\phi_1$  in the usual States-TPPI manner; quadrature detection in  $t_2$  by the usual echo anti-echo manner: for each  $t_2$  increment, two FIDs are acquired, one with  $G_5$  and  $\phi$  inverted, and stored separately (Weigelt, 1998). Field gradients are sine-bell shaped  $z$ -gradients with durations  $G_{1,\dots,8}$  of 0.7, 1, 0.7, 0.7, 2, 0.7, 0.503, 0.503 ms, and amplitudes of 14, 8, -10, 22, -30, 17, -4, 26 G/cm. The shaped pulse on the  $^{15}\text{N}$  channel is a 2-ms adiabatic hyperbolic secant inversion pulse applied at 155 ppm with an inversion bandwidth of  $\pm 80$  ppm.  $^{15}\text{N}$  and  $^{13}\text{C}$  decoupling during acquisition are applied using a GARP sequence with  $\gamma B_1/2\pi$  equal to 1.1 kHz and 3.6 kHz respectively. Pulse shapes, frequencies and durations are listed in Table 3. Corresponding Bruker pulse programs and NMRpipe macros are available at <http://spin.niddk.nih.gov/bax/>

function shifted by  $67^\circ$ , and the data in the directly detected  $^1\text{H}$  dimension  $F_3$  were apodized by a  $90^\circ$ -shifted squared sine-bell function, followed by zero filling at least two-fold in each dimension. For the ‘pseudo-3D’ spectra, recorded with homonuclear  $^1\text{H}$ - $^1\text{H}$  decoupling (hd) during the  $t_2$  constant-time evolution period (Figure 1), mirror-image linear prediction (Zhu and Bax, 1990) using 4 coefficients was used to double the number of time-domain points from 6 to 12 (after FT in the orthogonal dimensions had been carried out (Delaglio et al., 1995)), followed by apodization with a  $90^\circ$ -shifted squared sine-bell function, zero filling, and projection in the manner described previously (Boisbouvier et al., 2003), to yield a projected two-dimensional spectrum with digital resolutions of 4.8 Hz ( $F_1$ ) and 4.3 Hz ( $F_2$ ) for the experiments involving evolution of C5 magnetization during  $t_1$  (7.8 Hz and 4.3 Hz, respectively for the H6C6(C5) hd-TROSY-ECOSY experiment).

Doubling of the duration of the constant-time dimension ( $t_2$ ) by mirror-image linear prediction (using 4 to 6 coefficients) was also used for the data collected with the pulse scheme of Figure 2, followed by apodization with a  $90^\circ$ -shifted squared sine-bell function (such that the predicted data are strongly attenuated by the apodization), and zero filling, to yield 3D spectra with digital resolutions ranging from 1.0 to 2.8 Hz in  $F_1$ , from 5.2 to 10.8 Hz in  $F_2$ , and 7.8 Hz ( $F_3$ ).

An A-form helical model was generated for the terminal eight basepairs of the RNA molecule using the Biopolymer module of the program Insight II (Molecular Simulations, Inc.).

## Results and discussion

In this study, we use separate 3D pulse schemes for the measurement of pyrimidine and purine base RDCs. Particularly in the aligned state, the  $^1\text{H}$  multiplet structures of uridine and cytidine  $\text{H}_6$  and  $\text{H}_5$  resonances tend to be broad as a result of relatively large  $^1\text{H}$ - $^1\text{H}$  couplings. In the absence of homonuclear decoupling, this broad multiplet structure in E.COSY experiments (Griesinger et al., 1987) that measure passive splittings in the  $^1\text{H}$  dimension limits the precision of such measurements. The problem is compounded by cross-correlated relaxation effects, which introduce different intensities for the individual  $^1\text{H}$  multiplet components and can give rise to substantial systematic errors in the coupling measurement if these  $^1\text{H}$ - $^1\text{H}$  splittings are unresolved (see inset in Figure 4). The problem is par-

ticularly acute when attempting to measure one-bond  $^{13}\text{C}_5$ - $^1\text{H}_5$  couplings in the  $^1\text{H}$  dimension of a  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum, in the absence of  $^{13}\text{C}$  decoupling. Measurement of the  $^{13}\text{C}_5$ - $^1\text{H}_5$  splitting in the  $^{13}\text{C}$  dimension tends to be problematic as well because of the strong relaxation interference between  $^{13}\text{C}_5$  CSA and the  $^{13}\text{C}_5$ - $^1\text{H}_5$  dipolar interaction. Rapid transverse relaxation of the  $^{13}\text{C}_5$ - $\{^1\text{H}_5\}$  anti-TROSY component also adversely affects E.COSY experiments that measure  $^2\text{J}_{\text{C}_6\text{H}_5} + ^2\text{D}_{\text{C}_6\text{H}_5}$  couplings in the  $^{13}\text{C}$  dimension from  $\text{C}_5$ - $\text{C}_6$  cross peaks, where  $\text{H}_5$  is the passive spin, and such experiments were therefore not used. Instead, we rely on a recently proposed homonuclear decoupling scheme (Boisbouvier et al., 2003) to provide high resolution in the  $^1\text{H}$  dimension, and to eliminate the effect of cross-correlated relaxation (Figure 1). For purines, homonuclear  $^1\text{H}$ - $^1\text{H}$  couplings are less of a problem than for pyrimidines and, at least for  $\text{C}_8$ , relaxation interference between  $^{13}\text{C}_8$  CSA and the  $^{13}\text{C}_8$ - $^1\text{H}_8$  dipolar interaction is only moderate. In addition to the one bond  $^1\text{H}$ - $^{13}\text{C}$  couplings, measurements in purines focus on the  $^{13}\text{C}$ - $^{13}\text{C}$  couplings between  $\text{C}_5$  and  $\text{C}_4$ , and between  $\text{C}_5$  and  $\text{C}_6$ . For these measurements, the pulse scheme of Figure 2 is used. A brief discussion of the two types of pulse schemes is presented below.

### Measurement of couplings in pyrimidines

The pulse scheme of Figure 1 starts with magnetization transfer by the regular INEPT scheme from  $^1\text{H}$  to  $^{13}\text{C}$ . No phase alternation of the  $90^\circ_y$  ( $^1\text{H}$ ) pulse is used, such that for the downfield TROSY doublet component the  $^{13}\text{C}$  Boltzmann magnetization co-adds to the INEPT component (Pervushin et al., 1998). Depending on the frequencies at which the shaped pulses,  $a - e$ , are applied, the same pulse scheme can be used for magnetization that starts on  $\text{H}_5$  or  $\text{H}_6$ , and different nuclei are selected for decoupling to yield optimal resolution and sensitivity of the resonances corresponding to the interactions of interest. For example, the experiment can be optimized for measurement of the  $^{13}\text{C}_5$ - $^{13}\text{C}_4$  coupling, and simultaneous E.COSY measurement of the  $^{13}\text{C}_4$ - $^1\text{H}_5$  coupling, and is of the ‘out-and-back’ type (Ikura et al., 1990). As the experiment utilizes the homonuclear decoupling (hd) and TROSY elements, the experiment is referred to as H5C5(C4) hd-TROSY-ECOSY. The analogous experiment for measuring  $^{13}\text{C}_5$ - $^{13}\text{C}_6$  and  $^{13}\text{C}_6$ - $^1\text{H}_5$  couplings, utilizing different frequencies for the shaped pulses, is referred to as H5C5(C6) hd-

TROSY-ECOSY, where the nucleus between brackets refers to the passive spin used to generate the E.COSY displacement. Below, for brevity, we will drop the ‘hd-TROSY-ECOSY’ part of the experiment name.

In the H5C5(C4) experiment, the slowly relaxing downfield C<sub>5</sub> transverse magnetization, present at time point *i*, evolves for a period  $t_1$ . The shaped pulse, labeled *a*, is of the 180° selective inversion IBURP2 type (Geen and Freeman, 1991), and is applied to C<sub>6</sub> to remove any dephasing resulting from  $^1J_{C_5C_6}$ . The subsequent refocusing REBURP pulse, marked *b*, and delay  $\tau$  compensate for C<sub>5</sub> chemical shift evolution that occurred during pulse *a*, and allow time for pulsed field gradient phase-encoding, enabling gradient-enhanced magnetization transfer from C<sub>5</sub> (at time *ii*) to H<sub>5</sub> (at time *iii*). During the subsequent constant-time  $^1H$  evolution period, of total duration  $2T = 36$  ms, interactions with  $^{13}C_6$  as well as  $^{15}N$  are removed by application of the shaped pulse labeled *a* and the non-selective 180°  $^{15}N$  pulse. Pulses *c* and *d* are applied selectively to C<sub>5</sub>, such that the effect of  $^{13}C$ - $^{13}C$  dephasing during the  $\delta$  delays is eliminated and the spin-state of the passive carbon, C<sub>4</sub>, is preserved. Note that a very narrow spectral window is used in the  $t_2$  dimension (139 Hz) and that the resulting aliasing is undone in the subsequent unfolding procedure, where the dimensionality is reduced from three to two (Boisbouvier et al., 2003).

