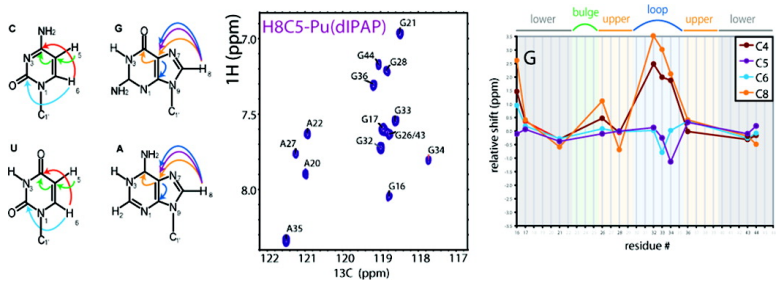
Article

C-Detection in RNA Bases: Revealing Structure–Chemical Shift Relationships

Christophe Fars, Irene Amata, and Teresa Carlomagno

J. Am. Chem. Soc., **2007**, 129 (51), 15814-15823 • DOI: 10.1021/ja0727417



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 12/04/2007

^{13}C-Detection in RNA Bases: Revealing Structure−Chemical Shift Relationships

Christophe Farès, Irene Amata, and Teresa Carlomagno*

Contribution from the Max-Planck-Institute for Biophysical Chemistry, Department of NMR-based Structural Biology, Am Fassberg 11, D-37077 Göttingen, Germany

Received April 26, 2007; E-mail: taco@nmr.mpibpc.mpg.de

Abstract: The chemical shifts of the unprotonated carbons in the proton-deficient nucleobases of RNA are rarely reported, despite the valuable information that they contain about base-pairing and base-stacking. We have developed ^{13}C-detected 2D-experiments to identify the unprotonated ^{13}C in the RNA bases and have assigned all the base nuclei of uniformly ^{13}C,^{15}N-labeled HIV-2 TAR-RNA. The ^{13}C chemical shift distributions revealed perturbations correlated with the base-pairing and base-stacking properties of all four base-types. From this work, we conclude that the information contained in the chemical shift perturbations within the base rings can provide valuable restraint information for solving RNA structures, especially in conformational averaged regions, where NOE-based information is not available.

Introduction

Nucleic acids, such as RNA, are universal molecules that play key roles in heredity, transcription, and translation. The understanding of their interactions is serving our insight into essential biochemical processes. However, their structural elucidation using NMR methodologies has not attained the same level of standardization as for proteins. To date, only 372 NMR-solved structures containing RNA have been deposited into the structure databanks, more than a factor of 10 less than for proteins (5361).[1−3]

The difficulties associated with NMR of oligonucleotides come on one hand from the poor variability in the residue types, causing a high level of signal overlap and making the assignment procedure a challenging task. Furthermore, nucleic acids are relatively deficient in protons along the backbone and on the nucleobase side-chains, which complicates the sequential assignment and measurements of long-range NOEs. Most elucidated RNA structures rely on relatively few key sequential NOEs or NOEs across hydrogen bonds. These are commonly complemented by torsion angles, H-bonds, and orientation restraints based on chemical shifts, *J*-couplings, and RDCs.

In the proton-deficient nucleobases of RNA, the unprotonated carbons ((Pu) C_2(G), C_4, C_5, C_6; and (Py) C_2, C_4) are at most used as relaying spins in the assignment procedure of their protonated counterparts ((Pu) C_2(A), C_8 and (Py) C_5, C_6).[4−7] These low-γ spins provide no useful internuclear distance or torsional constraints, and their RDC-based orientation constraints in the rigid aromatic base have only recently been used[8] to complement those more easily measured from the protonated C and N.[9−12] In fact, the chemical shift values for these spins are very rarely deposited in the Biological Magnetic Resonance Data Bank (BMRB).[13] Pulse sequences aimed at the measurement of unprotonated carbon chemical shifts have been recently proposed in the literature.[14,15]

Chemical shifts depend on the electronic environment and thus hold valuable information about the local structure. The use of these structure-chemical shift relationships is founded on extensive chemical shift databases or on accurate ab initio predictions. In fact, the extraction of semiquantitative structural information from chemical shift data is commonly used in protein backbone to derive secondary structure elements.[16,17] In nucleic acids, sugar pucker modes have also been predicted from canonical coordinates obtained from the chemical shift of all ribose carbons[15] and are based on ab initio calculations by Harbison and co-workers.[18−20] A statistical analysis was also carried out by Cromsigt et al.[21] to demonstrate the utility of ^{1}H

(1) Berman, H. M.; et al. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 899−907.
(2) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S. H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751−759.
(3) Murthy, V. L.; Rose, G. D. *Nucleic Acids Res.* **2003**, *31*, 502−504.
(4) Legault, P.; Farmer, B. T.; Mueller, L.; Pardi, A. *J. Am. Chem. Soc.* **1994**, *116*, 2203−2204.
(5) Marino, J. P.; Prestegard, J. H.; Crothers, D. M. *J. Am. Chem. Soc.* **1994**, *116*, 2205−2206.
(6) Simorre, J. P.; Zimmermann, G. R.; Mueller, L.; Pardi, A. *J. Biomol. NMR* **1996**, *7*, 153−156.
(7) Simon, B.; Zanier, K.; Sattler, M. *J. Biomol. NMR* **2001**, *20*, 173−6.
(8) Jaroniec, C. P.; Boisbouvier, J.; Tworowska, I.; Nikonowicz, E. P.; Bax, A. *J. Biomol. NMR* **2005**, *31*, 231−241.
(9) Boisbouvier, J.; Bryce, D. L.; O'Neil-Cabello, E.; Nikonowicz, E. P.; Bax, A. *J. Biomol. NMR* **2004**, *30*, 287−301.
(10) Hennig, M.; Carlomagno, T.; Williamson, J. R. *J. Am. Chem. Soc.* **2001**, *123*, 3395−3396.
(11) Yan, J. L.; Corpora, T.; Pradhan, P.; Bushweller, J. H. *J. Biomol. NMR* **2002**, *22*, 9−20.
(12) Zidek, L.; Wu, H. H.; Feigon, J.; Sklenar, V. *J. Biomol. NMR* **2001**, *21*, 153−160.
(13) Seavey, B. R.; Farr, E. A.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* **1991**, *1*, 217−36.
(14) Fiala, R.; Munzarova, M. L.; Sklenar, V. *J. Biomol. NMR* **2004**, *29*, 477−490.
(15) Furtig, B.; Richter, C.; Bermel, W.; Schwalbe, H. *J. Biomol. NMR* **2004**, *28*, 69−79.
(16) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171−180.
(17) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363−392.
(18) Ebrahimi, M.; Rossi, P.; Rogers, C.; Harbison, G. S. *J. Magn. Reson.* **2001**, *150*, 1−9.
(19) Rossi, P.; Harbison, G. S. *J. Magn. Reson.* **2001**, *151*, 1−8.
(20) Rossi, P.; Harbison, G. S. *Biophys. J.* **1999**, *76*, A319−a319.

10.1021/ja0727417 CCC: $37.00

chemical shifts for validating structures of RNA. Similar approaches could be devised for the carbons of the RNA bases. In density function theory (DFT) calculations[22] using the GIAO method, it was predicted that variations in chemical shifts of carbons can be expected for each residue type correlated with simple structural features such as base-pairing, base-stacking, and base orientation as defined by the glycosidic torsion angle χ (C2/C4−N1/N9−C1′-O4′).[22] Indeed, chemical shift variations of up to 4 ppm are to be expected for carbons in the base as a function of the glycosidic torsion angle, while base-pairing or base-stacking can account for chemical shifts variations up to 5 ppm.[22] The authors of this report, however, lack an important bank of experimental data to support their findings, partly due to the challenges associated with the measurement of these unprotonated aromatic carbon chemical shifts.

Recent advances in NMR spectroscopy include the development of the so-called *protonless* experiments, in which the sensitivity disadvantage of detecting low-γ nuclei is more than compensated for by a number of other advantages. Among them is the ability to bypass the fast-relaxing ^{1}H in large systems or in systems containing paramagnetic agents.[23−31] The development of protonless experiments has also been accompanied by important improvements in sensitivity of NMR probes to ^{13}C, through cryogenically cooled coils and preamplifiers, coil design, and higher magnetic fields.[32,33] These recent developments opened the way to direct detection of ^{13}C in liquid-state NMR.

In this work, we propose multidimensional ^{13}C-detected experiments that correlate the commonly assigned and non-exchangeable ^{1}H to unprotonated carbons in the aromatic moiety of all common base-types in RNA. With respect to other experiments previously proposed to measure the chemical shift of unprotonated carbons in RNA,[14,15] this new suite of ^{13}C-detected experiments has the advantage of minimizing the number of long delays during which the proton magnetization is transversal and therefore is very well suited for large RNAs or in the presence of line-broadening due to conformational averaging. In contrast, for small RNAs the previously proposed experiments[14] based on ^{1}H acquisition are to be preferred. During revision of this manuscript, a similar approach to the measurement of the chemical shifts of unprotonated carbons in RNA has been proposed by another group.[34] Using the new experiments, the nearly complete assignment of 10 different types of unprotonated carbons in RNA bases was achieved in a 30-residue RNA molecule corresponding to the trans activation response element of the human immunodeficiency virus type 2 (HIV-2 TAR-RNA). During the replication of the viral mRNA, this element is involved in the recognition of the complex composed of the trans acting Tat protein and the cyclin subunit (CycT1) of the positive transcription elongation factor complex (P-TEFb).[35] This key event in the elongation of the viral mRNA makes this element an ideal candidate for drug targeting, and thus it has been the object of numerous structural studies.[36−43] This well-characterized RNA element features two distinct stems, a pair of bulged out residues and a hexameric loop. The analysis of the data for each residue in this oligonucleotide shows that each carbon exhibits complementary sensitivities to the local base configuration and dynamics.

