

## Supporting Information

### **Direct NMR Evidence that Transient Tautomeric and Anionic States in dG•dT Form**

#### **Watson-Crick like Base Pairs**

Eric S. Szymanski<sup>1</sup>, Isaac J. Kimsey<sup>1†</sup>, and Hashim M. Al-Hashimi<sup>1,2\*</sup>

<sup>1</sup>Department of Biochemistry, Duke University Medical Center, Durham, NC, 27710

<sup>2</sup>Department of Chemistry, Duke University, Durham, NC, 27710

<sup>†</sup>Current Address: Nymirium, Ann Arbor, MI, 48109

### **Supplementary Methods**

#### ***Sample Preparation***

Selectively <sup>13</sup>C/<sup>15</sup>N site-labeled and site-modified DNA oligonucleotides were purchased from the Yale Keck Oligonucleotide Synthesis Facility and were synthesized using commercially available 2'-deoxyguanosine DMT-phosphoramidites (<sup>13</sup>C, 98%, <sup>15</sup>N, 98%) and 2'-deoxythymidine phosphoramidities (<sup>13</sup>C, 98%, <sup>15</sup>N, 98%) purchased from Cambridge Isotope Labs. Cartridge purified oligonucleotides were resuspended in water and exchanged into NMR buffer (25mM sodium chloride, 15mM sodium phosphate, 0.1 mM EDTA, at the desired pH) using an Amicon Ultra-4 (3 kDA cutoff) centrifugal concentrator to a final volume of  $\approx$  250  $\mu$ l. D<sub>2</sub>O was added to a final concentration of 10%.

#### ***NMR experiments***

The chemical shift assignments for all DNA constructs were obtained using imino [<sup>15</sup>N, <sup>1</sup>H], amino [<sup>15</sup>N, <sup>1</sup>H] and aliphatic [<sup>13</sup>C, <sup>1</sup>H] heteronuclear correlation experiments as described previously.<sup>1,2</sup> The data was collected on a 14.1 T Agilent spectrometer equipped with a Bruker

HCPN cryogenic probe. All data were processed and analyzed using the software NMRpipe<sup>3</sup> and SPARKY (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco). 1D <sup>15</sup>N  $R_{1\rho}$  RD experiments targeting imino nitrogen resonances of interest were carried out as previously described.<sup>4-6</sup> Raw data was processed using NMRpipe<sup>3</sup> to generate a series of peak intensities. On- and off- resonances RD profiles were recorded using spinlock powers ranging between 150 to 2000 Hz with the absolute offset ranging between 0-3.5X the applied spinlock power. Magnetization of the spins of interest was allowed to relax under an applied spinlock for 0 to 120ms.