Figure 3A presents a small region of the unfolded, projected H5C5(C4) spectrum, recorded for the aligned RNA sample. The vertical splittings ( $^{13}C$  dimension) correspond to  $^1J_{C_5C_4} + ^1D_{C_5C_4}$  and the horizontal displacement ( $^1H$  dimension) to  $^2J_{C_4H_5} + ^2D_{C_4H_5}$ . All couplings extracted from the spectrum recorded on the isotropic sample are available as Supporting Information; average values are listed in Table 1, together with their very small standard deviations.

By switching the frequencies at which pulses, labeled *a*, are applied from C<sub>6</sub> (147.4 ppm) to C<sub>4</sub> (163.9 ppm), an H5C5(C6) spectrum can be recorded in a manner fully analogous to that described above. In the  $^{13}C$  dimension, this spectrum displays the  $^1J_{C_5C_6} + ^1D_{C_5C_6}$  splitting, and in the  $^1H$  dimension, the  $^2J_{C_6H_5} + ^2D_{C_6H_5}$  splitting (Figure 3B). Again, under isotropic conditions, measured J couplings cluster in very narrow regions (Table 1).

Clearly, simply switching the frequencies of the shaped pulses, in the manner indicated in Table 2, also allows recording of H6C6(C5) spectra (Figure 3C), albeit at considerably worse resolution and sensitivity

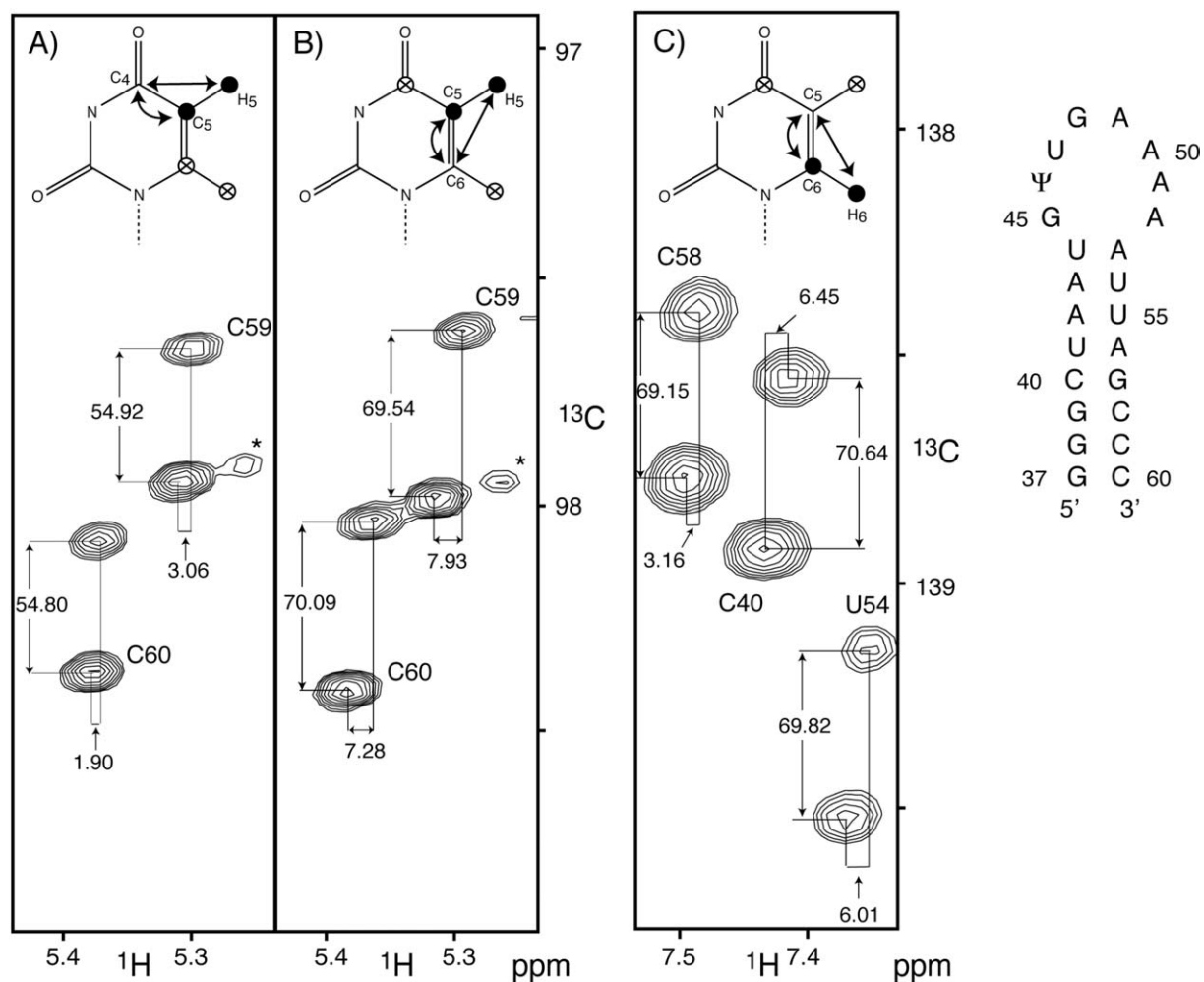
due to the less advantageous TROSY properties of C<sub>6</sub> compared to C<sub>5</sub>. The same  $^1J_{C_5C_6} + ^1D_{C_5C_6}$  splitting is available from both the H6C6(C5) spectrum and the H5C5(C6) spectrum. The close agreement between these independent measurements of  $^1D_{C_5C_6}$  (rmsd < 0.3 Hz, dominated by the larger uncertainty in the H6C6(C5)-derived coupling) testify to the reliability of these measurements.

The duration of pulse *a* in Figure 1 can also be adjusted to cover both C<sub>4</sub> and C<sub>6</sub> resonances simultaneously, resulting simply in the H5C5 hd-TROSY spectrum. Note that the H<sub>5</sub> frequency occurs in both the F<sub>2</sub> and F<sub>3</sub> dimensions of the 3D spectrum, once in the presence of homonuclear  $^1H$ - $^1H$  and heteronuclear  $^{13}C$  and  $^{15}N$  decoupling (F<sub>2</sub>), and in the detected dimension without such decoupling (F<sub>3</sub>). The coarse F<sub>3</sub> resolution is sufficient to unfold, in an automated manner, the spectrum that has been acquired at high resolution in the F<sub>2</sub> dimension, using only six complex data points (Boisbouvier et al., 2003). By altering the phase  $\psi$  during the TROSY transfer (Weigelt, 1998), either the downfield or upfield  $^{13}C_5$ - $^1H_5$  component can be selected (Figure 4). However, it should be noted that in order to minimize spectral overlap and to optimize sensitivity, the spectra containing the downfield and upfield components were recorded separately, in an interleaved manner, and the superposition is for display purposes only.

#### Measurement of couplings in purines

The absence of one-bond  $^{13}C$ - $^{13}C$  interactions involving C<sub>8</sub> or C<sub>2</sub> in purines allows for relatively straightforward measurement of the one-bond  $^1H$ - $^{13}C$  coupling for these sites from either a  $^1H$ - $^{13}C$  HSQC spectrum, recorded without  $^1H$  decoupling in the  $^{13}C$  dimension, or an IPAP version of this experiment (Ottiger et al., 1998). The latter method was used in the present study. For C<sub>8</sub>, relaxation interference between  $^{13}C$  CSA and  $^1H$ - $^{13}C$  dipolar coupling is moderate, resulting in acceptable resolution even for the broader upfield  $^{13}C$  doublet component. For C<sub>2</sub>, relaxation interference between the  $^{13}C$ - $^1H$  dipolar interaction and the  $^{13}C$  CSA is more severe. Therefore, it can be advantageous to measure the  $^1H$ - $^{13}C$  coupling in the  $^1H$  dimension of the spectrum, particularly when the sample is dissolved in D<sub>2</sub>O and the  $^1H$  density in the vicinity of adenine H<sub>2</sub> is low.