The dependence of the chemical shifts of protonated carbons on secondary structure elements in the TAR-RNA was compared to that observed for the chemical shift data deposited in the Biological Magnetic Resonance Data Bank (BMRB).[13] The small number of unprotonated carbons chemical shifts submitted to the BMRB[13] and the fact that most of the available data refer to special RNA structures, as for example the UUCG tetraloop,[15] did not allow a comprehensive comparison for unprotonated carbons. However, the trends observed in the chemical shift−structure relationship in the TAR-RNA are largely confirmed by the analysis of the chemical shifts of protonated carbons reported in the BMRB,[13] further supporting the usefulness of the chemical shift data as structural restraints.

Materials and Methods

The uniformly ^{13}C,^{15}N-labeled 30-mer oligonucleotide encompassing the trans activation response (TAR) element of HIV-2 RNA (^{16}GGCCAGA^{23}U^{25}UGAGCCUGGGAGCUCUCUGGC^{46}C) was enzymatically synthesized from labeled NTPs as described previously.[44] The HIV-2 TAR-RNA lacks the C in position 24 present in the HIV-1 TAR-RNA. The RNA was dissolved in Na_3PO_4 buffer (20 mM; pD 6.6) with D_2O (99.95%), heated to 80 °C for 2 min, allowed to anneal at 0 °C for 2 min, and transferred to a Shigemi (Allison Park, PA) sample tube. The total sample volume was 250 μL for a final concentration of 1.5 mM. All experiments were performed maintaining the sample temperature at 298 K.

NMR Spectroscopy. All spectra were acquired on Bruker DRX800 spectrometer (Karlsruhe, Germany) equipped with a 5-mm ^{1}H[^{13}C/^{15}N] triple resonance probehead featuring cryogenically cooled preamplifiers and rf coils on ^{1}H and ^{13}C channels and *z*-axis gradient. New ^{1}H-excited/^{13}C-detected 2D NMR experiments with selectivity for pyrimidine (Py) or for purine (Pu) bases are presented: H6C2-Py, H6C4-Py(IPAP), H5c5C4-Py(IPAP), H8c5C4-Pu(IPAP), H8c5C6-Pu(IPAP), and H8C5-Pu(dIPAP) (Figure 1). The name of these pulse programs refers to the correlated atoms (from excited to detected). The suffixes IPAP (in-phase/antiphase) or dIPAP (double IPAP) refer to the ^{13}C−^{13}C operators

(21) Cromsigt, J. A.; Hilbers, C. W.; Wijmenga, S. S. *J. Biomol. NMR* **2001**, *21*, 11−29.
(22) Xu, X. P.; Au-Yeung, S. C. F. *J. Phys. Chem. B* **2000**, *104*, 5641−5650.
(23) Bermel, W.; Bertini, I.; Felli, I. C.; Kummerle, R.; Pierattelli, R. *J. Am. Chem. Soc.* **2003**, *125*, 16423−16429.
(24) Bermel, W.; Bertini, I.; Felli, I. C.; Kummerle, R.; Pierattelli, R. *J. Magn. Reson.* **2006**, *178*, 56−64.
(25) Bermel, W.; Bertini, I.; Felli, I. C.; Lee, Y. M.; Luchinat, C.; Pierattelli, R. *J. Am. Chem. Soc.* **2006**, *128*, 3918−3919.
(26) Bermel, W.; Bertini, I.; Felli, I. C.; Piccioli, M.; Pierattelli, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 25−45.
(27) Bertini, I.; Duma, L.; Felli, I. C.; Fey, M.; Luchinat, C.; Pierattelli, R.; Vasos, P. R. *Angew. Chem., Int. Ed.* **2004**, *43*, 2257−2259.
(28) Bertini, I.; Felli, I. C.; Kummerle, R.; Moskau, D.; Pierattelli, R. *J. Am. Chem. Soc.* **2004**, *126*, 464−465.
(29) Bertini, I.; Jimenez, B.; Piccioli, M.; Poggi, L. *J. Am. Chem. Soc.* **2005**, *127*, 12216−12217.
(30) Bertini, I.; Jimenez, B.; Piccioll, M. *J. Magn. Reson.* **2005**, *174*, 125−132.
(31) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. *ChemBioChem* **2005**, *6*, 1536−1549.
(32) Kovacs, H.; Moskau, D.; Spraul, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 131−155.
(33) Shimba, N.; Kovacs, H.; Stern, A. S.; Nomura, A. M.; Shimada, I.; Hoch, J. C.; Craik, C. S.; Dotsch, V. *J. Biomol. NMR* **2004**, *30*, 175−9.
(34) Fiala, R.; Sklenar, V. *J. Biomol. NMR* **2007**, *39*, 153−163.
(35) Wei, P.; Garber, M. E.; Fang, S. M.; Fischer, W. H.; Jones, K. A. *Cell* **1998**, *92*, 451−62.
(36) Aboul-ela, F.; Karn, J.; Varani, G. *Nucleic Acids Res.* **1996**, *24*, 3974−81.
(37) Brodsky, A. S.; Williamson, J. R. *J. Mol. Biol.* **1997**, *267*, 624−39.
(38) Colvin, R. A.; White, S. W.; Garcia-Blanco, M. A.; Hoffman, D. W. *Biochemistry* **1993**, *32*, 1105−12.
(39) Du, Z.; Lind, K. E.; James, T. L. *Chem. Biol.* **2002**, *9*, 707−12.
(40) Faber, C.; Sticht, H.; Schweimer, K.; Rosch, P. *J. Biol. Chem.* **2000**, *275*, 20660−20666.
(41) Jaeger, J. A.; Tinoco, I. *Biochemistry* **1993**, *32*, 12522−12530.
(42) Raghunathan, D.; Sanchez-Pedregal, V. M.; Junker, J.; Schwiegk, C.; Kalesse, M.; Kirschning, A.; Carlomagno, T. *Nucleic Acids Res.* **2006**, *34*, 3599−608.
(43) Richter, S.; Cao, H.; Rana, T. M. *Biochemistry* **2002**, *41*, 6391−7.
(44) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783−8798.

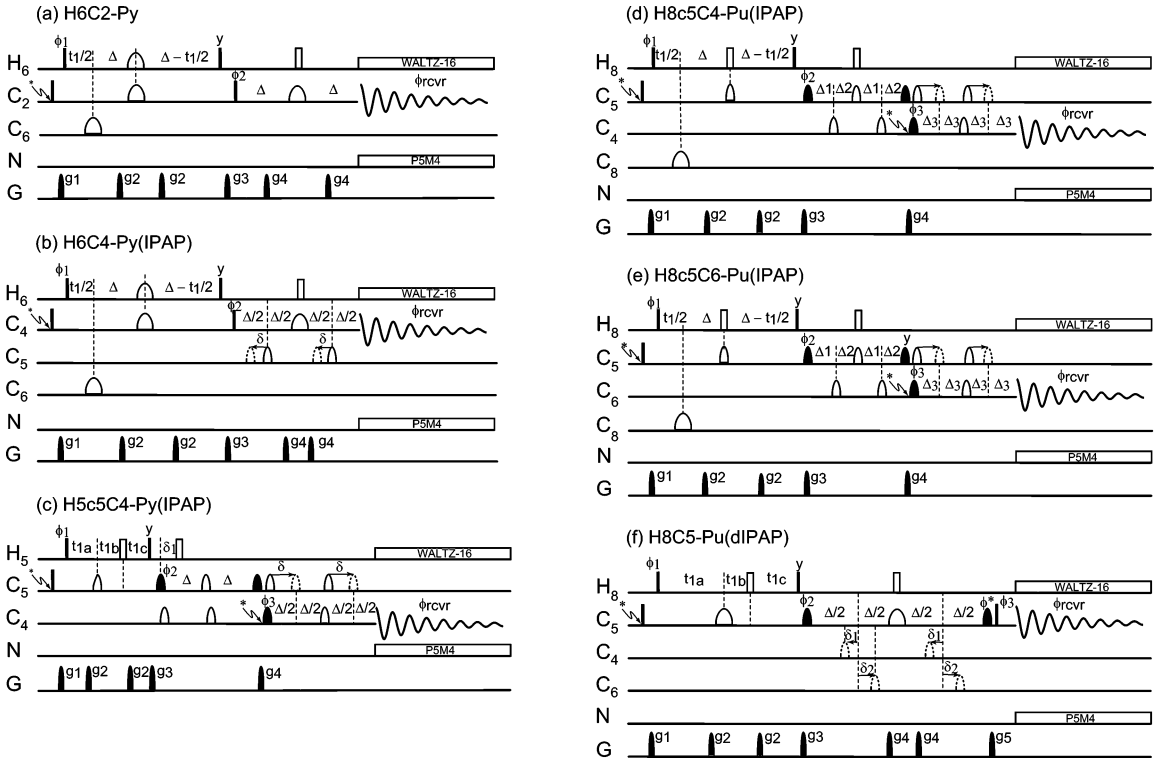


Figure 1. Pulse sequence for the carbon-detected experiments. Filled- and open-pulse symbols represent 90° and 180° rf pulses, and rectangular and rounded pulses symbolize hard and selective pulses. Carbon hard pulses apply to all carbon "channels". The hard 180°$_x$ pulses on the carbon channel are applied as a composite 59.4°$_x$298°$_{-x}$59.4°$_x$. All pulse phases are along *x* except where otherwise indicated. The ^{13}C carrier frequency is set and/or switched to the middle of the specified carbon resonance region as denoted by the asterisk (∗). Where necessary, the antiphase experiments are obtained by substituting the full-line pulses with the dashed-line pulses, as indicated by arrows. Frequency discrimination is performed in all cases in States-TPPI mode by incrementing the phase of the pulse labeled ϕ_1. The other experimental parameters are described in the Materials and Methods.

that are observed in an interleaved way to allow for post-acquisition homonuclear decoupling, according to Ottiger et al.[45] and Duma et al.[46] Details for these experiments are described in the following section. Additional HSQC spectra of the RNA bases were acquired at the two pD values of 6.6 and 7.4 to verify the effect of potential partial protonation of the adenine N1 on the chemical shifts.[4]

All processing including the decoupling was performed using NMRPipe[47] and the spectra were analyzed in Sparky.[48] All reported proton and carbon chemical shift assignments from these experiments are referenced to the methyl group protons and carbons of 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS), prepared in the same buffer, as an external standard.