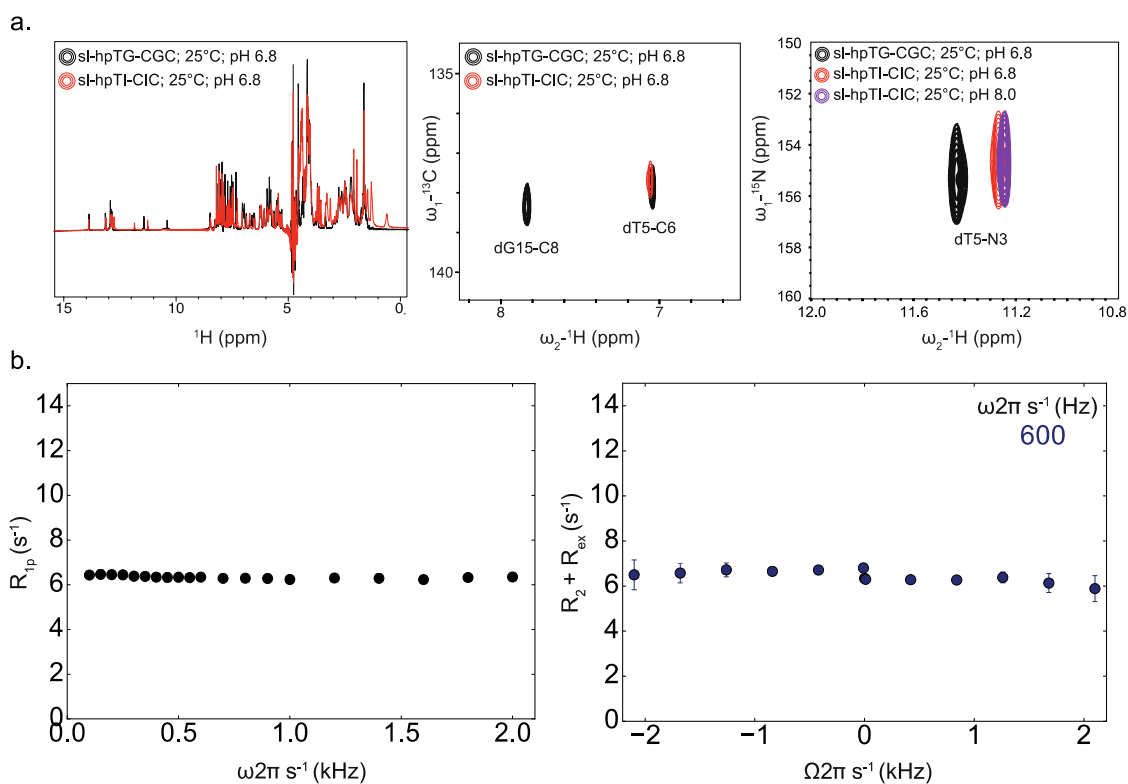
<sup>15</sup>N  $R_{1\rho}$  RD experiments targeting amino dG-N2 resonances were carried out using the same experiment used to measure imino dG-N1 and dT-N3 RD data.<sup>2,6,7</sup> In both dG•dC WC and WB dG•dT, the amino appears as a single NMR resonance due to intermediate-fast exchange of the non-equivalent H-bonded and non-H-bonded protons (dG•dC) or because the amino group is not H-bonded yielding two equivalent protons (dG•dT). Therefore, analogues to imino experiment,<sup>2,6,7</sup> matched <sup>1</sup>H and <sup>15</sup>N fields ( $\approx 90$  Hz) were applied during the cross polarization period to generate <sup>15</sup>N magnetization. In general, to avoid unwanted evolution of double-quantum coherence during CP, the sum of the rf amplitudes needs to be greater than the one-bond N-H scalar coupling ( $^1J_{\text{NH}}$ ) i.e. rf amplitude  $> ^1J_{\text{NH}}/3$ .<sup>8</sup> Note that when calibrating the duration of the cross polarization period for AX<sub>2</sub> systems with equivalent protons, optimal transfer durations will be  $\approx 3/4 \tau$  ( $\tau = 1/|J|$ ) as compared to an optimal transfer duration of  $\tau$  in AX or nonequivalent AX<sub>2</sub> spin systems.<sup>8</sup> For equivalent AX<sub>2</sub> spin system, the magnetization evolution starting from protons H<sub>1x</sub> and H<sub>2x</sub> after a duration of  $3/4 \tau$  generates  $S_x + 4I_{1z}I_{2z}S_x$ . Some magnetization will therefore be lost to the generation of the three spin coherence.<sup>8,9</sup> On-resonance continuous wave <sup>1</sup>H irradiation ( $\omega_{1H} \approx 5000$  Hz) is used to suppress DD/CSA cross-