Measurement of dipolar couplings in purines other than the one-bond  $^{13}C$ - $^1H$  interactions can be considerably more challenging due to the inherently much



**Figure 3.** Small sections from the projected HC(C) hd-TROSY-ECOSY spectra, used for measurement of  $^1D_{CC}$  and  $^2D_{CH}$  RDC in pyrimidine bases. Spectra for the aligned 24-nt RNA sample were recorded at 750 MHz, 25 °C, with total measuring times of 12 h each, using the pulse scheme of Figure 1. The  $C_5H_5$  (panels A and B) and  $C_6H_6$  (panel C) TROSY components ( $^{13}C$  downfield,  $^1H$  upfield) appear as ECOSY doublets, due to the passive coupling with their adjacent  $^{13}C$ . Shown are examples for (A)  $C_4$ , (B)  $C_6$ , and (C)  $C_5$ . Shape *a* (Table 2) is adjusted to decouple the interaction with  $C_6$  (A), or  $C_4$  (B, C). Correlations marked by asterisks correspond to impurities. Identical acquisition and processing parameters were used for spectra (A) and (B): time domain matrices of  $256^* \times 6^* \times 256^*$  data points, with acquisition times of 102 ms ( $t_1$ ), 36 ms ( $t_2$ ) and 57 ms ( $t_3$ ). For the  $H_6C_6(C_5)$  hd-TROSY-ECOSY experiment, the acquisition time in the carbon dimension has been adjusted to 64 ms (100 complex increments) to take into account the faster relaxation rate of the  $C_6$  TROSY component compared to the favorable TROSY relaxation rates for  $C_5$ . After zero-filling and Fourier transformation, the 3D matrix has been projected along  $F_3$  to form a  $1024 \times 512$  ( $F_1, F_2$ ) 2D matrix for panels A and B,  $1024 \times 256$  for panel C, using the processing scheme described by Boisbouvier et al. (2003). Spectra are plotted at identical contour levels. Insets illustrate the spins for which the frequency is measured ( $\bullet$ ), the  $^1H$  and  $^{13}C$  spins decoupled ( $\times$ ), and the splittings observed ( $\leftrightarrow$ ). Correlations are labeled by residue number and measured splittings are marked in Hz: (A), ( $C_5C_4$ ,  $H_5C_4$ ); (B), ( $C_5C_6$ ,  $H_5C_6$ ); (C), ( $C_6C_5$ ,  $H_6C_5$ ).

smaller size of such interactions. Here, we focus on the measurement of  $C_4$ - $C_5$  and  $C_5$ - $C_6$  couplings, which offer several advantages relative to other potentially accessible weaker couplings, such as two-bond  $^1H$ - $^{15}N$  and one-bond  $^{13}C$ - $^{15}N$  interactions. First, the orientational distribution of the  $C_4$ - $C_5$  and  $C_5$ - $C_6$  vectors relative to the  $^{13}C$ - $^1H$  vectors is favorable. Second,

the relaxation properties of the quaternary  $C_4$ ,  $C_5$  and  $C_6$  carbons in purines are favorable, resulting in relatively long  $T_2$  values of *ca.* 60 ms (for a rotational correlation time of *ca.* 5 ns). Thirdly, the relatively short C-C distances for these pairs results in dipolar interaction constants that are only 8.3 times lower than for a one-bond  $^{13}C$ - $^1H$  pair (Figure 5). Potential chal-

Table 1. Experimental  $^1J_{CH}$ ,  $^1J_{CC}$  and  $^2J_{CH}$  spin-spin coupling constants in the 24-nucleotide RNA sample<sup>a</sup>.

Coupling	Average	Experiment	Precision <sup>a</sup>
Uridine (5 bases)			
C <sub>6</sub> C <sub>5</sub>	66.10 ± 0.26	H5C5(C6) hd-TROSY-ECOSY	(0.07)
C <sub>6</sub> H <sub>5</sub>	4.25 ± 0.21	H5C5(C6) hd-TROSY-ECOSY	(0.09)
H <sub>6</sub> C <sub>5</sub>	2.30 ± 0.23	H6C6(C5) hd-TROSY-ECOSY	(0.11)
H <sub>5</sub> C <sub>4</sub>	1.35 ± 0.07	H5C5(C4) hd-TROSY-ECOSY	(0.09)
C <sub>5</sub> C <sub>4</sub>	64.36 ± 0.26	H5C5(C4) hd-TROSY-ECOSY	(0.09)
C <sub>5</sub> H <sub>5</sub>	176.47 ± 0.35	H5C5 hd-TROSY	(0.05)
C <sub>6</sub> H <sub>6</sub>	180.17 ± 0.36	H6C6 hd-TROSY	(0.18)
PseudoUridine (1 base)			
C <sub>6</sub> C <sub>5</sub>	71.78	H6C6(C5) hd-TROSY-ECOSY	(0.18)
H <sub>6</sub> C <sub>5</sub>	0.90	H6C6(C5) hd-TROSY-ECOSY	(0.11)
C <sub>6</sub> H <sub>6</sub>	178.62	H6C6 hd-TROSY	(0.25)
Cytosine (4 bases)			
C <sub>6</sub> C <sub>5</sub>	67.36 ± 0.13	H5C5(C6) hd-TROSY-ECOSY	(0.09)
C <sub>6</sub> H <sub>5</sub>	4.29 ± 0.07	H5C5(C6) hd-TROSY-ECOSY	(0.11)
H <sub>6</sub> C <sub>5</sub>	2.96 ± 0.17	H6C6(C5) hd-TROSY-ECOSY	(0.11)
H <sub>5</sub> C <sub>4</sub>	1.63 ± 0.19	H5C5(C4) hd-TROSY-ECOSY	(0.11)
C <sub>5</sub> C <sub>4</sub>	54.65 ± 0.13	H5C5(C4) hd-TROSY-ECOSY	(0.10)
C <sub>5</sub> H <sub>5</sub>	173.68 ± 0.25	H5C5 hd-TROSY	(0.05)
C <sub>6</sub> H <sub>6</sub>	179.29 ± 0.33	H6C6 hd-TROSY	(0.15)
Adenine (7 bases) <sup>c</sup>			
C <sub>6</sub> C <sub>5</sub>	74.71 ± 0.49	H2C2C5(C6)-, H8C8C6(C5)-TROSY	(0.10)
C <sub>5</sub> C <sub>4</sub>	65.12 ± 0.61	H2C2C5(C4)-, H8C8C4(C5)-TROSY	(0.07)
C <sub>2</sub> H <sub>2</sub> <sup>b</sup>	200.27 ± 0.58	IPAP-HSQC	(0.19)
C <sub>8</sub> H <sub>8</sub> <sup>b</sup>	214.41 ± 0.66	IPAP-HSQC	(0.19)
Guanine (3 bases) <sup>d</sup>			
C <sub>6</sub> C <sub>5</sub>	85.98 ± 0.33	H8C8C6(C5) TROSY	(0.17)
C <sub>5</sub> C <sub>4</sub>	63.61 ± 0.35	H8C8C4(C5) TROSY	(0.11)
C <sub>8</sub> H <sub>8</sub> <sup>b</sup>	214.85 ± 0.61	IPAP-HSQC	(0.47)

<sup>a</sup>For each base, the average value and the standard deviation of the measured J splitting, the experiment used for the measurement, and the corresponding experimental precision are given. Pyrimidine couplings have been measured at  $B_0 = 17.6$  T, and purine data at  $B_0 = 14.1$  T; no correction for magnetic field induced alignment is included.

<sup>b</sup>measured with an IPAP-HSQC experiment (Ottiger et al., 1998).

<sup>c</sup>No precise measurements could be made for A52 due to conformational exchange broadening and partial overlap.

<sup>d</sup>Outside the stem region, G-H<sub>8</sub> resonances had exchanged with solvent deuterons and were vanishingly weak.

lenges, however, involve the proximity between the spectral regions in which the C<sub>4</sub> (149–154 ppm) and C<sub>6</sub> (157–161 ppm) resonances are found.