2D ^{1}H-Excited ^{13}C-Detected Experiments. Figure 1 displays the pulse sequences to observe the $C2_{Py}$ (H6C2) or the $C4_{Py}$ (H6C4 or H5c5C4) in the pyrimidine rings, as well as the $C4_{Pu}$ (H8c5C4), the $C6_{Pu}$ (H8c5C6) or the $C5_{Pu}$ (H8C5) of the purine rings. Figure 2 displays the corresponding spectra.

All experiments start with a π/2-pulse-gradient pair on ^{13}C to purge any longitudinal carbon magnetization. When necessary, transfers are accomplished through ^{1}H-^{13}C INEPT or selective ^{13}C-^{13}C INEPT steps. The ^{1}H chemical shift t_1-indirect evolution and the first ^{1}H-^{13}C INEPT are combined into a semiconstant time or constant time evolution step. Except for the H6C2 experiment, the final coherence is allowed to evolve either in a form that is in-phase or antiphase with respect to their bonded ^{13}C neighbor just prior to detection, to allow for homonuclear decoupling using the method devised by Ottiger et al.[45] and Duma et al.[46]

Detection is accompanied with ^{1}H decoupling and ^{15}N broadband decoupling, using WALTZ-16 on ^{1}H (ν_{rf} = 2.5 kHz) and broadband P5M4[49] (WURST, 20 kHz, $\nu_{rf,max}$ = 2.8 kHz) on ^{15}N. Pulsed-field gradients were applied with the following shapes, intensities, and durations: g1[sinebell, 1.0 ms, 21.5 G/cm], g2 [sinebell, 1.0 ms, 18.5 G/cm], g3[1.0 ms, 8.5 G/cm], g4[sinebell, 1.0 ms, 30 G/cm], g5[sinebell, 1.0 ms, 11.5 G/cm]. The following description lists the relevant delays and shapes [with flip angle, duration, carrier frequency] for individual experiments: (a) H6C2-Py: The shaped pulses on carbon are a Q3 [180°, 1.4 ms, 140 ppm] on the C_6 and two Q3 [180°, 1 ms, 155 ppm] on the C_2, while the shaped pulse on proton is a Q3 [180°, 1 ms, 7.7 ppm] on the H_6. The initial delay Δ is 13.6 ms. Phase cycling consists of ϕ_1, (x, −x); ϕ_2, (x, x, −x, −x); ϕ_{rec}, (x, −x, −x, x). (b) H6C4-Py-(IPAP): The C_4-shaped pulses are both Q3 [180°, 1 ms, 168.0 ppm], the C_5-shaped pulses are both Q3 [180°, 1 ms, 101 ppm], and the C_6-shaped pulse is Q3 [180°, 1.4 ms, 140 ppm]. The shaped pulse on H_6 is a Q3 [180°, 1 ms, 7.7 ppm]. If the sample is dissolved in H_2O, the last π proton pulse should be selective on the H_6, in order to decouple the ∼6 Hz large $^2J_{C4H4a}$ coupling with H_{4a} of cytidines at ∼6.5 ppm. Initial delay Δ is 13.6 ms for the Us, while for the Cs, a shorter delay of 9 ms is recommended, due to the fastest relaxation of the C_4 nuclei in Cs. Phase cycling consists of ϕ_1, (x, −x); ϕ_2, (x, x, −x, −x); ϕ_{rec}, (x, −x, −x, x). (c) H5c5C4-Py(IPAP): Two separate experiments were performed for Cs and Us to optimize the C4−C5 transfer for the different $^1J_{C4C5}$ (65.1 Hz for Us and 54.7 Hz for Cs). Proton chemical shift evolution (t_1) is performed in a semiconstant time way, using increments $\Delta t_{1a} = \Delta t_1/2$, $\Delta t_{1b} = n^*\Delta t_1/2$, and $\Delta t_{1c} = (1 - n)^*\Delta t_1/2$, such that $t_{1c} = 0$ in the final increment. On C_5, the filled-shaped pulses are Q5 [90°, 2 ms, 105(U) or 96.5(C) ppm] and the open-shape pulses are one reburp [180°, 1 ms, 105(U) or 96.5(C) ppm] and three Q3 [180°, 1 ms, 105(U) or 96.5(C) ppm]. On C_4, the filled-shaped pulses

(45) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373−378.
(46) Duma, L.; Hediger, S.; Lesage, A.; Emsley, L. *J. Magn. Reson.* **2003**, *164*, 187−195.
(47) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(48) Goddard, T. D.; Kneller, D. G. *SPARKY3*; University of California, San Francisco, 2001.
(49) Tycko, R.; Pines, A.; Guckenheimer, J. *J. Chem. Phys.* **1985**, *83*, 2775−2802.

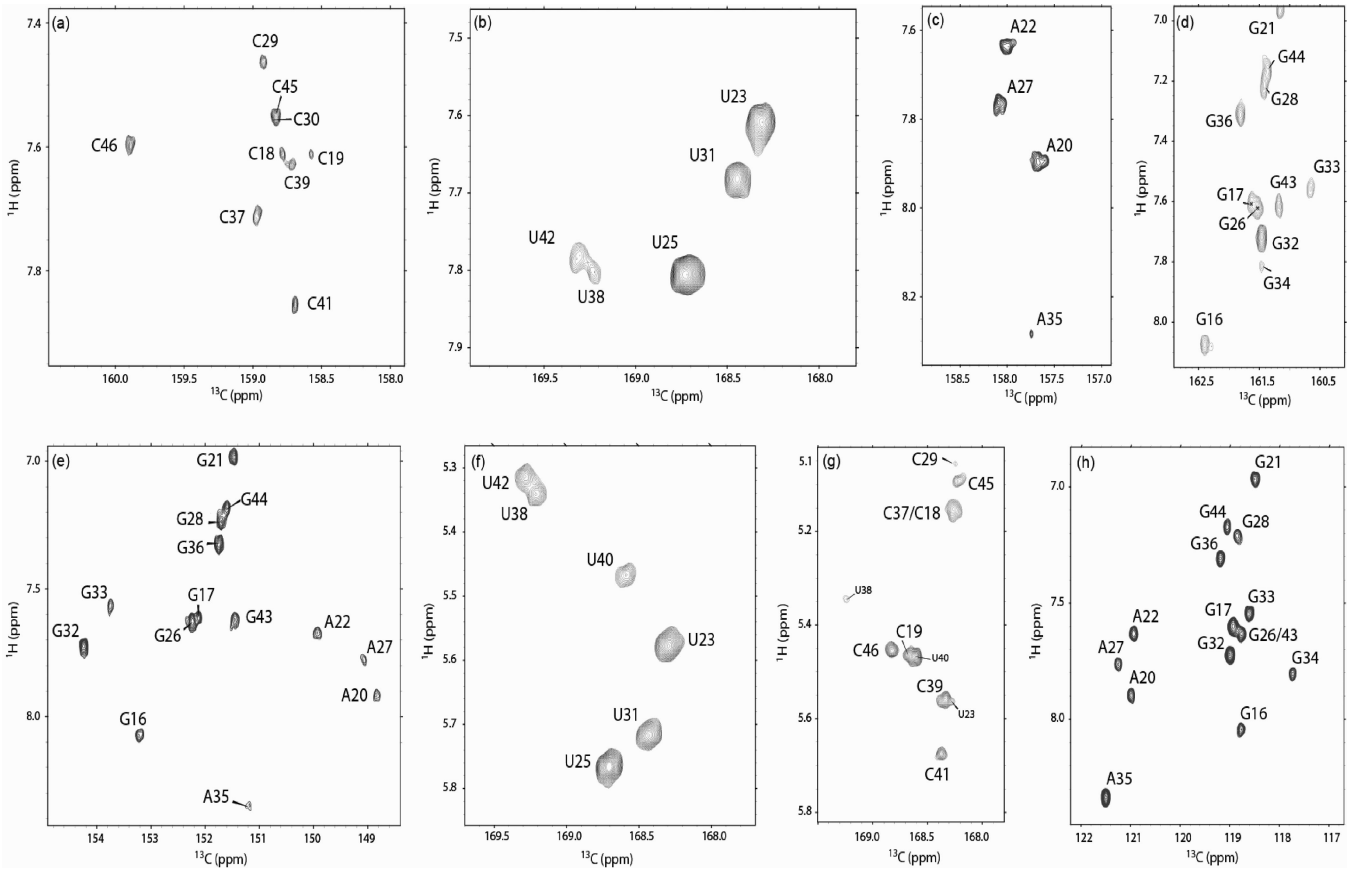


Figure 2. 2D ^{13}C-detected experiments for unprotonated carbons in uniformly labeled HIV-2 TAR-RNA base residues. Labels indicate unambiguous assignments. (a) (H6C2-Py) Here the H6−C2 region of the Cs is shown; the Us′ signals are visible between 153 and 154 ppm (region not shown); (b) (H6C4-Py(IPAP)) here the H6−C4 uridines correlation is shown. The same experiment performed with a shorter delay Δ allows detection of the Cs, which are generally less intense than the Us due to the faster C_4 relaxation rate. (c) (H8c5C6-Pu(IPAP)) selective for the adenosines; (d) (H8c5C6-Pu(IPAP)) selective for the guanosines; (e) H8c5C4-Pu(IPAP); (f) (H5c5C4-Py(IPAP)) selective for the uridines; (g) (H5c5C4-Py(IPAP)) selective for the cytidines; (h) H8C5-Pu(dIPAP). The spectra were recorded at 18.8 T (ν_H = 800.15 MHz) for a 1.5 mM sample at 298 K. The spectral width and the time resolution in the direct dimension (^{13}C) were 20 ppm and 1024 points, respectively, with a recycling delay of 3 s for all the experiments. The indirect (^{1}H) spectral width, number of transients, and number of increments were as follows for each experiment: (a) 0.8 ppm and 512 transients × 16 increments, (b) 0.8 ppm and 256 transients × 16 increments (×2, interleaved), (c) 1.6 ppm and 256 transients × 48 increments (×2, interleaved), (d) 1.6 ppm and 256 transients × 48 increments (×2, interleaved), (e) 1.6 ppm and 320 transients × 48 increments (×2, interleaved), (f) 0.8 ppm and 256 transients × 48 increments (×2, interleaved), (g) 0.8 ppm and 256 transients × 48 increments (×2, interleaved), (h) 1.6 ppm and 64 transients × 76 increments (×4, interleaved). The spectra were all processed with NmrPipe,[47] as previously described in Ottiger et al.[45] and in Duma et al.[46]