correlated cross-relaxation as described previously.<sup>10</sup> In addition to DD/CSA interactions, AX<sub>2</sub><sup>11</sup> and AX<sub>3</sub><sup>12</sup> systems can be subject to N<sub>2</sub>H<sub>a</sub>/N<sub>2</sub>H<sub>b</sub> dipole-dipole (DD/DD) cross-correlated cross relaxation and the generation of doubly antiphase magnetization from in-phase magnetization during the relaxation period. Assuming isotropic tumbling, the DD<sub>N<sub>2</sub>H<sub>a</sub></sub>/DD<sub>N<sub>2</sub>H<sub>b</sub></sub> cross-correlation contribution scales as  $\approx \tau_c (1/2)(3\cos^2 \theta - 1)$ ,<sup>13</sup> in which  $\tau_c$  is the correlation time of the molecule and  $\theta$  is the angle between the two bonds. As noted previously, the angle between the N-H bonds in amino NH<sub>2</sub> group ( $\approx 120^\circ$ ), results in a small angular term ( $\approx -0.125$ ) and DD<sub>N<sub>2</sub>H<sub>a</sub></sub>/DD<sub>N<sub>2</sub>H<sub>b</sub></sub> cross-correlation contribution for a  $\approx 10$  bp duplex with  $\tau_c \approx 5$  ns is estimated to be small  $\approx 0.5$  s<sup>-1</sup> and safely be ignored.<sup>13,14</sup> This is supported by lack of any significant bi-exponential behavior in the magnetization decay curves (Figure S2) even when using longer relaxation times (135 ms) or increased sampling (61 delay points) (Figure S2d) and by the observation of an approximately 1:2:1 multiplet in a coupled <sup>15</sup>N/<sup>1</sup>H HSQC (Figure S3).

