Supporting Information

Geometric approximation: a new computational approach to characterize protein dynamics from NMR adiabatic relaxation dispersion experiments

Fa-An Chao*, R. Andrew Byrd*

Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD, 21702-1201, USA.

Address correspondence to: fa-an.chao@nih.gov or byrdra@mail.nih.gov

Geometric approximation method (detailed implementation)

A fundamental problem in extracting discrete dynamics parameters from NMR relaxation data is the requirement to have analytical solutions to the complex behavior of spin dynamics. Analytical solutions have been derived for CPMG methods[1,2]; however, these enable description of a rather limited range of timescales (10^2~10^3 sec^{-1}). The same situation also occurs for conventional $R_{1\rho}$ experiments[3,4], although they are more sensitive to faster dynamics (10^3~10^4 sec^{-1}). Adiabatic $R_{1\rho}$ and $R_{2\rho}$ NMR experiments are sensitive to exchange processes over an extremely wide range of dynamics (10^2~10^5 sec^{-1})[5]; however, no analytical solutions exist to extract the dependence on kinetic parameters in all time regimes. We propose a novel solution to this problem by building approximate 6-dimensional solution surfaces of adiabatic $R_{1\rho}$ and $R_{2\rho}$ rates, which are defined on a physically meaningful domain (offset, k_{ex}, $\Delta\omega$, p_a, R_1, and R_2). The 6-dimensional solution surfaces are built on a library of solution points, which are analogous to “lookup-tables” of solutions. Once the solution surfaces are built, the dynamic parameters can be determined by a Monte Carlo search of the surfaces to extract the set of kinetic parameters that match the observable adiabatic $R_{1\rho}$ and $R_{2\rho}$ rates. Significantly, the same solution surfaces can be used for data analysis of any biological sample, provided the same adiabatic pulses are used.

First of all, the proper library is calculated using a multi-processor cluster. Based on the concept described in the main text, each 6-dimensional solution surface can be

decomposed into three 4-dimensional surfaces (f(), g(), and h() in equation [2] in the main text). Therefore, the grid is placed on the four dynamic parameters (offset, k_{ex}, Δω, and p_a). After analyzing the relaxation dispersion profiles with respect to different dynamic parameters for our HARD experiments, the grid spacing in the offset dimension is chosen as 0.5k s^{-1}, the grid spacing in the k_{ex} dimension (on the logarithmic scale) is chosen as $x10^{0.1}$, the grid spacing in the Δω dimension is chosen as 0.2k s^{-1}, and the grid spacing in the p_a dimension is chosen as 5%. The solution libraries consist of 41 points in the offset dimension (spanning the domain of -10k to 10k s^{-1}), 41 points in the k_{ex} dimension (spanning the domain of 0.1 to 1000k s^{-1}), 81 points in the Δω dimension (spanning the domain of -8k to 8k s^{-1}), and 10 points in the p_a dimension (spanning the domain of 55 to 100%). Each solution point in the library was calculated as the effective relaxation rate of the bulk magnetization (A+B) after a single train of the composite adiabatic pulses, and precise numerical integration was performed on the Bloch-McConnell equation using Euler's method including the second order correction term. The third order correction term was used to estimate the deviations of the numerical integration from the true values, which are less than 0.00001% for $R_{1\rho}$ rates and less than 0.001% for $R_{2\rho}$ rates in average. Computation of the library required approximately 22 days, using a 128-core Linux cluster computer system. The final library consists of approximately 20 million points for each of the adiabatic $R_{1\rho}$ and adiabatic $R_{2\rho}$ relaxation rates. This 40 million point library is approximately 660 MB in size.

After the library is calculated, a module can be built to approximate a higher resolution on the whole surface based on the library. Following the concept described in the main text, the approximation can be accomplished with locally defined polynomial functions. Because the two slope functions (f() and g()) for intrinsic relaxation rates (R_1 and R_2) are very smooth with respect to the four dynamic parameters (offset, k_{ex}, Δω, and p_a), the two 4-dimensional surfaces (f() and g()) are approximated with zero order polynomial functions. But the third 4-dimensional surface (h()) will be approximated with second order polynomial functions based on the following general equation.

$$R(x_1, x_2, x_3, x_4) = a_{00} + a_{n0}x_n + \cdots + a_{nm}x_n x_m + \cdots \qquad [1]$$

$$a_{00} = R(0,0,0,0)$$

…

$$a_{n0} = 0.5 \cdot [R(0,0,1,0) - R(0,0,-1,0)] \quad if\ n = 3$$

…

$$a_{nm} = 0.5 \cdot [R(0,0,1,0) + R(0,0,-1,0)] - R(0,0,0,0) \quad if\ n = m = 3$$

$$a_{nm} = R(0,1,1,0) - R(0,1,0,0) - R(0,0,1,0) + R(0,0,0,0) \quad if\ n = 3, m = 2$$

…

The solution surfaces, built by this method, were compared with the simulated solution surfaces (directly calculated by numerical integration of the Bloch-McConnell equation): for 10,000 randomly chosen points, the average deviation of the approximated value from the simulated surface is 0.05% and 0.06% for adiabatic $R_{1\rho}$ and $R_{2\rho}$ rates, respectively (**Tab. 1**).