are Q5 [90°, 2 ms, 168.7 ppm], and the open shapes are, in order, two iburp2 [180°, 1 ms, 168.7 ppm], and reburp [180°, 1.8 ms, 168.7 ppm]. Initial delays are $\delta_1 = t_{1a} = t_{1c} = 1.25$ ms, $t_{1b} = 0$ and $\Delta = 3.8$(U) or 4.6(C) ms $= \delta$*2. Phase cycling consists of ϕ_1, (x, −x); ϕ_2, (x, x, −x, −x); ϕ_3, (x, x, x, x, −x, −x, −x, −x); ϕ_{rec}, (x, −x, −x, x, −x, x, x, −x). (d) H8c5C4-Pu(IPAP): On C_5, the filled-shaped pulses are Q5 [90°, 2 ms, 114.7 ppm] and the open-shaped pulses are all Q3 [180°, 1 ms, 114.7 ppm]. On C_4, the filled-shaped pulse is Q5 [90°, 2 ms, 149.7 ppm] and the open-shaped pulses are Q3 [180°, 1.8 ms, 149.7 ppm]. The C_8-shape pulse is Q3 [180°, 1.5 ms, 136.5 ppm]. Initial delays are $\Delta = 13.8$ ms, $\Delta_1 = \Delta/2 - 1/8^1J_{C4C5}$, $\Delta_2 = \Delta/2 + 1/8^1J_{C4C5}$ with $^1J_{C4C5} = 64$ Hz, $\Delta_3 = 1/8^1J_{C4C5}$. If high resolution is desired in the proton dimension, evolution of the H8 chemical shifts can be performed in the semiconstant time way, as in panel f. If the sample is dissolved in H_2O, the last π proton pulse should be selective on the H_8, to decouple the ∼6 Hz large $^2J_{C5H6b}$ coupling with amino proton H_{6b} of adenosines. The same also applies to the experiments in panels e and f. Phase cycling consists of ϕ_1, (x, −x); ϕ_2, (x, x, −x, −x); ϕ_3, (x, x, x, x, −x, −x, −x, −x); ϕ_{rec}, (x, −x, −x, x, −x, x, x, −x). (e) H8c5C6-Pu(IPAP): Two experiments were performed to optimize the C5−C6 transfer for the different $^1J_{C5C6}$ (74.8.1 Hz for A and 87.8 Hz for G). On C_5, the filled-shaped pulse is Q5 [90°, 2 ms, 120 ppm], and the open-shaped pulses are all Q3 [180°, 1 ms, 120 ppm]. On C_6, the filled-shaped pulse is Q5 [90°, 2 ms, 157.5(A) or 161(G) ppm] and the open-shaped pulses are all Q3 [180°, 1.8 ms, 157.5(A) or 161(G) ppm]. The C_8-shaped pulse is Q3 [180°, 1.5 ms, 136.5 ppm]. Initial delays are $\Delta = 13.8$ ms, $\Delta_1 = \Delta/2 - 1/8^1J_{C5C6}$, $\Delta_2 = \Delta/2 + 1/8^1J_{C5C6}$ with $^1J_{C5C6} = 74.8$ Hz for A and 87.8 Hz for G, $\Delta_3 = 1/8^1J_{C5C6}$. If high resolution is desired in the proton dimension, evolution of the H8 chemical shifts can be performed in the semiconstant time way, as in panel f. Phase cycling consists of ϕ_1, (x, −x); ϕ_2, (x, x, −x, −x); ϕ_3, (x, x, x, x, −x, −x, −x, −x); ϕ_{rec}, (x, −x, −x, x, −x, x, x, −x). (f) H8C5-Pu(dIPAP): On C_5, the filled-shaped pulses are eburp2 [90°, 2.25 ms, 122.5 ppm] and the open-shaped pulses are, in order, iburp2 [180°, 1.8 ms, 122.5 ppm] and reburp2 [180°, 2.025 ms, 122.5 ppm]. The C_4-shaped pulses are, in order, iburp2 [180°, 3.5 ms, 151.5 ppm] and iburp2 [180°, 3.5 ms, 151.5 ppm, time-reversed]. The C_6-shaped pulses are, in order, iburp2 [180°, 3.5 ms, 159.5 ppm], iburp2 [180°, 3.5 ms, 159.5 ppm, time-reversed]. Initial delays are $\Delta = t_{1a} = t_{1c} = 11.6$ ms, $t_{1b} = 0$, $\delta_1 = 1/8^1J_{C4C5} = 1.42$ ms and $\delta_2 = 1/8^1J_{C5C6} = 1.92$ ms. The double-IPAP experiments are achieved by shifting the frequency offset of the dashed pulses (shaped) of the C_4 and C_6 channels and by adjusting the phase ϕ*, as follows: IP_{C4}/IP_{C6}, shaped pulses at 179.5/179.5 ppm, $\phi^* = y$; AP_{C4}/IP_{C6}, shaped pulses at 151.0/179.5 ppm, $\phi^* = x$; IP_{C4}/AP_{C6}, shaped pulses at 179.5/159.6 ppm, $\phi^* = -x$; AP_{C4}/AP_{C6}, shaped pulses at 151.0/159.6 ppm, $\phi^* = y$. Phase cycling consists of ϕ_1, (x, −x); ϕ_2, (x, x, −x, −x); ϕ_3, (x, x, x, x, −x, −x, −x, −x); ϕ_{rec}, (x, −x, −x, x, −x, x, x, −x).

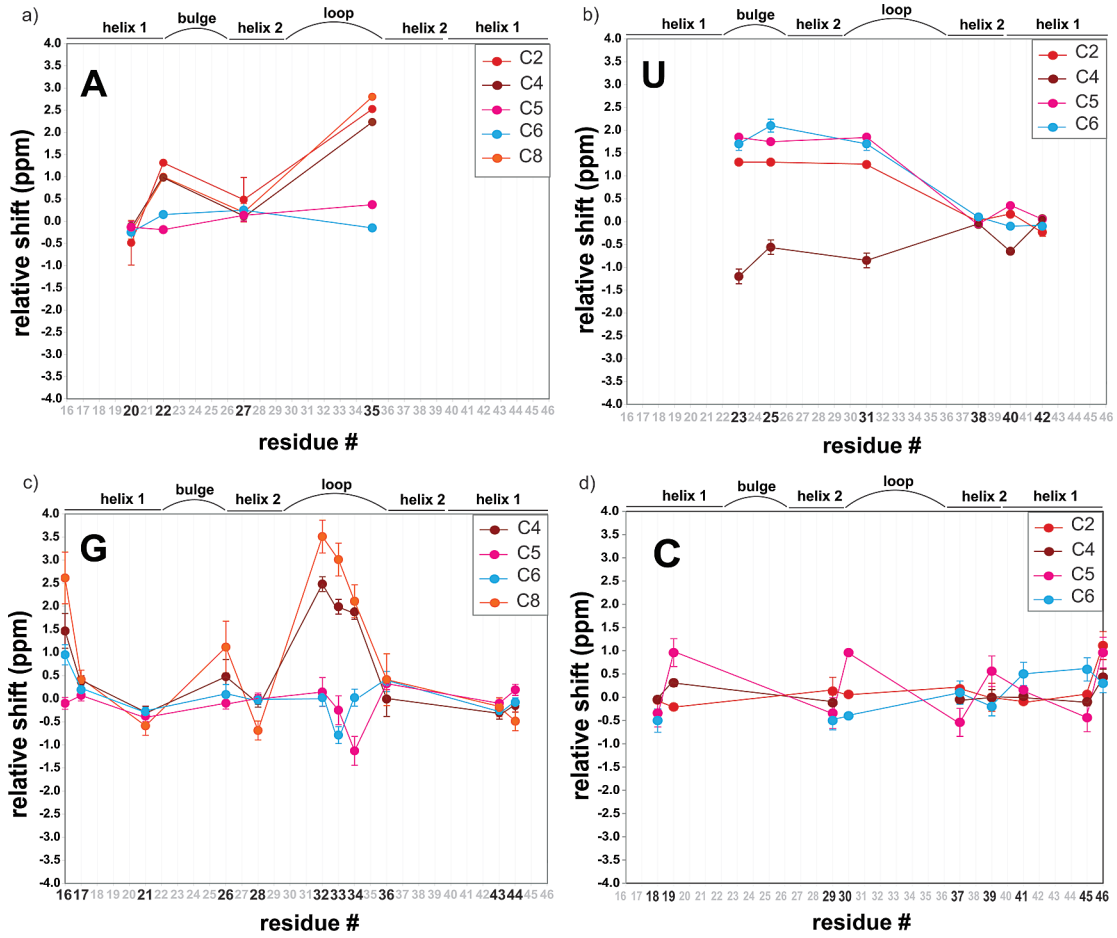


Figure 3. Sequence and residue specific chemical-shift differences measured for all available carbon atoms in HIV-2 TAR-RNA. The values are presented as the difference between the true chemical shift and the mean base-paired, base-stacked chemical shift measured for the same atom and the same nucleotide type from within the stems of HIV-2 TAR-RNA. The standard deviations are calculated separately for residues belonging to the helices, to the bulge/loop and to the terminal helical regions, when at least two values are available. In some cases the standard deviations are so small that cannot be represented. Presented together are the carbon atoms of (a) adenosines, (b) uridines, (c) guanosines, and (d) cytidines. The schemes of the secondary structure refer to Figure 4.