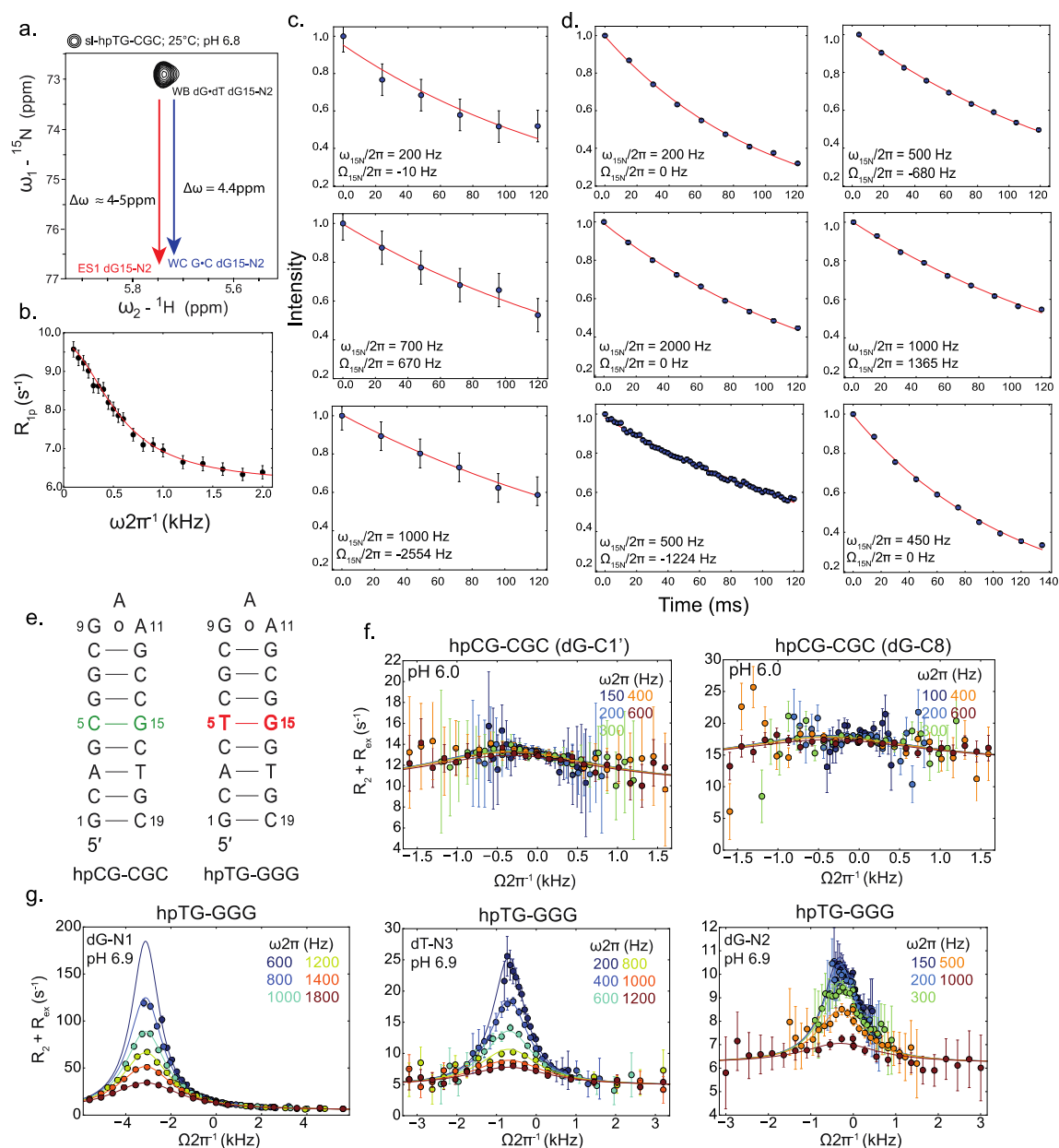
### ***R<sub>1ρ</sub> Data Analysis***

*R<sub>1ρ</sub>* values were calculated by fitting the decay of peak intensity versus relaxation delay to a mono-exponential. Errors in *R<sub>1ρ</sub>* were estimated using spectra noise and duplicate data points. Measured on- and off- resonance *R<sub>1ρ</sub>* data were globally fit to algebraic equations describing chemical exchange<sup>15-17</sup> using a Levenberg-Marquardt method as described previously.<sup>2</sup>

Data was fit assuming two-state or three-state exchange to the two-state or three-state Laguerre equation<sup>15-17</sup> and was also directly fit with two-state or three-state star-like topology exchange using numerical integration of the Bloch-McConnell equations.<sup>2,18</sup> Exchange parameters obtained using Laguerre and Bloch-McConnell were in excellent agreement.

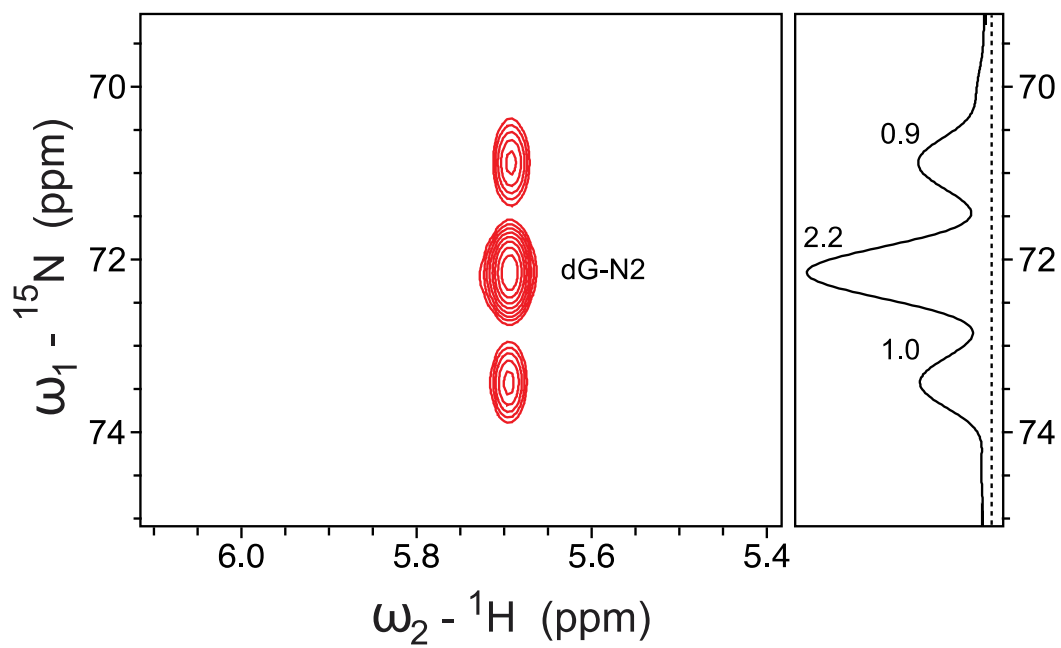


**Supporting Figure 1.** (a) Comparison of the hpTG-CGC NMR spectra with and without the guanine to inosine substitution. Shown are overlays of 1D  $^1\text{H}$  (left), 2D  $^{13}\text{C}/^1\text{H}$  HSQC (center), and 2D  $^{15}\text{N}/^1\text{H}$  HSQC (right) spectra for dG•dT (black) and dI•dT (red and purple). (b) On- (left) and off- (right) resonance RD profiles measured for dT5-N3 in the dI•dT mismatch at 10°C, pH 8.0 showing no signs of chemical exchange.