Provided spectral resolution is sufficient, we find the most reliable measurements are those starting from H<sub>8</sub> (Figure 2). In the first step, magnetization is trans-

ferred from H<sub>8</sub> to C<sub>8</sub>. The second 90° <sup>1</sup>H pulse is again applied with phase y, such that the Boltzmann <sup>13</sup>C magnetization adds to the downfield <sup>13</sup>C doublet component. In these H8C8C6(C5) and H8C8C4(C5) experiments, the upfield doublet component rapidly decays during the subsequent delay, 2T = 40 or 34 ms,

Table 2. Parameters for the shaped pulses used in the HC(C) hd-TROSY-ECOSY pulse scheme presented in Figure 1<sup>a</sup>

Experiment	H5C5(C4)	H5C5(C6)	H6C6(C5)	H6C6	H5C5
Splittings	C <sub>5</sub> C <sub>4</sub> , H <sub>5</sub> C <sub>4</sub>	C <sub>5</sub> C <sub>6</sub> , H <sub>5</sub> C <sub>6</sub>	C <sub>6</sub> C <sub>5</sub> , H <sub>6</sub> C <sub>5</sub>	H <sub>6</sub> C <sub>6</sub>	H <sub>5</sub> C <sub>5</sub>
Shape <i>a</i>	IBURP2	IBURP2	IBURP2	IBURP2	IBURP2
	1.0 ms	1.0 ms	1.0 ms	1.0 ms	0.65 ms
	147.4 ppm	163.9 ppm	163.9 ppm	100.9 ppm	154.9 ppm
Shape <i>b</i>	REBURP	REBURP	REBURP	REBURP	REBURP
	2.56 ms	2.56 ms	2.56 ms	2.56 ms	2.56 ms
	100.9 ppm	100.9 ppm	140.0 ppm	140.0 ppm	100.9 ppm
Shape <i>c</i>	REBURP	REBURP	REBURP	REBURP	REBURP
	1.60 ms	1.60 ms	1.60 ms	1.60 ms	1.60 ms
	100.9 ppm	100.9 ppm	140.0 ppm	140.0 ppm	100.9 ppm
Shape <i>d</i>	90°-Sinc	90°-Sinc	90°-Sinc	90°-Sinc	90°-Sinc
	0.27 ms	0.27 ms	0.27 ms	0.27 ms	0.27 ms
	100.9 ppm	100.9 ppm	140.0 ppm	140.0 ppm	100.9 ppm
Shape <i>e</i>	REBURP	REBURP	REBURP	REBURP	REBURP
	6.0 ms	6.0 ms	4.0 ms	4.0 ms	6.0 ms
	5.0 ppm	5.0 ppm	7.4 ppm	7.4 ppm	5.0 ppm

<sup>a</sup>For each shaped pulse, the shape form, the duration, and the frequency are indicated. Parameters are given for a spectrometer operating at  $B_0 = 17.6$  T. Each required pulse duration scales inversely with the strength of the magnetic field.

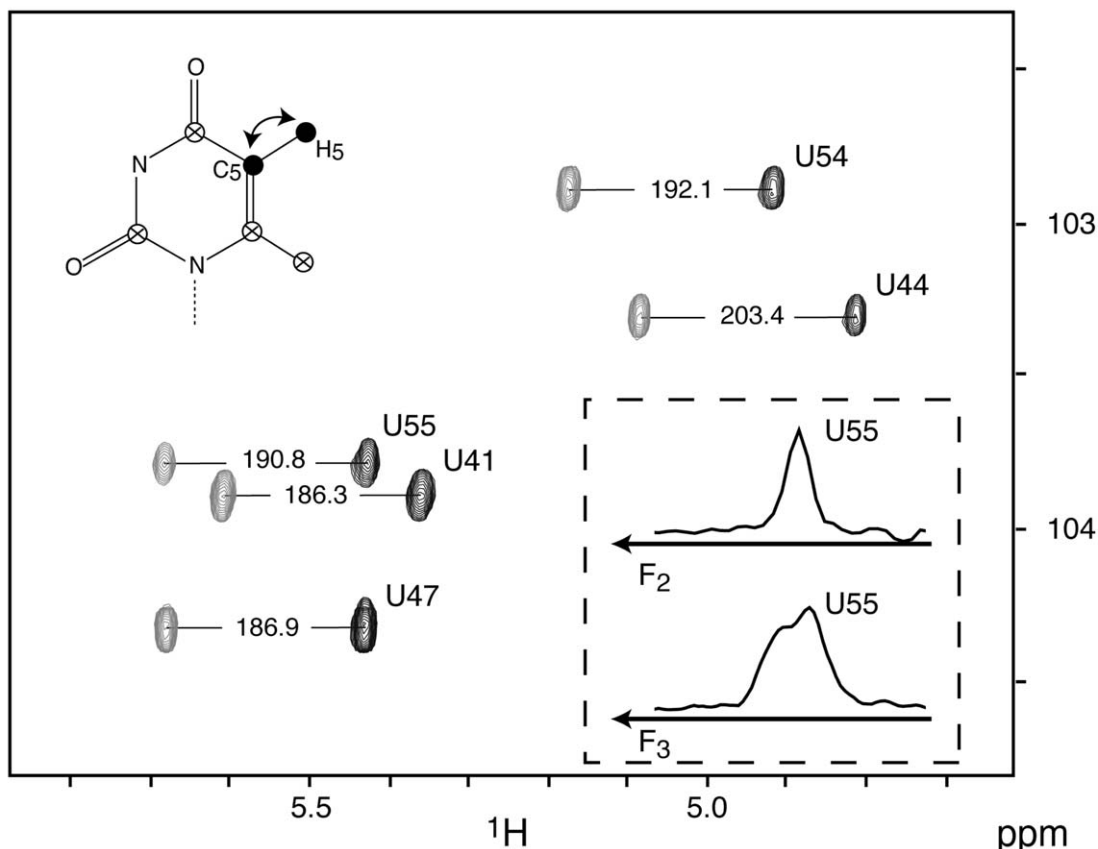
where C<sub>8</sub> magnetization becomes antiphase with respect to either C<sub>6</sub> or C<sub>4</sub>, depending on the frequency at which pulses marked ‘*a*’ are applied. The <sup>3</sup>J<sub>C<sub>8</sub>C<sub>6</sub> and <sup>2</sup>J<sub>C<sub>8</sub>C<sub>4</sub> couplings are relatively large, 8–12 Hz (Ippel et al., 1996), which, when combined with the favorable C<sub>8</sub> TROSY relaxation behavior, makes this process reasonably efficient. The subsequent 90°<sub>φ<sub>2</sub></sub> pulse transforms this magnetization into C<sub>6</sub> or C<sub>4</sub> magnetization, which then evolves for a duration t<sub>1</sub>, while coupled to C<sub>5</sub>. After the subsequent reverse pathway, and TROSY evolution during the constant-time evolution period 2T, magnetization is transferred back to H<sub>8</sub> and detected during t<sub>3</sub>. In the final 3D spectra, the one-bond <sup>1</sup>J<sub>C<sub>4</sub>C<sub>5</sub> (Figure 6A) or <sup>1</sup>J<sub>C<sub>6</sub>C<sub>5</sub> (Figure 6B) coupling constants can be measured with good accuracy. In the absence of linear prediction, the lower limit estimate for the random error in the measurement of the splitting approximately equals 0.7 × LW/SN, where LW is the line width and SN is the signal-to-noise ratio (Kontaxis et al., 2000). Note that this empirically derived relation is relatively insensitive to the type of apodization, but assumes identical phases for the two doublet components and the absence of potential systematic errors resulting from cross correlation and the like. With the moderate degree of linear prediction and</sub></sub></sub></sub>

subsequent strong apodization used in the <sup>1</sup>H dimension of the CT hd-TROSY-ECOSY experiments, LW in this dimension is reduced by about 35% relative to the spectrum derived without linear prediction, but the accuracy of the peak position in our experience does not improve significantly and can even worsen in cases where less strong apodization or more extensive linear prediction is used.

For the isotropic samples, the uncertainty in peak position estimated in the above manner corresponds to random errors in J that are smaller than 0.1 Hz for most purines (Table 1), and random errors of about 0.2 Hz in the dipolar contributions calculated from the difference in splitting between aligned and isotropic samples. Average measured isotropic J couplings for the purines are listed in Table 1. Accuracy of the measured RDCs, as determined by a ‘self consistency’ analysis (Zidek et al., 2003), will be shown to be comparable to the above estimate of the random error.