All ^{13}C-detected experiments described here have been designed with the following common features. First, the excited nucleus is a non-exchangeable ^{1}H. Besides the prerequisite of correlating the unprotonated carbon to a known assignment, there is an advantageous sensitivity gain relative to a truly protonless ^{13}C experiment of up to a theoretical factor of $\gamma_H/\gamma_C \approx 4$. Furthermore, the repetition time of such an experiment is dictated by the ^{1}H longitudinal relaxation rate, which is typically smaller than that of carbons. A second common feature is that the detected nucleus is the unprotonated carbon itself. The sensitivity loss from detection of low-γ nuclei is counterbalanced by a number of other advantages. The experiments are typically shorter and simpler since they do not require the relaying of the magnetization back to a potentially distant ^{1}H. This eliminates additional delays and pulses, both sources of magnetization losses due to relaxation and to rf-field inhomogeneity, and offset effects. In particular, as discussed in details in the Results and Discussion section, the number of delays during which the magnetization of the fast-relaxing protons is transversal are minimized, providing a considerable gain in signal-to-noise for large RNAs as compared to proton-detected experiments (Table S1).[14] Using probeheads with cryogenically cooled ^{13}C rf-coil, and high-field magnets, these experiments can be completed in about 12 h each.

Residue-Specific ^{13}C Chemical-Shift Sensitivity in the TAR-RNA. Figure 3 displays the relative chemical shift for each assigned ^{13}C in the bases of HIV-2 TAR-RNA as a function of base position in the known structure (Figure 4). Each shift is represented as the difference of the measured value (Table 1) to the average chemical shift of the same base-type present within the upper and lower stem of TAR-RNA and known to form a stable Watson−Crick-type base-pair and to form a stable π-stack between the preceding and following bases; these exclude the base-pairs at the beginning and at the end of each stem. This chemical shift will be referred to as the mean base-paired, base-stacked chemical shift ($\delta^*(Cx_y)$), and the difference of the observed chemical shifts to $\delta^*(Cx_y)$ is given as $\Delta\delta(Cx_y)$, where x is the carbon position and y the nucleobase-type. In most cases, the structural and dynamic properties of each residue, as detected in the large number of high-resolution NMR and biochemical studies on TAR-RNA in free and bound forms, can account for the chemical shift difference, demonstrating that the chemical shifts of base carbon are indeed good indicators of structural features. To assist with the interpretation of the chemical shift perturbations, the salient properties of the atypical base residues of TAR are summarized in Figure 4. In Figure 5, we report a similar analysis for the base carbon chemical shifts deposited in the BMRB[13] database, as described in the following paragraph.

Analysis of ^{13}C Chemical Shifts in the BMRB. ^{13}C chemical shifts have been deposited in the Biological Magnetic Resonance Data Bank (BMRB)[13] for only 38 RNA sequences with available structures. In our analysis, we excluded the RNA-protein complexes and the RNAs that do not contain long regular A-form helices, such as pseudoknots. The total number of RNA molecules used in this analysis is thus reduced to 29 (ranging in size from 14 to 102 nt). A list of the structures used for this analysis is available in Table S2 of the Supporting Information. A sound statistical analysis to derive the correlation between secondary

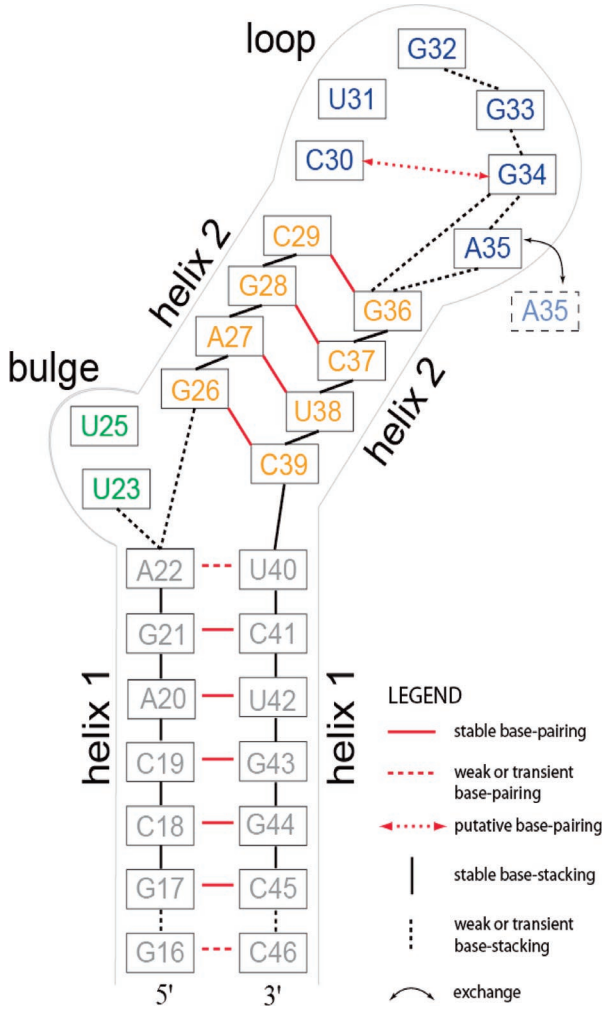


Figure 4. Sequence and secondary structure of the TAR-RNA. The dynamic properties are also summarized. The red color refers to the base-pair properties, whereas solid and dashed lines indicate stable base-pairing and weak or transient base-pairing, respectively. Similarly, black lines describe stable base-stacking, while black dashed lines refer to weak or transient base-stacking. The red and the black double-headed arrows indicate putative base-pairing and exchange, respectively. This figure is based on the structural analysis of several HIV-1 and HIV-2 TAR-RNA structures in the free and bound form from the literature.[36−43]

structure elements and ^{13}C chemical shifts would require a much larger number of data; nevertheless, to support the trends observed for the ^{13}C chemical shifts of residues belonging to differently structured RNA regions in the TAR-RNA, we performed a survey of the BMRB[13] chemical shifts, even though mostly only protonated carbons are reported.

The analysis was performed as described in the following. First, we divided the residues of all structures in the following three groups: (1) helical, when the bases are involved in Watson−Crick-type base pairs and show stable π-stacking between two base-pairs; (2) terminal, if they are located at the terminus of helical regions (stacked only on one side); (3) disordered, for residues belonging to loops or bulges (no base pairing and partial or no stacking). Residues involved in non-canonical base-pairs have not been considered. Second, for each of the three groups, and for each atom type, Cx_y (x = atom number, y = nucleobase), we calculated the difference ($\langle\Delta\delta(Cx_y)\rangle$) between the average value of the carbon chemical shifts belonging to that group and the average value of the carbon chemical shifts belonging to helical regions ($\delta^*(Cx_y)$), as described above for the TAR-RNA. To avoid possible discrepancies in the referencing of the carbon chemical shifts in the different studies, δ^* and $\langle\Delta\delta\rangle$ were calculated independently in each structure. The chemical shift differences for each atom type in each group of all the 29 structures were combined in an overall average and plotted together in Figure 5. The error bars represent the standard deviations calculated for the chemical shifts of each group separately. When only one chemical shift value is present for a certain atom type of a specific group (as for the C6$_{Gua}$ in disordered regions), no standard deviation is reported.

Table 1. ^{13}C Chemical Shifts of Nucleobases in HIV-2 TAR-RNA

chemical shifts[a]		C2	C4	C5	C6	C8
			(ppm)[b]			
G16	helix 1[d]		153.22	118.74	162.39	*139.00*
G17			152.12	118.91	161.63	*136.80*
C18		158.74	168.32	*97.20*	*140.70*	
C19		158.59	168.67	*98.50*	*140.70*	
A20		*152.43/152.44*[e]	148.84	120.97	157.66	*139.30*
G21			151.46	118.47	161.16	*135.80*
A22		*154.16/154.23*[e]	149.93	120.92	158.00	*140.50*
U23	bulge	154.03	168.29	*105.40*	*143.40*	
U25		154.03	168.71	*105.30*	*143.80*	
G26	helix 2		152.23	118.74	161.53	*137.50*
A27		*153.38/153.37*[e]	149.06	121.24	158.09	*139.70*
G28			151.70	118.84	161.42	*135.70*
C29		156.93	168.25	*97.20*	*140.70*	
C30	loop	158.86		*98.50*	*140.80*	
U31		153.98	168.43	105.40	143.40	
G32			154.24	118.97	161.46	*139.90*
G33			153.75	118.59	160.66	*139.40*
G34			153.63	117.71	161.46	*138.50*
A35		*155.12/155.41*[e]	151.18	121.48	157.70	*142.30*
G36	helix 2		151.74	119.17	161.81	*136.80*
C37		158.80	168.31	*97.00*	*141.30*	
U38		152.92	169.21	*103.50*	*141.80*	
C39		158.71	168.36	*98.10*	*141.00*	
U40	helix 1	152.92	168.60	*103.90*	*141.60*	
C41		158.71	168.38	*97.70*	*141.70*	
U42		152.50	169.30	*103.60*	*141.60*	
G43			151.45	118.74	161.18	*136.20*
G44			151.60	119.03	161.36	*135.90*
C45		158.86	168.33	*97.10*	*141.80*	
C46		159.91	168.80	*98.50*	*141.50*	
			δ^* (ppm)[c]			
A		*152.91/152.91*[e]	148.95	121.11	157.88	*139.50*
U		152.73	169.69	*103.55*	*141.70*	
G			151.76	118.84	161.44	*136.39*
C		158.80	168.37	*97.54*	*141.20*	

[a] Reported chemical shifts are relative to the methyl carbons of 2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt (DSS). [b] Chemical shifts in italics belong to protonated carbons and were previously assigned. [c] Mean base-paired, base-stacked, chemical shift. [d] A detailed scheme of the secondary structure is shown in Figure 4. [e] Measured at pD 6.6 and 7.5, respectively.