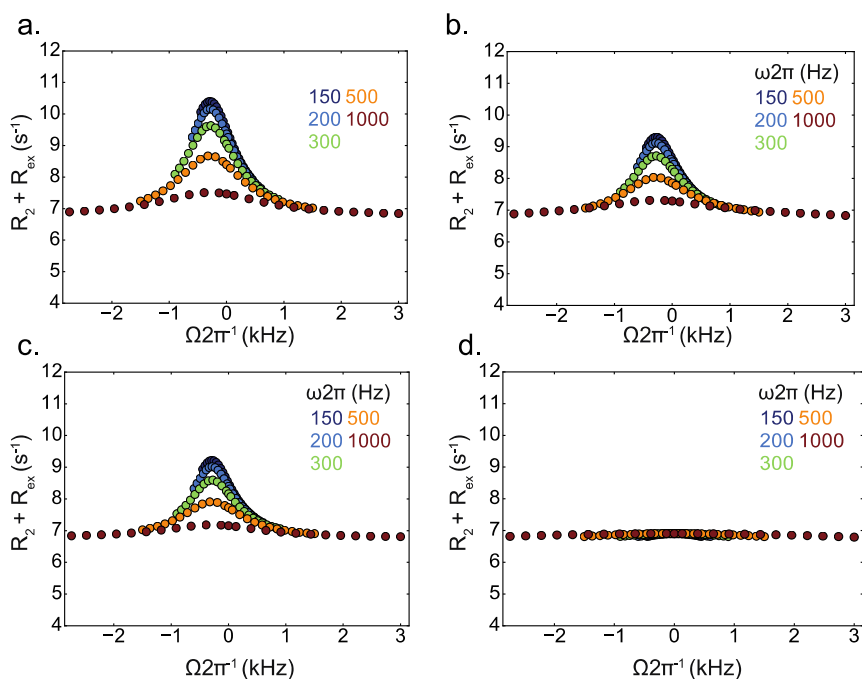


**Supporting Figure 2.** (a) 2D  ${}^{15}\text{N}/{}^1\text{H}$  HSQC spectra of site-labeled dG•dT (hpTG-CGC) showing the two equivalent non-bonded protons of the amino group as a single resonance. Blue arrow indicates the approximate position of dG•dC dG-N2 chemical shift. Red arrow indicates expected change in chemical shift upon transition from non-H-bonded wobble dG•dT to H-bonded ES dG•dT (b) Representative on-resonance RD profile for dG15-N2 measured for the

hpTG-CGC construct at pH 6.9, 25°C with spinlock powers ranging between 0 to 2 kHz. (c) Representative  $^{15}\text{N}$  dG-N2 relaxation decays showing mono-exponential behavior as measured in a dG•dC WC bp in hpCG-CGC. The large uncertainty is due to poor signal to noise arising from exchange broadening of the two motionally averaged guanine amino protons. (d) Representative  $^{15}\text{N}$  dG-N2 relaxation decays measured in hpTG-GGG showing mono-exponential behavior at a high and low spinlock powers with high and low offset values (see inset). Error bars are within symbols. (e) hpCG-CGC and hpTG-GGG DNA constructs used in this study. (f) dG-N2 RD profiles measured in a dG•dC WC bp in hpCG-CGC showing no evidence for chemical exchange due to Hoogsteen bp formation at pH = 6.0.<sup>6</sup> (g) RD profiles measured for dG-N1, dG-N2, and dT-N3 in hpTG-GGG. Solid line indicates global fit to the RD data in which the population ( $p_b$ ) and exchange rate ( $k_{ex}$ ) are shared.