For several adenines in the stem-loop sequence studied, overlap of loop H<sub>8</sub>C<sub>8</sub>C<sub>4</sub> or H<sub>8</sub>C<sub>8</sub>C<sub>6</sub> correlations in the 3D spectrum prevents accurate measurement of the C<sub>6</sub>-C<sub>5</sub> and C<sub>4</sub>-C<sub>5</sub> couplings. In these cases, an alternate method can be used to measure these couplings, which transfers magnetization from





**Figure 4.**  $C_5H_5$  correlation spectra for uridine nucleotides in the aligned 24-nt RNA sample, obtained using the hd-TROSY experiment (Figure 1), for  $^1D_{CH}$  RDC measurement. The shaped pulse **a** is chosen to decouple both  $C_6$  and  $C_4$  nuclei (Table 2), and two spectra are acquired in interleaved mode, one with  $\psi$  inverted (Figure 1). The two spectra (red, downfield; black, upfield) are superimposed. Spectra are plotted at identical contour levels. Correlations are labeled by nucleotide number and measured J+D splittings are given in Hz. Spectra were recorded at 750 MHz, 25 °C, with total measuring times of 6 h each. Identical acquisition and processing parameters were used: time domain matrices of  $128^* \times 6^* \times 256^*$  data points, with acquisition times of 51 ms ( $t_1$ ), 36 ms ( $t_2$ ) and 57 ms ( $t_3$ ). After zero-filling and 3D Fourier transformation, the 3D matrix has been projected along  $F_3$  to form a  $1024 \times 256$  ( $F_1, F_2$ ) 2D matrix, using the processing scheme described by Boisbouvier et al. (2003). Insets show the spins selected ( $\bullet$ ), the decoupled  $^1H$  and  $^{13}C$  spins ( $\times$ ), and the observed splitting ( $\leftrightarrow$ ). Also shown are cross sections parallel to the  $F_2$  and  $F_3$  axes through the upfield U55- $H_5$  doublet component, illustrating the effect of  $^1H$ - $^1H$  decoupling. Note the asymmetry in the unresolved  $F_3$  doublet structure, resulting from cross-correlated relaxation.

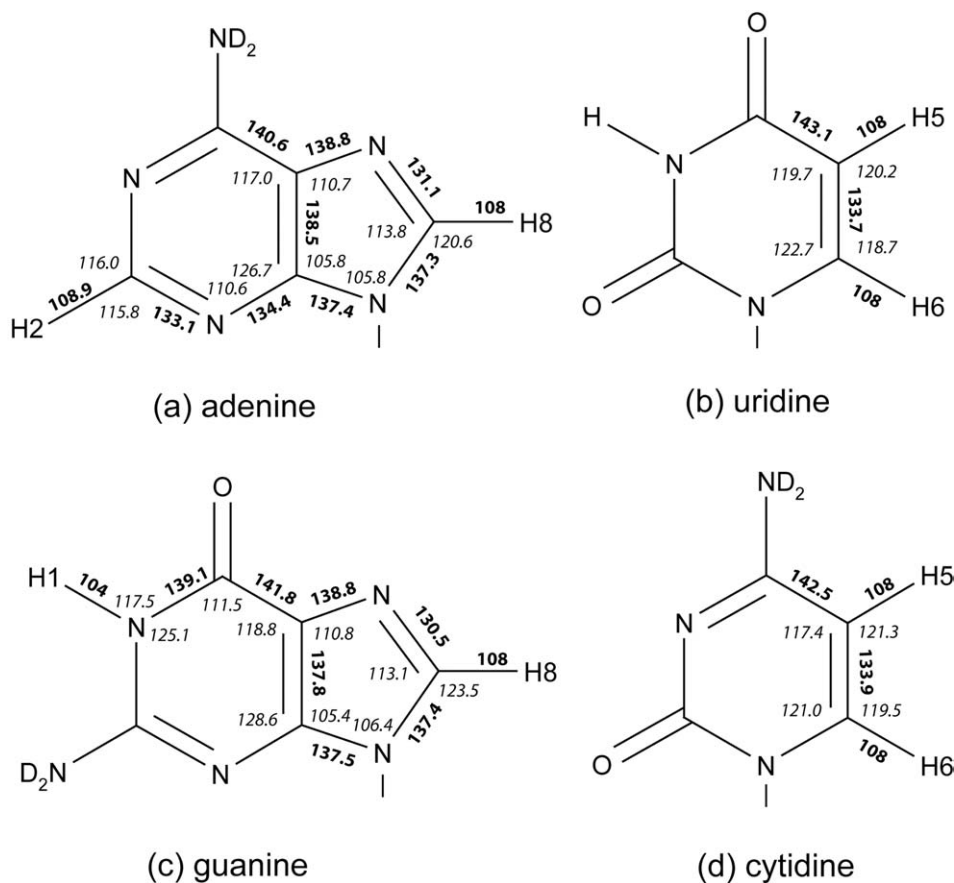
$H_2$  to  $C_2$  and on to  $C_5$  via the large (11.2 Hz)  $^3J_{C_2C_5}$  coupling. This experiment is carried out with the same pulse sequence of Figure 2, but the shaped pulses are applied at different frequencies (Table 3). During  $C_5$  evolution, either  $C_4$  or  $C_6$  is decoupled by means of an IBURP2 pulse (Table 3), covering a narrow bandwidth of only  $\pm 1.2$  ppm. Examples of small regions of cross sections through the 3D H2C2C5(C4) and H2C2C5(C6) spectra are shown in Figures 6C,D.

Note that due to the proximity of  $C_2$  to the  $C_4$  and  $C_6$  regions in the  $^{13}C$  spectrum, this type of measurement can generate small phase anomalies for nucleotides with resonances near the edges of these regions and despite the high sensitivity obtained in the

H2C2C5(C4) and H2C2C5(C6) experiments, the accuracy of the corresponding couplings then is lowered. Similarly, in uncommon cases where  $C_4$  or  $C_6$  resonances overlap with the  $C_2$  region, the H2C2C5(C4) and H2C2C5(C6) experiments will not yield reliable results.

#### *Evaluation of dipolar coupling accuracy.*

Assuming the geometries of nucleic acid bases to be fixed and planar, with bond lengths and angles as marked in Figure 5, there are only three independent dipolar couplings for any given base (Zidek et al., 2003; Bryce and Bax, 2004). We have used the geo-

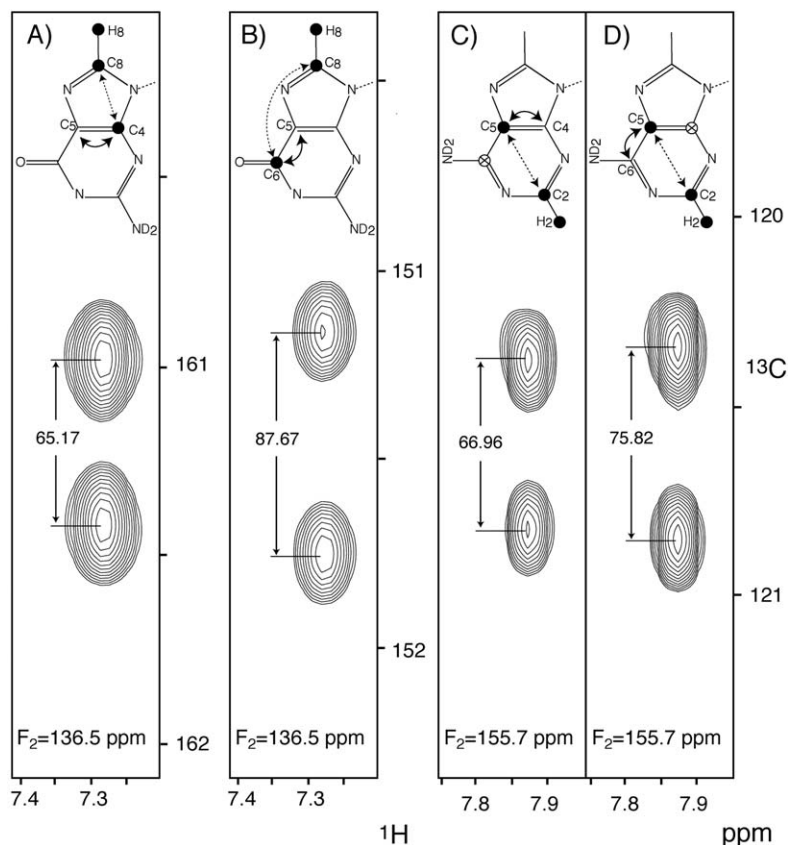


**Figure 5.** Relevant aspects of base geometries used for fitting dipolar couplings. (a) adenine, (b) uridine, (c) guanine, and (d) cytidine. Bond lengths (bold) are marked in pm, and angles (italicized) are given in degrees. For cytidine, the C<sub>5</sub>-C<sub>6</sub> bond was taken to lie at zero degrees, which places C<sub>6</sub>-H<sub>5</sub> at 26.2°, C<sub>5</sub>-H<sub>5</sub> at 58.7°, C<sub>4</sub>-H<sub>5</sub> at 92.2°, C<sub>4</sub>-C<sub>5</sub> at 117.4°, C<sub>6</sub>-H<sub>6</sub> at 119.5°, and C<sub>5</sub>-H<sub>6</sub> at 153.5°. For uridine, the C<sub>5</sub>-C<sub>6</sub> bond was taken to lie at zero degrees, which places C<sub>6</sub>-H<sub>5</sub> at 26.5°, C<sub>5</sub>-H<sub>5</sub> at 59.9°, C<sub>4</sub>-H<sub>5</sub> at 94.2°, C<sub>4</sub>-C<sub>5</sub> at 119.7°, C<sub>6</sub>-H<sub>6</sub> at 118.7°, and C<sub>5</sub>-H<sub>6</sub> at 152.8°. For adenine, the C<sub>4</sub>-C<sub>5</sub> bond was taken to lie at zero degrees, which places C<sub>2</sub>-H<sub>2</sub> at 58.5°, C<sub>8</sub>-H<sub>8</sub> at 91.0°, and C<sub>5</sub>-C<sub>6</sub> at 117.0°.