The analysis used the following number of entries. For protonated carbons: 135 (C2$_{Ade}$), 136 (C8$_{Ade}$), 238 (C8$_{Gua}$), 78 (C5$_{Ura}$), 105 (C6$_{Ura}$), 169 (C5$_{Cyt}$), and 178 (C6$_{Cyt}$). For unprotonated carbons: 13 (C2$_{Gua}$), 8 (C5$_{Gua}$), 13 (C6$_{Gua}$), 13 (C4$_{Cyt}$). Moreover, the analyzed unprotonated carbon chemical shifts for the cytidines come from only two structures, with 5, 6, and 2 elements belonging to the helical, terminal, and disordered groups, respectively, and are therefore of limited significance. Similarly, for the guanosines, the unprotonated data are from eight structures (3, 2, and 3 for C2$_{Gua}$, C5$_{Gua}$ and C6$_{Gua}$, respectively), with a total of 22, 10 and 2 entries in the helical, terminal, and disordered groups, divided in 8, 4, and 1 (helical, terminal, and disordered) values for C2$_{Gua}$; 6, 2, and 0 (helical, terminal, and disordered) values for C5$_{Gua}$; 8, 4, and 1 (helical, terminal, and disordered) values for C6$_{Gua}$.

Furthermore, the graphs presented here should not be understood as an attempt to propose a chemical shift index for RNA, as such a task would require the analysis of a much larger number of data than those available in the BMRB[13] database so far. However, as will be discussed in details in the following section, some of the structure−chemical shift

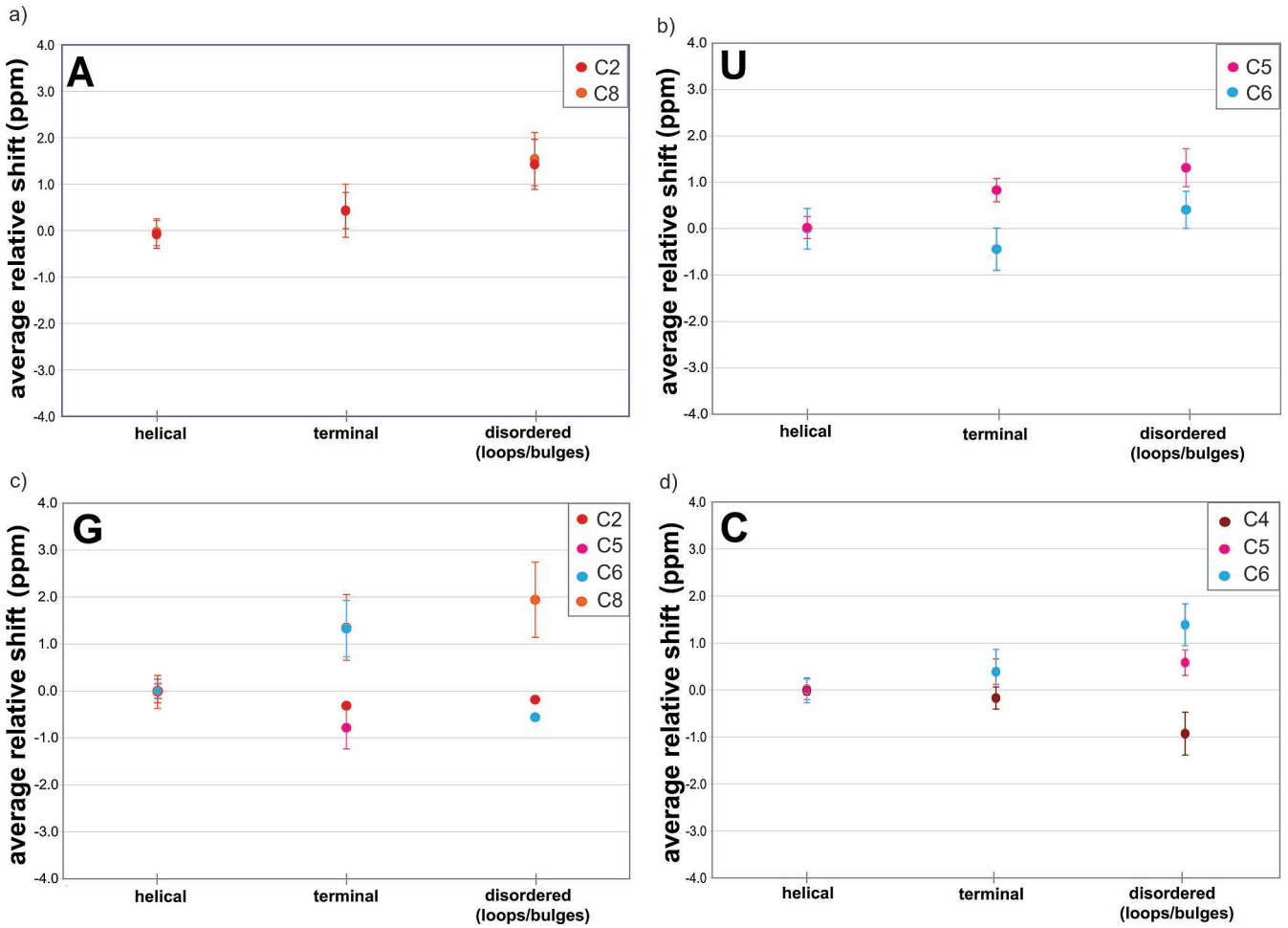


Figure 5. Structural dependence of the ^{13}C chemical shifts of RNA bases available. The individual nucleobases for which chemical shifts are reported in the BMRB[13] are classified into three major groups: (1) "helical", if the bases form Watson−Crick-type base-pairs and stable π-stacking between adjacent residues; (2) "terminal", if they are located at the terminal end of a helical region; and (3) "disordered", for residues belonging to loops or bulges. The average shifts relative to those from helical regions are depicted separately for the different types of RNA bases: (a) adenosine, (b) uridines, (c) guanosine, and (d) cytidines. Residues involved in non-canonical base-pairs, from RNA-protein complexes or from structures lacking a long regular A-form helix, have been not considered in this analysis. The error bars represent the standard deviations in the distribution of relative chemical shifts for each group. When only one chemical shift value is present (as for the $C2_{Gua}$ and $C6_{Gua}$ in disordered regions), no standard deviation is reported.

relationship trends observed for the TAR-RNA are confirmed by the analysis of the chemical shifts of the database.

Results and Discussion

Choice of NMR Experiments for the ^{13}C Assignment of Unprotonated Carbons. The NMR experiments described above were performed on a 1.5 mM sample of ^{13}C- and ^{15}N-uniformly ^{13}C, ^{15}N-labeled HIV-2 TAR-RNA. Selected processed spectra are depicted in Figure 2. These show that for a 30-mer RNA molecule, resolution in a 2D experiment is adequate enough to resolve and assign all targeted ^{13}C chemical shifts. In most cases, a common experiment is recommended to detect both types of purines (A and G) or of pyrimidines (C and U). However, for the H5c5C4 transfer in pyrimidines and for the H8c5C6 transfer in purines (Figure 2c,e), two different experiments are needed to optimize the efficiency for each nucleotide type, owing to the large differences in the $^1J_{C4C5}$ couplings for U and C and in the $^1J_{C5C6}$ couplings for A and G.

The new suite of carbon-detected experiments provides the assignment of the chemical shifts of all unprotonated carbons in RNA nucleobases with the only exception of the $C2_{Gua}$. However, such assignments could also be achieved via experiments that use conventional proton detection. The decision about which experiment to use to measure the chemical shifts of protonless carbons largely depends on the size of the RNA, on the solvent (D_2O vs H_2O) and on the desired correlation. In Table S1, we compare the efficiency of the proton-detected experiments proposed by Fiala et al.[14] with the carbon-detected experiments proposed here for an RNA of the size of ca. 30 nucleotides and for an RNA of double size at 800 MHz. Only experiments correlating the carbons to non-exchangeable protons are considered. The magnetization transfer efficiency for both proton- and carbon-detected experiments takes into consideration the size of the *J*-couplings in the INEPT-like steps and the relaxation losses. Therefore, the sensitivity of the carbon-detected experiments has to be divided by 8 before comparing it to that of the proton-detected experiments owing to the lower gyromagnetic ratio of the carbon nuclei ($(\gamma_C/\gamma_H)^{3/2} = 1/8$). The analysis reported in Table S1 reveals that for RNAs of ca. 30 nt ($\tau_c = 5$ ns) the carbon detected experiments achieving the H8−C4 correlation in purines or the H6−C2 correlations in pyrimidines perform only slightly worse than the proton-detected counterparts. However, for RNAs with $\tau_c = 10$ ns, a signal-to-noise gain is to be expected in the carbon-detected experi-

ments with respect to their proton-detected counterparts for the H8−C4 correlation in purines and the H6−C4 correlation in pyrimidines (Table S1). This gain is due to the reduced number of delays during which the fast-relaxing proton magnetization is in the transverse plane in the carbon-detected experiments. For RNAs with $\tau_c = 15$ ns the carbon-detected experiments are to be preferred to the proton-detected ones in all cases with the only exception of the H5-C5-C4 correlation, as the proton-detected version of this experiment does not suffer from large relaxation losses.