**Supporting Figure 3.** A coupled 2D  $^{15}\text{N}/^1\text{H}$  HSQC showing the triplet splitting pattern of the dG-N2 ( $\text{NH}_2$ ) resonance in the hpTG-GGG construct at pH 8.0, 25 °C. The intensity ratio is 0.9:2.2:1.0.



**Supporting Figure 4.** Bloch-McConnell simulations showing limited contributions to dG-N2 RD profiles due to exchange with anionic ES2. Shown are simulated dG-N2 RD profiles due to exchange with ES1 and ES2 assuming exchange parameters deduced previously<sup>2</sup> based on N1/3 RD data measured in hpTG-GGG and an assumed  $\Delta\omega_{B(dGN2)} = 4.5$  p.p.m. RD profiles are shown for (a) 2-state exchange at low pH = 6.9 with  $p_{ES1} = 0.36\%$ ;  $k_{ex}(GS-ES1) = 3000$  s<sup>-1</sup>;  $\Delta\omega_{ES1} = 4.5$  p.p.m;  $R_1 = 4.7$ ;  $R_2 = 6.5$ . (b) 3-state exchange with star-like topology at high pH = 8.0 with  $p_{ES1} = 0.23\%$ ;  $p_{ES2} = 0.22\%$ ;  $k_{ex}(GS-ES1) = 1500$  s<sup>-1</sup>;  $k_{ex}(GS-ES2) = 50000$  s<sup>-1</sup>;  $\Delta\omega_B = 4.5$  p.p.m;  $\Delta\omega_C = 4.5$  p.p.m;  $R_1 = 4.7$ ;  $R_2 = 6.5$ . The individual contributions to the 3-state RD profiles due to (c) tautomeric ES1 obtained by setting  $p_{ES2} = 0\%$  and (d) anionic ES2 obtained by setting  $p_{ES1} = 0\%$ . The smaller RD observed in (b) as compared to (a) is due to a slightly smaller tautomeric  $p_{ES1}$  at higher pH.



Individual dG-N2 Fits		
Construct	hpTG-CGC; pH 6.9, 25°C	hpTG-GGG; pH 6.9, 25°C
pB (%)	0.239 ± 0.04	0.536 ± 0.037
$k_{\text{ex}}$ (s <sup>-1</sup> )	2486 ± 208	3087 ± 100
$\Delta\omega$ (p.p.m.)	3.81 ± 0.37	4.28 ± 0.17
R1	4.42 ± 0.02	4.3 ± 0.01
R2	6.14 ± 0.05	6.07 ± 0.05
Red. $\chi^2$	0.45	0.37
Global dG-N1, dG-N2, dT-N3 Shared Fits		
Construct	hpTG-CGC; pH 6.9, 25°C	hpTG-GGG; pH 6.9, 25°C
pB (%)	0.17 ± 0.002	0.366 ± 0.004
$k_{\text{ex}}$ (s <sup>-1</sup> )	2743 ± 45	2797 ± 41
$\Delta\omega$ (dG-N1)	36.05 ± 0.15	51.44 ± 0.11
$\Delta\omega$ (dG-N2)	4.85 ± 0.19	5.15 ± 0.08
$\Delta\omega$ (dT-N3)	18.36 ± 0.08	11.4 ± 0.12
Red. $\chi^2$	1.42	0.41

**Supporting Table 1.** Exchange parameters obtained from individual or globally fitting on- and off resonance  $R_{1\rho}$  relaxation dispersion data to the 2-state Laguerre equation as previously described; fitting errors are given as standard error.<sup>2</sup>