metric parameters of Clowney et al. for the heavy atoms, which result from a statistical survey of high-resolution neutron and X-ray diffraction structures (Clowney et al., 1996). Base protons were added such that the corresponding C-H bond vector bisects the obtuse angle formed by the surrounding heavy atoms, with the exception of the C<sub>2</sub>-H<sub>2</sub> and C<sub>8</sub>-H<sub>8</sub> vectors in adenine, for which adenosine neutron diffraction coordinates were used (Klooster et al., 1991). The seven dipolar couplings measured for each pyrimidine can be fit to a three-variable cosinusoidal function (eq 4 of Bryce and Bax (2004); see also Supporting Information Figure 5), using the assumed rigid structures and the fact that these RDCs are correlated. In the absence of degenerate, parallel or antiparallel orientations (as nearly applies for C<sub>5</sub>-C<sub>4</sub> and C<sub>6</sub>-H<sub>6</sub> interactions), a perfect fit is always possible if only three couplings

are used as input (Supporting Information Figure 5). A residual in the fit, for the case where more than three couplings are used, reflects either experimental error in the data or inaccuracies in the geometry used. Subsuming rapid internal intrabase vibrations and librations in the average base structure, our mode of analysis is independent of all other types of internal dynamics in the oligonucleotide, as the data for each base are fitted separately.

Shown in Figure 7A are the results of the pyrimidine fits, displayed in a single correlation graph. All dipolar couplings have been appropriately scaled relative to the <sup>1</sup>D<sub>CH</sub> coupling, using the average bond lengths of Figure 5, and equal weights were used for all couplings in the fits. With seven measured data points and only three adjustable parameters in each fit, the residual in the fit is 1.15 Hz and rep-



**Figure 6.** 2D strips taken from the 3D HCC(C)-TROSY spectrum, showing measurement of the  $^1J_{C_6C_5}$  and  $^1J_{C_4C_5}$  splittings in purine bases. Spectra for the aligned 24-nt RNA sample were recorded at 600 MHz using a cryogenic probe, 25 °C, with total measuring times of 24 h for panels A and B and 14 h for panels C and D, using the pulse scheme of Figure 2. Panels A and B have been extracted from H8C8C4(C5) and H8C8C6(C5) TROSY experiments respectively, at the C<sub>8</sub> frequency of G57 (F<sub>2</sub> dimension). For the C<sub>8</sub>→C<sub>4</sub> and C<sub>8</sub>→C<sub>6</sub> transfers, shaped pulse **a** was optimized to invert selectively C<sub>6</sub> and C<sub>4</sub> respectively, and shaped pulse **b** was not applied. The chemical shifts of C<sub>6</sub> and C<sub>4</sub> are very similar for a given nucleotide type. In case of overlap in the C<sub>8</sub>H<sub>8</sub> plane, these couplings can be extracted for adenine bases from H2C2C5(C4) and H2C2C5(C6) TROSY experiments (panels C and D respectively). The planes C and D have been extracted, at the C<sub>2</sub> frequency of A49 (F<sub>2</sub> dimension). For the two experiments starting from C<sub>2</sub>, shaped pulse **a** is not applied during the transfer, and shaped pulse **b** is adjusted to invert selectively C<sub>6</sub> (panel C) or C<sub>4</sub> (panel D) during the frequency labelling of C<sub>5</sub>. On each panel the measured C<sub>5</sub>C<sub>4</sub> (A and C) or C<sub>5</sub>C<sub>6</sub> (B and D) splittings are displayed. The time domain 3D matrix sizes: 50\* × 30\* × 512\* data points, with acquisition times of 35 ms (t<sub>1</sub>), 22 ms (t<sub>2</sub>) and 64 ms (t<sub>3</sub>) for the H8C8C4(C5) TROSY experiment (A); 33\* × 38\* × 512\* data points, with acquisition times of 36 ms (t<sub>1</sub>), 27 ms (t<sub>2</sub>) and 64 ms (t<sub>3</sub>) for the H8C8C6(C5) TROSY experiment (B); and 18\* × 22\* × 512\* data points, with acquisition times of 36 ms (t<sub>1</sub>), 33 ms (t<sub>2</sub>) and 64 ms (t<sub>3</sub>) for the H2C2C5(C4) and H2C2C5(C6) TROSY experiment (C, D). Insets mark the spins for which the frequency is edited (●), the proton and carbon spins decoupled (×), the splittings observed (↔), and the long-range magnetization transfer pathway (dashed arrows).

resents a lower limit for the combined error resulting from measurement error and uncertainty in the base geometry.

A second more stringent test for evaluating the accuracy of the couplings is to use only six dipolar couplings as input values for the fit, and to predict the seventh coupling (Figure 7B). This process is repeated seven times per pyrimidine, each time leaving out a different coupling. The discrepancy between the measured and predicted values now reflects the

sum of the error in the measurement of the seventh datapoint and the error in the prediction, which is based on the fit containing six error-prone experimental data points to an imperfect mean structure. This latter procedure therefore yields an upper limit for the error in the measured dipolar couplings (rmsd of fit is 2.01 Hz). Note that this error is only 2.8% of the total range (72 Hz) spanned by  $^1D_{CH}$  couplings for the alignment tensor values applicable for this oligonucleotide ( $D_a^{CH} = -20.6 \pm 0.5$  Hz;  $R =$

Table 3. Shaped pulse parameters for the HCC(C)-TROSY pulse scheme of Figure 2<sup>a</sup>

Experiment splittings	H2C2C5(C4) C <sub>5</sub> C <sub>4</sub>	H2C2C5(C6) C <sub>5</sub> C <sub>6</sub>	H8C8C6(C5) C <sub>6</sub> C <sub>5</sub>	H8C8C4(C5) C <sub>4</sub> C <sub>5</sub>
Shape <i>a</i>	None	None	IBURP2 5.0 ms 151.0 ppm	IBURP2 5.0 ms 159.3 ppm
Shape <i>b</i>	IBURP2 11.0 ms 158.2 ppm	IBURP2 11.0 ms 149.5 ppm	None	None
Shape <i>c</i>	REBURP 2.0 ms 121 ppm	REBURP 2.0 ms 121 ppm	REBURP 2.0 ms 155 ppm	REBURP 2.0 ms 155 ppm
T duration <sup>b</sup>	18 ms	18 ms	20 ms	17 ms

<sup>a</sup>For each shaped pulse, the shape form, the duration, and the frequency are indicated. Parameters are given for a spectrometer operating at  $B_0 = 14.1$  T. Each required pulse duration scales inversely with the strength of the magnetic field.

<sup>b</sup>Duration of T in ms. Total duration of the constant-time evolution period equals 2T.

$0.31 \pm 0.03$  (O’Neil-Cabello et al., 2004b)), corresponding to cross-validated Q values of less than 10% (see eq 1 of (Ottiger and Bax, 1999)). It is also worth noting that the different types of couplings are predicted about equally well in Figure 7B, suggesting that the sum of the error in the prediction and measurement for a given coupling is very similar for different types of couplings.