Adenosines. Adenosines in TAR-RNA occur only four times at positions 20, 22, 27, and 35. The A35, which lies at the 3′ extremity of the 6-nt loop, is the only adenosine which does not have a base-pairing partner. In a number of TAR structures, including the free HIV-1 TAR (pdb:1anr)[36] and argininamide-bound HIV-2 TAR (pdb:1aju),[37,50] this base undergoes exchange between two conformations, one where it forms a π-stack with the following base G36 and one where it extends away from the stem letting the preceding G34 stack on top of G36. In complex with synthetic ligands,[42] the equilibrium is shifted to the extended conformation. The features of this irregular base are translated in high chemical-shift values for the $C2_{Ade}$, $C4_{Ade}$, and $C8_{Ade}$ which are, respectively, +2.53, +2.23, and +2.80 ppm above the mean base-paired values for this residue (see Figure 3a). On the other hand the $C5_{Ade}$ and $C6_{Ade}$ seem to be much less sensitive to the lack of regular secondary structure.

Protonation of adenosines can occur at pH values close to 6.5 and dramatically influence the carbon chemical shift values in the base ring.[4] To verify the protonation state of the A35 base in our study, the ^{13}C-^{1}H correlations were followed at pD 6.6 and 7.4. A small chemical shift change of +0.3 ppm occurs for the $C2_{Ade35}$ when increasing the pD by 0.8 unit, indicating that, at pD 6.6, A35 is only transiently protonated. The chemical shift reported in Figure 3 belongs to the unprotonated state of A35 at high pD. Thus, a large effect of protonation on the anomalous chemical shifts of A35 can be safely excluded.

The residue A22 has a Watson−Crick base-pairing partner (U40), although a number of studies have shown that this base-pair is not stable, as indicated by the failure to observe the U40 imino-proton at room temperature. Additionally, the A22 stacks only loosely on the 3′ side with U23 (pdb:1anr) or with G26 (pdb:1aju). This transient base-pairing is also translated in the chemical shift of its five carbons. As for A35, the $C2_{Ade}$, $C4_{Ade}$, and $C8_{Ade}$ are shifted positively but by smaller amounts (+1.45, +0.98, and +1.0 ppm). On the other hand, the other two carbons, $C5_{Ade}$ and $C6_{Ade}$, do not show significant effects on their chemical shift (−0.19 and +0.15 ppm).

Besides the presence of bound ions or cofactors, three dominant structural factors, namely the lack of pairing, the lack of stacking, and the distribution of glysosidic angles can influence the base carbons chemical shifts, whereby the value of the χ angle and the degree of stacking are often correlated. DFT predictions[22] indicate that $C4_{Ade}$ and $C8_{Ade}$ shift *positively* upon breakage of the base-pair, whereas $C2_{Ade}$, $C5_{Ade}$, and $C6_{Ade}$ shift *negatively*. Also, the largest chemical shift changes are expected for $C6_{Ade}$ and $C8_{Ade}$, while $C2_{Ade}$, $C4_{Ade}$, and $C5_{Ade}$ are predicted to be much less sensitive to base-pairing. In addition, the chemical shift differences are more pronounced and in the same direction if base stacking is lacking as well. Some of these DFT figures[22] largely disagree with our observations. Our analysis points to the $C2_{Ade}$, $C4_{Ade}$, and $C8_{Ade}$ as the best indicators for the lack of pairing and stacking, while these factors seem to have opposite, partially canceling effects or even no effect on the $C5_{Ade}$ and $C6_{Ade}$ chemical shifts.

More insight is obtained from the direct comparison of A22 and A35. Residue A22 is likely to stably stack upon G21, while the hydrogen bonds with U40 are only partially formed. The fact that $C2_{Ade}$, $C4_{Ade}$, and $C8_{Ade}$ have smaller shifts in A22 compared to A35 suggests that base-stacking (transiently present for A35 and present only on 5′ side for A22) and base-pairing (absent for A35 and weakened for A22) have cumulative effects on these carbons. For the TAR-RNA in complex with an aminoglycoside analogue,[42] where the A22-U40 base-pair is further weakened with respect to free TAR and the A22 base is partially tilted, both the $C8_{Ade}$ and the $C2_{Ade}$ experience a positive shift with respect to free TAR, strengthening the hypothesis that both base-pairing and base-stacking disruptions shift the $C2_{Ade}$ and $C8_{Ade}$ chemical shifts toward positive values.

A deeper understanding of structure-chemical shift relationships would have to await an exhaustive statistical analysis of the chemical shifts of RNAs with different structures, supported by DFT calculations.[22] From our data we can conclude that $C2_{Ade}$, $C8_{Ade}$, and $C4_{Ade}$ are sensitive to both base-stacking and base-pairing and that the two effects are cooperative. These nuclei can be reliably used to recognize irregular structural regions. A smaller distribution is observed for the chemical shifts of $C5_{Ade}$ and $C6_{Ade}$, due either to opposite effects of base-pairing and base-stacking or to no effect. If an extensive chemical shift analysis of other RNA sequences confirms the first hypothesis (opposite canceling effects of base-pairing and base-stacking on $C5_{Ade}$ and $C6_{Ade}$), these nuclei, in combination with $C2_{Ade}$, $C4_{Ade}$, and $C8_{Ade}$ could be used to distinguish between lack of base-stacking and lack of base-pairing in non-A-form structural region.

The analysis of the protonated carbon chemical shifts reported in the database (Figure 5a) confirms that the $C2_{Ade}$ and $C8_{Ade}$ chemical shifts experience a positive shift by disruption of both base stacking (terminal residues vs helical residues) and base pairing (disordered residues vs terminal and helical residues).

Uridines. Of the 6 uridines present in TAR-RNA, two are located in the bulge (U23 and U25), one in the loop (U31), and three in the upper or lower stem (U38, U40, U42). For all atom types, the chemical shifts are correlated to the position of the uridine in the structure. Indeed, for the unpaired and unstacked Us of the bulge and loop, the $C2_{Uri}$, $C6_{Uri}$, $C5_{Uri}$, and $C4_{Uri}$ are shifted by about +1.5, +2.10, +1.75, and −0.56 ppm, respectively, compared to the Us of the stem (see Figure 3b).

Some insight on the relative importance of base-pairing is shed by U40 of the stem, which forms an unstable base-pair with A22. Despite partial base-pairing disruption, U40 remains stacked between C39 and C41. This base-pair weakening is observed through an important negative $C4_{Uri}$ and slight positive $C2_{Uri}$ and $C5_{Uri}$ chemical shift differences relative to U38 and U42 ($\Delta\delta(C4_{Uri40}) = -0.65$ ppm, $\Delta\delta(C2_{Uri40}) = +0.19$ ppm, $\Delta\delta(C5_{Uri40}) = +0.35$ ppm), but no significant change in the $C6_{Uri}$. These values follow the trends of the Us from the bulge and the loop and suggest a larger effect of base-pairing

(50) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, *257*, 76−80.

disruption on the $C4_{Uri}$ compared to the $C2_{Uri}$, and the protonated $C5_{Uri}$ and $C6_{Uri}$. This observation is in agreement with the fact that for the TAR-RNA bound to an aminoglycoside analogue,[42] where the A22-U40 base pair is further weakened, but U40 remains stably stacked between C39 and C41, the chemical shifts of the $C5_{U40}$ and $C6_{U40}$ do not exhibit any significant change.

In agreement with the trend observed for the TAR-RNA, the analysis of the database chemical shift shows important positive shifts for the $C5_{Uri}$ and $C6_{Uri}$ belonging to residues in disordered regions and minor shifts for the uridine residues located at helix termini (Figure 5b).

Again here, our experimental observations are only partially in agreement with the DFT studies,[22] where $C4_{Uri}$ is expected to be lower by −2 ppm in unpaired and unstacked Us and the $C2_{Uri}$, $C5_{Uri}$, and $C6_{Uri}$ should be largely unaffected. One possible way to reconcile this discrepancy would be to assume that the large changes in the $C2_{Uri}$, $C5_{Uri}$, and $C6_{Uri}$ chemical shifts in the bulge Us are mainly due to changes in the χ angles. This hypothesis is partially in agreement with the large dependence of the $C6_{Uri}$ chemical shift on χ angle values reported in the same study,[22] while the same calculations do not recapitulate the experimental data for $C2_{Uri}$ and $C5_{Uri}$.

Guanosines. Of the 11 guanosines of TAR-RNA, 3 are positioned inside the loop (G32, G33, G34), and 3 are involved in termination base-pairs of the stems (G16, G26, G36), while the other 5 are confined within the stable A-form lower and upper stems (G17, G21, G28, G43, G44). Such a mixture of conformations is convenient to identify the local environmental factors that are best sensed by the observed ^{13}C chemical shifts of the guanosines.

The very dynamic guanosines from the loop exhibit the most prominent ^{13}C chemical shifts of atoms $C8_{Gua}$ and $C4_{Gua}$ (Figure 3c). These are strongly positively shifted for residues G32, G33, and G34, with $\Delta\delta(C8_{Gua})$ varying between +2.11 and +3.51 ppm and $\Delta\delta(C4_{Gua})$ between +1.88 and +2.48 ppm. Furthermore, the gradually decreasing chemical shift differences ($\Delta\delta(C4/8_{Gua32}) > \Delta\delta(C4/8_{Gua33}) > \Delta\delta(C4/8_{Gua34})$) correlate well with the progressive sensitivity to ribonuclease[38] of these Gs and thus with their stacking and base-pairing propensity. Indeed, the G34, G33 and G32 were described to form a transient stacking prolongation of the upper stem with decreasing stability.[38,41] Also, G34 may be involved in a weak base-pair with C30,[43] whereas no base-pairing partners for the G33 and G32 have been described. Interestingly, for the TAR-RNA bound to an aminoglycoside analogue,[42] where the G34 is more stably stacked on the top of G36, the chemical shift of the $C8_{Gua34}$ moves toward that of the stably stacked and paired residues.