Residue (spin)	On-resonance spinlock power ( $\omega$ )/ Off-resonance spinlock power ( $\omega$ ) & $\pm$ {offset ( $\Omega$ )}
T5 (N3) hpTG-CGC	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 200 Hz & $\pm$ {10, 64, 128, 192, 256, 320, 384, 448, 512, 576, 640, 704} 400 Hz & $\pm$ {10, 127, 254, 381, 508, 635, 762, 889, 1016, 1143, 1270, 1397} 600 Hz & $\pm$ {10, 191, 382, 573, 764, 955, 1146, 1337, 1528, 1719, 1910, 2101} 800 Hz & $\pm$ {10, 243, 476, 709, 942, 1175, 1408, 1641, 1874, 2107, 2340, 2573} 1000 Hz & $\pm$ {10, 318, 636, 954, 12721590, 1908, 2226, 2544, 2862, 3180, 3498} 1250 Hz & $\pm$ {10, 398, 796, 1194, 1592, 1990, 2388, 2786, 3184, 3582, 3980, 4378}
T5 (N3) hpTG-GGG	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 200 Hz & $\pm$ {10, 55, 110, 165, 220, 275, 330, 385, 440, 495, 550, 605} 400 Hz & $\pm$ {10, 109, 218, 327, 436, 545, 654, 763, 872, 981, 1090, 1199} 600 Hz & $\pm$ {10, 164, 328, 492, 656, 820, 984, 1148, 1312, 1476, 1640, 1804} 800 Hz & $\pm$ {10, 218, 436, 654, 872, 1090, 1308, 1526, 1744, 1962, 2180, 2398} 1000 Hz & $\pm$ {10, 273, 546, 819, 1092, 1365, 1638, 1911, 2184, 2457, 2730, 3003} 1200 Hz & $\pm$ {10, 327, 654, 981, 1308, 1635, 1962, 2289, 2616, 2943, 3270, 3597}
G15 (N1) hpTG-CGC	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 400 Hz & $\pm$ {10, 109, 218, 327, 436, 545, 654, 763, 872, 981, 1090, 1199} 600 Hz & $\pm$ {10, 164, 328, 492, 656, 820, 984, 1148, 1312, 1476, 1640, 1804} 1000 Hz & $\pm$ {10, 273, 546, 819, 1092, 1365, 1638, 1911, 2184, 2457, 2730, 3003} 1250 Hz & $\pm$ {10, 341, 682, 1023, 1364, 1705, 2046, 2387, 2728, 3069, 3410, 3751} 1500 Hz & $\pm$ {10, 409, 818, 1227, 1636, 2045, 2454, 2863, 3272, 3681, 4090, 4499}
G15 (N1) hpTG-GGG	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 600 Hz & $\pm$ {10, 164, 328, 492, 656, 820, 984, 1148, 1312, 1476, 1640, 1804} 800 Hz & $\pm$ {10, 218, 436, 654, 872, 1090, 1308, 1526, 1744, 1962, 2180, 2398} 1000 Hz & $\pm$ {10, 273, 546, 819, 1092, 1365, 1638, 1911, 2184, 2457, 2730, 3003} 1200 Hz & $\pm$ {10, 327, 654, 981, 1308, 1635, 1962, 2289, 2616, 2943, 3270, 3597} 1400 Hz & $\pm$ {10, 382, 764, 1146, 1528, 1910, 2292, 2674, 3056, 3438, 3820, 4202} 1800 Hz & $\pm$ {10, 491, 982, 1473, 1964, 2455, 2946, 3437, 3928, 4419, 4910, 5401}
G15 (N2) hpTG-CGC hpTG-GGG	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 150 Hz & $\pm$ {10, 41, 82, 123, 164, 205, 246, 328, 369, 410, 451} 200 Hz & $\pm$ {10, 55, 110, 165, 220, 275, 330, 385, 440, 495, 550, 605} 300 Hz & $\pm$ {10, 82, 164, 246, 328, 410, 492, 574, 656, 738, 820, 902} 500 Hz & $\pm$ {10, 136, 272, 408, 544, 680, 816, 952, 1088, 1224, 1360, 1496} 1000 Hz & $\pm$ {10, 273, 546, 819, 1092, 1365, 1638, 1911, 2184, 2457, 2730, 3003}
G15 (N2) hpCG-CGC	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 200 Hz & $\pm$ {10, 55, 110, 165, 220, 275, 330, 385, 440, 495, 550, 605} 500 Hz & $\pm$ {10, 136, 272, 408, 544, 680, 816, 952, 1088, 1224, 1360, 1496} 700 Hz & $\pm$ {10, 185, 360, 535, 710, 885, 1060, 1235, 1410, 1585, 1760, 1935}
G15 (N2) hpTG-CGC; 10°C	100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 2000 Hz/ 600 Hz & $\pm$ {10, 360, 710, 1060, 1410, 1760, 2110}
G15 (C1')	150 Hz & $\pm$ {10, 41, 82, 123, 164, 205, 246, 287, 328, 369, 410, 451} 200 Hz & $\pm$ {10, 55, 110, 165, 220, 275, 330, 385, 440, 495, 550, 605} 300 Hz & $\pm$ {10, 82, 164, 246, 328, 410, 492, 574, 656, 738, 820, 902} 400 Hz & $\pm$ {10, 109, 218, 327, 436, 545, 654, 763, 872, 981, 1090, 1199} 600 Hz & $\pm$ {10, 164, 328, 492, 656, 820, 984, 1148, 1312, 1476, 1640, 1804}
G15 (C8)	100 Hz & $\pm$ {10, 27, 54, 81, 108, 135, 162, 189, 216, 243, 270, 297} 200 Hz & $\pm$ {10, 55, 110, 165, 220, 275, 330, 385, 440, 495, 550, 605} 300 Hz & $\pm$ {10, 82, 164, 246, 328, 410, 492, 574, 656, 738, 820, 902} 400 Hz & $\pm$ {10, 109, 218, 327, 436, 545, 654, 763, 872, 981, 1090, 1199} 600 Hz & $\pm$ {10, 164, 328, 492, 656, 820, 984, 1148, 1312, 1476, 1640, 1804}