For adenine, only four couplings have been measured. A fit of only four couplings to a model adenine structure results in a very good fit (Supporting Information Figure 1), but this is largely artifactual as the number of adjustable parameters in the fit (three) is only slightly smaller than the number of experimental dipolar couplings (four). When using three experimental data points to cross-validate the fourth, the scatter is considerably larger (4.5 Hz; Supporting Information Figure 2). In this case, the rms error is dominated by the propagated error resulting from the fit of three error-containing datapoints to parameterize the equation that describes the orientational dependence of dipolar coupling in the plane of the base. This latter conclusion is supported by repeating the same procedure for pyrimidines: For example, if three uridine couplings are used to predict a fourth, the rms difference between the predicted and observed dipolar couplings increases from 2.0 Hz (Figure 7B) to 4.6 Hz (Supporting Information Figure 3), a number very similar to that seen for adenine. Similar results are observed when the procedure is applied to cytidines

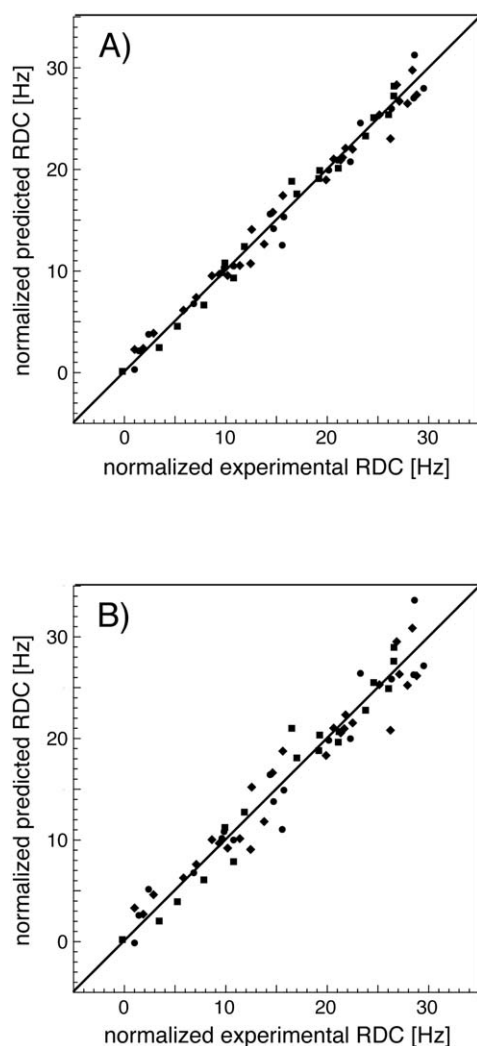
(Supporting Information). These results confirm that the experimental accuracy for adenine is comparable to that for pyrimidine base dipolar couplings.

For guanine, only three couplings have been measured, so no independent method for estimating the accuracy of these couplings is available. However, considering the similarity in coupling pattern and sensitivity obtained for G and A base couplings, their error is expected to be very similar.

For the 8-base pair helical stem, the quality of the fit to a model A-form helix, generated using the program Insight, is comparable for all three types of couplings,  $^1D_{CH}$ ,  $^2D_{CH}$ , and  $^1D_{CC}$  (3.4, 3.4, and 3.0 Hz rmsd, respectively, when each is normalized to the one-bond  $^1D_{CH}$  interaction and each class of couplings is fit separately. If all are fit simultaneously, the rmsd increases to 4.6 Hz. This rmsd is considerably higher than the upper limit for the measurement error estimated above, suggesting that the couplings will allow refinement of the helical structure beyond that of a model helix.

### Concluding remarks

Efforts were made to refine the pyrimidine H<sub>5</sub> and H<sub>6</sub> proton positions by systematically varying the C<sub>5</sub>-H<sub>5</sub>/C<sub>6</sub>-H<sub>6</sub> bond angles and/or C-H distances and evaluating the RDC cross-validation statistics by generating correlation graphs of the type shown in Fig-



**Figure 7.** Correlation between experimental and predicted RDC data for pyrimidines. The fit (A) and cross-validation procedure (B) are based on the fact that the RDCs are correlated (Zidek et al., 2003; Bryce and Bax, 2004). (■), (●) and (◆) correspond to the  $^1D_{CC}$ ,  $^1D_{CH}$  (●) and  $^2D_{CH}$  couplings (◆), respectively. Geometries used are from Figure 5. RDCs are normalized to the aromatic one-bond C-H interaction. (A) Correlation resulting from a fit of  $^1D_{C6H6}$ ,  $^1D_{C5H5}$ ,  $^1D_{C4C5}$ ,  $^1D_{C5C6}$ ,  $^2D_{C6H5}$ ,  $^2D_{C4H5}$ , and  $^2D_{C5H6}$  for C40, U41, U44, U47, U54, U55, C58, C59 and C60 to separate cosinusoidal functions for each base. Pearson's correlation coefficient  $R_p = 0.991$ , and the rmsd between experimental and predicted pyrimidine RDCs is 1.15 Hz. (B) Cross-validation of pyrimidine data, where six RDCs are used to predict the 7th. Pearson's correlation coefficient  $R_p = 0.9733$ , and the rmsd between experimental and predicted RDCs is 2.0 Hz.

ure 7B. However, we were unable to obtain a statistically significant improvement when changing the  $N_1C_6H_6$  angle or the  $C_6C_5H_5$  angle from the ideal values marked in Figure 5. Similarly, changing the C-H bond lengths relative to the standard values marked in Figure 5 did not improve cross validation statistics either when using the constraint that all C-H bond lengths should change by the same fraction relative to their starting value, as expected if librational effects for these bonds are comparable, or by allowing the  $C_5H_5$  and  $C_6H_6$  distances to be completely independent variables. Therefore, the pyrimidine structural parameters shown in Figure 5 are very close to optimal for the analysis of dipolar couplings in nucleic acids. For purines, the number of dipolar couplings measured is insufficient to make a similar evaluation, although all indications are that these geometries are close to optimal too.

Our cross validation results indicate that for each type of normalized measured dipolar coupling, the upper limit for the measurement error is less than *ca* 10% of  $D_a^{CH}$ . This error is small compared to the rms difference between observed dipolar interactions and those predicted by the very best NMR or crystal structures (when these couplings are not used as input restraints). Therefore, the error in the measured couplings is unlikely to be a limiting factor in obtaining the highest possible accuracy of the final structure, calculated from these couplings. Instead, it is likely that the accuracy at which a structure can be determined from such dipolar couplings will be limited by the lack of comparably accurate distance restraints, small deviations from the commonly used and tightly imposed idealized angular restraints, the relative scarcity of restraints related to the phosphodiester torsion angles, and by potential differential dynamics along the oligonucleotide chain.

Even though only three dipolar couplings carry independent information for a planar structure, having more such couplings available will reduce the small impact any potential measurement error could have, as demonstrated in Supporting Information Figure 5. Moreover, the opportunity to check for self-consistency between the overdetermined set of couplings facilitates troubleshooting in cases of assignment or bookkeeping errors (Zidek et al., 2003). Although each set of couplings requires recording of a separate 3D data set, each of these is of good sensitivity and requires only narrow bandwidths in the indirectly detected dimensions, and therefore can be recorded relatively rapidly.

For the 24-nt structure used in the present study, nearly complete sets of couplings were obtained owing to the near-absence of overlap in the NMR spectra, which benefit in resolution from the homonuclear decoupling and TROSY features in the corresponding pulse sequences. For larger oligonucleotides, overlap is likely to increase, resulting in a smaller number of measurable couplings, i.e., in a less degenerate set of couplings. If necessary, the experiments are readily extended to three dimensions, alleviating the potential overlap problem, or complementary base type specific labeling schemes may be used. Moreover, considering that only three couplings are needed to define the base orientation—even fewer in the common case where dipolar couplings for the ribose restrict the  $C_1'$ -N bond vector orientation—a set of base dipolar couplings that is less complete than that of the present study will often be perfectly adequate for structure determination purposes.

The accuracy at which couplings can be measured scales approximately inversely with line width, and also inversely with the signal-to-noise ratio of the NMR data. If the line width is dominated by transverse relaxation, and not by unresolved scalar or dipolar couplings, it will scale approximately linearly with the size of the system. This also applies to the constant-time  $^1\text{H}$  dimension, used for homonuclear  $^1\text{H}$ - $^1\text{H}$  decoupling, where faster transverse relaxation dictates shorter maximum constant-time durations. However, even if the random error in the coupling measurement were three-fold higher than in our current study, it would still be less than about  $\pm 8\%$  of the total range applicable for a given type of coupling, and such couplings would therefore remain quite useful in the structure determination process.