Seemingly, the $C4_{Gua}$ and $C8_{Gua}$ chemical shifts are the strongest reporters of dynamics related to base-pairing and base-stacking. The positive shift of the $C8_{Gua}$ chemical shifts upon disruption of base pairing and stacking is confirmed by the analysis of the database chemical shifts (Figure 5).

In many ways, the stem-closing G16, G26, and G36 also exhibit a similar trend for these two chemical shifts. The 5′ terminal G16 shows large $C4_{Gua}$ and $C8_{Gua}$ shifts ($\Delta\delta(C4_{Gua16}) = +1.5$ ppm, $\Delta\delta(C8_{Gua16}) = +2.61$ ppm), which can be attributed to the fact that it is part of a fraying base-pair with C46 and is only stacked on its 3′ side. To a lesser extent, the G26 and G36 also exhibit positive shift differences ($\Delta\delta(C8_{Gua26}) = +1.11$ ppm, $\Delta\delta(C8_{Gua36}) = +0.41$ ppm, $\Delta\delta(C4_{Gua26}) = +0.47$ ppm, $\Delta\delta(C4_{Gua36}) = +0.41$ ppm), which can be recapitulated by the lack of stacking on one side. Our data suggest that as for the adenosines, the effect of base pairing and base-stacking disruption on the C4 and C8 chemical shift is cooperative. Remarkably, the DFT calculations[22] agree well with the strong positive shift of the protonated $C8_{Gua}$, but fail to predict the positive shift of unprotonated C4.

The other two assigned carbons, $C5_{Gua}$ and $C6_{Gua}$, show a much narrower chemical shift distribution; however, the outliers suggest a much more subtle sensitivity to local environment. The largest shifted $C6_{Gua}$ are those of G16 ($\Delta\delta(C6_{Gua16}) = +0.95$ ppm) and G33 ($\Delta\delta(\Delta C6_{Gua33}) = -0.79$ ppm). These divergent trends for two residues, which show similar shifts for $C8_{Gua}$ and $C4_{Gua}$, suggest that there might be opposing influences acting on the $C6_{Gua}$ chemical shift. This hypothesis would explain why G32, one of the most disordered of all guanosine side chain (with no base-pairing and no base-stacking), has a very similar $C6_{Gua}$ chemical shift to those of the stems (G21, G28, G42, G43) and is supported by the analysis of the $C6_{Gua}$ chemical shifts of the database, where a positive shift is observed for the residues at the termini of helical regions and a moderate negative shift for those in disordered loop regions. In contrast, the DFT calculations[22] predict a negative effect for the base-paring disruption and no effect for the stacking disruption on the $C6_{Gua}$.

Among the $C5_{Gua}$ chemical shifts, the single outlier is assigned to the loop residues G34 ($\Delta\delta(C5_{Gua34}) = -1.13$ ppm). The fact that none of the other dynamic loop or stem-closing Gs shows a similar chemical shift difference suggests that $C5_{Gua}$ is either insensitive to conventional base-pairing and stacking or that their effects cancel each other. Unfortunately the lack of additional data for this nucleus does not allow us to confirm or discard any of the two hypotheses.

The four assigned carbons of guanosines exhibit a strong sensitivity on the structural environment and, as for the previous two residue-types, complementary dependencies, which could be used as a robust structural prediction tool for H-bonding, base-stacking and base orientation.

Cytidines. The HIV-2 TAR-RNA is also rich in cytidines; however, these are mostly distributed in the A-form helical stems, and are all involved in a canonical base-pair with the exception C30, positioned at the 5′ end of the hexameric loop. Both C29 and C46 form a termination base-pair of the upper and lower stems and thus only have a single stacking partner on their 5′-side. Concordantly, a full assignment of all cytidine carbon chemical shifts (with the exception of the $C4_{Cyt}$ of C30) show a relatively narrow distribution compared to the other base types (Figure 3d), with no dominant outliers. The apparently wider dispersions of the $C5_{Cyt}$ and $C6_{Cyt}$ chemical shifts are not statistically different from those observed for the ensemble of helical cytidines in the database analysis.

The focus can only be placed on the terminal C46 and C29 and on the single disordered C30. The largest $C2_{Cyt}$ and $C4_{Cyt}$ shifts are +1.11 and +0.43 ppm, respectively, and both belong to the terminating C46. These significant positive shifts for the unprotonated carbons may be influenced by the transient pairing or weakened stacking of this fraying residue. However, the importance of this effect is neither revealed by similar significant

shifts for residue C29 or C30 nor supported by DFT calculations.[22] Here, the absence of data entries in the BMRB[13] for the $C2_{Cyt}$ atoms and the rather negative effect seen on the few $C4_{Cyt}$ (Figure 5d) in disordered regions does not allow any further conclusions. The database chemical shifts show however positive shifts for $C5_{Cyt}$ and $C6_{Cyt}$ in unstructured regions. This result is reproduced in TAR only by the higher $C5_{Cyt30}$ chemical shift. The DFT study[22] predicts instead a negative change of $C5_{Cyt}$ and a positive change of $C6_{Cyt}$ for disordered regions.

In conclusion, because of the poor structural dispersion of cytosines in TAR-RNA, the chemical shift dependencies on structural features cannot be as clearly described as for the other residues types. Nevertheless, we may empirically identify $C5_{Cyt}$ as the best potential indicator of base disorder. Naturally, this assumption and the relevance of other carbon chemical shifts can only be confirmed by a more extensive sampling.

Conclusions

In this article, six new experiments are proposed to directly detect and assign the unprotonated carbons of the base residues of nucleic acids by correlating them to the commonly assigned protons. By means of these experiments, we have been able to assign all base-carbon resonances for the 30-mer HIV-2 TAR-RNA, with the sole exception of the $C2_{Gua}$. The new methods are based on the excitation of ^{1}H and on the direct observation of ^{13}C, and have exquisite sensitivity and resolution at high fields using cryogenically cooled coils and preamplifiers. The novelty of these experiments is that they give straightforward access to a new set of ^{13}C chemical shifts to complement the relatively poor concentration of ^{1}H in the bases and therefore the paucity of structural indicators. These carbons may play important roles in common NMR methodologies such as measurement of RDCs in the base or for the measurement of distances to paramagnetic centers attached to RNA and DNA. With respect to the proton-detected experiments previously proposed to measure the unprotonated carbon chemical shifts,[14,15] the new carbon detected experiments are better suited for large RNAs (>50 nt) because long delays, during which the fast relaxing proton magnetization is transversal, are minimized.

Both unprotonated and protonated carbons chemical shifts have the potential to directly provide easily interpretable structural information about the local environment of the bases. By combining the information of the 10 new types of ^{13}C chemical shifts with those of the typically assigned protonated ^{13}C, chemical shift information could be translated to simple structural restraints such as base-pairing and base-stacking.

Indeed we find a strong correlation between most base-carbon chemical shifts and the structural features of all nucleotide types. In particular the $C2_{Ade}$, $C4_{Pu}$, and $C8_{Pu}$ are very good indicators of irregular structural elements, as seen by a cumulative effect of the disruption of both base-pairing and base-stacking. However, the relative contribution of each of these structural features is difficult to assess on the basis of these chemical shifts alone. Fortunately, the $C5_{Pu}$ and $C6_{Pu}$ nuclei seem to exhibit a partially canceling dependence on base-pairing and stacking and thus may complement the other shifts in the identification of the most dominant structural irregularity. Similarly for uridines, the $C4_{Uri}$ nucleus seems to be a good indicator of base-pairing, while the $C2_{Uri}$ and $C5_{Uri}$ are sensitive to both pairing and stacking. Most of the tendencies in the correlation between secondary structure elements and chemical shifts of the base carbons observed for the TAR-RNA could be confirmed by the analysis of the chemical shifts reported in the BMRB[13] database. An exception to this is the data for the cytidines, which partially exhibit a different behavior for TAR compared to the database.

The combination of the chemical shifts of protonated and unprotonated carbon nuclei, which have either a cooperative or a mutually canceling dependence on the diverse structural descriptors, such as base-pairing and stacking, could provide valuable structural information at each nucleotide site. From these, an approach could be conceived whereby the secondary structure is determined on the basis of a chemical shift "fingerprint" extracted from the collection of available (protonated and unprotonated) ^{13}C values. However, before the chemical shift information can be used for structural determination, the effect of the local structural context of each base on the chemical shifts has to be understood quantitatively either through DFT calculations and/or statistical analyses. Unfortunately, very few unprotonated aromatic carbon chemical shifts in RNA have been reported and deposited into the Biological Magnetic Resonance Data Bank (BMRB[13]) to date, which makes a statistical analysis meaningless. Also the number of protonated carbon chemical shifts contained in the database is far from being sufficient to develop a chemical shift index for RNA. It is our belief that a better understanding of the structure−chemical shift relationships of all base atoms, including the unprotonated carbons, through the expansion of the aromatic chemical shift databanks, will allow the development of RNA and DNA secondary structure determination methods based on chemical shift-mediated structural information. The six new experiments proposed here will hopefully contribute to this project.

Acknowledgment. The authors thank Dr. D. Lee for helpful discussions and C. Schwiegk for the synthesis of TAR RNA. This work is supported by the Deutsche Forschungsgemeinschaft (SFB416 to T.C.) and by the MPG.

Supporting Information Available: Complete ref 1. Comparison of the theoretical absolute sensitivities of the experiments proposed in this work, using ^{13}C detection, and of those proposed from Fiala et al.,[14] using ^{1}H detection. Experiments correlating the same nuclei are directly compared (Table S1). A list of the BMRB[13] entries of the RNA structures used in the ^{13}C chemical shift analysis is given in Table S2. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0727417