**Supporting Table 2.** On/off-resonance parameters used to collect  $R_{1\rho}$  relaxation dispersion profiles.

1. Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*, 2<sup>nd</sup> Edition; Academic Press: Burlington, MA, 2008.
2. Kimsey, I. J.; Petzold, K.; Sathyamoorthy, B.; Stein, Z. W.; Al-Hashimi, H. M. *Nature* **2015**, *519*, 315.
3. Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
4. Hansen, A. L.; Nikolova, E. N.; Casiano-Negroni, A.; Al-Hashimi, H. M. *J. Am. Chem. Soc.* **2009**, *131*, 3818.
5. Korzhnev, D. M.; Orekhov, V. Y.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127*, 713.
6. Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. *Nature* **2011**, *470*, 498.
7. Nikolova, E. N.; Gottardo, F. L.; Al-Hashimi, H. M. *J. Am. Chem. Soc.* **2012**, *134*, 3667.
8. Pelupessy, P.; Chiarparin, E. *Concepts Magn. Reson.* **2000**, *12*, 103.
9. Chiarparin, E.; Pelupessy, P.; Bodenhausen, G. *Mol. Phys.* **1998**, *95*, 759.
10. Korzhnev, D. M.; Skrynnikov, N. R.; Millet, O.; Torchia, D. A.; Kay, L. E.; *J. Am. Chem. Soc.* **2002**, *124*, 10743
11. Mulder, F. A. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 967.
12. Baldwin, A. J.; Kay, L. E. *J. Biomol. NMR* **2012**, *53*, 1.
13. Kumar, A.; Royappa, G.; Madhu P. K. *Prog. Nucl. Magn. Reson. Spectrosc.* **2000**, *37*, 191.
14. Ernst, M.; Ernst, R. R. *J. Magn. Reson.* **1994**, *110*, 202.
15. Miloushev, V. Z.; Palmer III, A. G. *J. Magn. Reson.* **2005**, *177*, 221.
16. Palmer III, A. G.; Massi, F. *Chem Rev.* **2006** *106*, 1700
17. Trott, O.; Palmer III, A. G. *J. Magn. Reson.* **2004**, *170*, 104.
18. Xue, Y.; Kellogg, D.; Kimsey, I. J.; Sathyamoorthy, B.; Stein Z.; McBrairty, M; Al-Hashimi, H. M. *Methods Enzymol.* **2015**, *559*, 39.