The fact that isotropic J values reported in this study cluster in very narrow regions suggests that it may not be necessary to record the isotropic base couplings for each study, unless extremely precise data are essential. Conversely, if measurements on an isotropic sample show values that deviate significantly, this may serve as a warning that experiments may not be properly optimized. However, it should also be noted that the isotropic coupling values reported here include the effect of weak alignment resulting from magnetic susceptibility anisotropy. Ignoring the contribution from the loop residues to the susceptibility-induced alignment of the 8-basepair stem region, at 750 MHz the effect of the magnetic field induced dipolar contribution to the observed  $^1\text{J}_{\text{CH}}$  splitting in base  $^{13}\text{C}$ - $^1\text{H}$  sites is about  $-1$  Hz, and for  $^{13}\text{C}$ - $^{13}\text{C}$

couplings the alignment contribution is correspondingly smaller.

Future work will focus on how well a nucleic acid structure can be defined by such an abundance of dipolar couplings, when supplemented by extensive ribose dipolar couplings (Miclet et al., 2003; O'Neil-Cabello et al., 2004a) and  $^{31}\text{P}$  CSA restraints (O'Neil-Cabello et al., 2004b).

## Acknowledgements

We thank Frank Delaglio for assistance in data analysis. This work was supported by fellowships from the Human Frontier Science Program (J.B.) and the Natural Sciences and Engineering Research Council of Canada (D.L.B.). EPN acknowledges support from the Robert A. Welch Foundation C-1277 and National Science Foundation MCB-0078501.

## Supporting Information Available

Figures showing (1) predicted (best-fit) vs experimental RDCs for adenine; (2) cross-validation correlation plots for adenine; (3) cross-validation correlation plots for cytidine and uridine bases using only four RDCs, for comparison with adenine results; (4) cross-validation using 3 data points and fits using 7 data points to the three-variable cosinusoidal function for cytidines; A table listing the individual isotropic J values. This material is available at: <http://kluweronline.com/issn/0925-2738>

## References

- Al-Hashimi, H.M., Gorin, A., Majumdar, A., Gosser, Y. and Patel, D.J. (2002) *J. Mol. Biol.*, **318**, 637–649.
- Bachmann, P., Aue, W.P., Müller, L. and Ernst, R.R. (1977) *J. Magn. Reson.*, **28**, 29–39.
- Barbic, A., Zimmer, D.P. and Crothers, D.M. (2003) *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 2369–2373.
- Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J.D. and Williamson, J.R. (1992) *Nucl. Acids Res.*, **20**, 4515–4523.
- Boisbouvier, J., Delaglio, F. and Bax, A. (2003) *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 11333–11338.
- Bondensgaard, K., Mollova, E.T. and Pardi, A. (2002) *Biochemistry*, **41**, 11532–11542.
- Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. and Simorre, J.P. (1998) *J. Am. Chem. Soc.*, **120**, 11845–11851.
- Bryce, D.L. and Bax, A. (2004) *J. Biomol. NMR*, **28**, 273–287.
- Cabello-Villegas, J., Winkler, M.E. and Nikonowicz, E.P. (2002) *J. Mol. Biol.*, **319**, 1015–1034.

- Chou, J.J., Gaemers, S., Howder, B., Louis, J.M. and Bax, A. (2001) *J. Biomol. NMR*, **21**, 377–382.
- Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998) *J. Am. Chem. Soc.*, **120**, 10571–10572.
- Clowney, L., Jain, S.C., Srinivasan, A.R., Westbrook, J., Olson, W.K. and Berman, H.M. (1996) *J. Am. Chem. Soc.*, **118**, 509–518.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
- D'Souza, V., Dey, A., Habib, D. and Summers, M.F. (2004) *J. Mol. Biol.*, **337**, 427–442.
- Farmer, B.T., Pardi, A. and Mueller, L. (1995) *J. Cell. Biochem.*, **16**–16.
- Fiala, R., Czernek, J. and Sklenar, V. (2000) *J. Biomol. NMR*, **16**, 291–302.
- Green, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93–141.
- Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1987) *J. Magn. Reson.*, **75**, 474–492.
- Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.
- Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry* **29**, 4659–4667.
- Ippel, J.H., Wijmenga, S.S., deJong, R., Heus, H.A., Hilbers, C.W., deVroom, E., vanderMarel, G.A. and vanBoom, J.H. (1996) *Magn. Reson. Chem.*, **34**, S156–S176.
- Ishii, Y., Markus, M.A. and Tycko, R. (2001) *J. Biomol. NMR*, **21**, 141–151.
- Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.
- Klooster, W.T., Ruble, J.R., Craven, B.M. and McMullan, R.K. (1991) *Acta Crystallogr. Sect. B- Struct. Commun.*, **47**, 376–383.
- Kontaxis, G., Clore, G.M. and Bax, A. (2000) *J. Magn. Reson.*, **143**, 184–196.
- Kung, H.C., Wang, K.Y., Goljer, I. and Bolton, P.H. (1995) *J. Magn. Reson. Ser. B*, **109**, 323–325.
- Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z.R. and Summers, M.F. (2003) *J. Mol. Biol.* **326**, 529–542.
- Leeper, T., Leulliot, N. and Varani, G. (2003) *Nucl. Acids Res.*, **31**, 2614–2621.
- Lukavsky, P.J., Kim, I., Otto, G.A. and Puglisi, J.D. (2003) *Nat. Struct. Biol.*, **10**, 1033–1038.
- MacDonald, D. and Lu, P. (2002) *Curr. Opin. Struct. Biol.*, **12**, 337–343.
- Masse, J.E., Bortmann, P., Dieckmann, T. and Feigon, J. (1998) *Nucl. Acids Res.*, **26**, 2618–2624.
- Meier, S., Haussinger, D. and Grzesiek, S. (2002) *J. Biomol. NMR*, **24**, 351–356.
- Miclet, E., O'Neil-Cabello, E., Nikonowicz, E.P., Live, D. and Bax, A. (2003) *J. Am. Chem. Soc.*, **125**, 15740–15741.
- Nikonowicz, E.P., Sirt, A., Legault, P., Jucker, F.M., Baer, L.M. and Pardi, A. (1992) *Nucl. Acids Res.*, **20**, 4507–4513.
- O'Neil-Cabello, E., Bryce, D.L., Nikonowicz, E.P. and Bax, A. (2004a) *J. Am. Chem. Soc.*, **126**, 66–67.
- O'Neil-Cabello, E., Wu, Z., Bryce, D.L., Nikonowicz, E.P. and Bax, A. (2004b) *J. Biomol. NMR*, **30**, 61–70.
- Ono, A., Tate, S., Ishido, Y. and Kainosho, M. (1994) *J. Biomol. NMR*, **4**, 581–586.
- Ottiger, M. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 187–191.
- Ottiger, M., Delaglio, F. and Bax, A. (1998) *J. Magn. Reson.*, **131**, 373–378.
- Palmer, A.G., Cavanagh, J., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **93**, 151–170.
- Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1998) *J. Am. Chem. Soc.*, **120**, 6394–6400.
- Ruckert, M. and Otting, G. (2000) *J. Am. Chem. Soc.*, **122**, 7793–7797.
- Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2000) *J. Biomol. NMR*, **18**, 303–309.
- Steff, R., Wu, H.H., Ravindranathan, S., Sklenar, V. and Feigon, J. (2004) *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 1177–1182.
- Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
- Tjandra, N., Tate, S., Ono, A., Kainosho, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 6190–6200.
- Tycko, R., Blanco, F.J. and Ishii, Y. (2000) *J. Am. Chem. Soc.*, **122**, 9340–9341.
- Ulmer, T.S., Ramirez, B.E., Delaglio, F. and Bax, A. (2003) *J. Am. Chem. Soc.*, **125**, 9179–9191.
- Vermeulen, A., Zhou, H.J. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 9638–9647.
- Weigelt, J. (1998) *J. Am. Chem. Soc.*, **120**, 10778–10779.
- Wu, H.H., Yang, P.K., Butcher, S.E., Kang, S., Chanfreau, G. and Feigon, J. (2001) *Embo J.*, **20**, 7240–7249.
- Yan, J.L., Corpora, T., Pradhan, P. and Bushweller, J.H. (2002) *J. Biomol. NMR*, **22**, 9–20.
- Zhu, G. and Bax, A. (1990) *J. Magn. Reson.*, **90**, 405–410.
- Zidek, L., Padrt, P., Chmelik, J. and Sklenar, V. (2003) *J. Magn. Reson.*, **162**, 385–395.
- Zidek, L., Wu, H.H., Feigon, J. and Sklenar, V. (2001) *J. Biomol. NMR*, **21**, 153–160.
- Zimmer, D.P. and Crothers, D.M. (1995) *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 3091–3095.