This geometric approximation to the complete solution of the Bloch-McConnell equation forms the basis of the acronym for our approach: geoHARD (geometric approximation on Heteronuclear Adiabatic Relaxation Dispersion).

Finally, we designed an algorithm, which combines geometric approximation with Monte Carlo sampling, to extract information on spin dynamics at slow (μs-ms) time scales. Because the intrinsic R_1 rates are determined by experiments and the offset values are known from the spectra, only k_{ex}, $\Delta\omega$, p_a, and intrinsic R_2 rates will be treated as independent variables. Therefore, adiabatic $R_{1\rho}$ and $R_{2\rho}$ relaxation rates under different adiabatic pulses can be rapidly calculated for an arbitrary set of dynamic parameters (k_{ex}, $\Delta\omega$, p_a, and R_2) using the geometric approximation. The algorithm aims to minimize the distance function between the observed relaxation rates and the calculated ones by varying the dynamic parameters. Once the minimum is reached, a set of the dynamic parameters giving relaxation rates closest to the observed ones will be reported. In order to achieve the global minimum, 100 different runs of Monte Carlo sampling[6] from randomized initial dynamic parameters are performed for a single calculation, and each run uses simulated annealing containing 10,000 steps of the random grid search combined with 5,000 steps of random-walk. (Random-walk: both direction and step size are randomized during the sampling.) Thus, in each calculation, 1.5 million solution points are sampled in a solution surface for a given data point. Among 100 results from 100 different runs of Monte Carlo sampling, the 10 results with the best scores were used for statistical analysis. For the case of data from two magnetic fields, a complete data set contained 20 experimental data points (relaxation rates using HS1, HS2, HS4, HS6, and HS8 pulses for both adiabatic $R_{1\rho}$ and adiabatic $R_{2\rho}$ experiments at two magnetic field strengths). In the analysis by geometric approximation, 30 million solution points were sampled for each data set, requiring only 5 minutes using a single-core Linux computer. By comparison, in the analysis by numerical integration, each solution point for an adiabatic $R_{1\rho}$ ($R_{2\rho}$) rate requires 1 sec (5 sec) leading to a total equivalent time of 1.5 million minutes to perform such massive sampling using a single-core Linux computer. Therefore, it is an impossible task to analyze the relaxation data by numerical integration without utilizing supercomputing resources.

In order to test the new algorithm, we simulated 300 sets of relaxation data of adiabatic $R_{1\rho}$ and adiabatic $R_{2\rho}$ experiments with randomized dynamic parameters using the

Bloch-McConnell equation. The simulated relaxation data (adiabatic $R_{1\rho}$, adiabatic $R_{2\rho}$, and R_1 rates) were analyzed using the new algorithm to extract the dynamic parameters and compare to the input values. These tests reveal that data sets from a single magnetic field are sufficient to obtain accurate results *in the absence of errors* (**Fig. 2**). Furthermore, in the presence of random errors, the algorithm provides accurate and reliable dynamic parameters if data sets from two magnetic fields are available (**Fig. 3**), and adiabatic relaxation data from more than two magnetic fields continues to improve the precision and accuracy of the fit results. Since adiabatic $R_{1\rho}$ and $R_{2\rho}$ experiments exhibit differential sensitivity to exchange time regimes, the algorithm provides better discrimination among conformational dynamics at different time scales. In our tests, the higher precision is equivalent to the higher accuracy in the fit k_{ex} values (**Fig. 3a**, **3d**, and **3g**).

For most of this discussion, it has been assumed that the experimental system resides only in the fast exchange regime and a resonance in the two-dimensional spectrum represents the total magnetization (A+B). However, it may occur that the system is in the slow exchange regime ($k_{ex} << \Delta\omega$) with a small population of state B, which is difficult to observe. In this case, experimentally, only the relaxation dispersion of the main component (A) is measured, and the solution surfaces for the decay of the total magnetization (A+B) become impractical. Therefore, another set of the libraries and solution surfaces were built based on the same procedures but for the case of slow exchange and detection of only the A component. The same tests are carried out to show that accurate and reliable dynamic parameters can be extracted by the new set of solution surfaces (**Fig. S11**). The relaxation dispersion of the A component still carries information about dynamics in the system based on our simulated tests (**Fig. S11a**, **S11b**, and **S11c**). In the real application, a given experimental data set will be separately analyzed by these two sets of solution surfaces, and the best-fit results will be selected to provide accurate results (**Fig. S12**).

The library and the script for data analysis will be available on request from the authors.

Optimization of proton-decoupling scheme

A density matrix analysis was used to evaluate the effects of N-H^N spin coupling and the efficiency of radio-frequency (RF) decoupling schemes to eliminate the coupling and yield an effective single spin system[7]. Here, we evaluated the behavior of the spin system via a quantum mechanical simulation, in which we expanded the density matrix from the conventional (16X16) matrix to a (32X32) matrix to include consideration of chemical exchange effects and all possible relaxation channels. Our expanded density matrix (32X32) includes the effects of RF pulses, chemical shift evolution, two-site

exchange, J-coupling, and all secular terms of dipole-dipole and chemical shift anisotropy relaxation[8,9]. Cross-correlated relaxation channels are included based on Redfield's theory[10]. Simulations, showing evolution of the density matrix, demonstrate that the effects of proton coupling on ^{15}N relaxation during adiabatic pulses are relatively small when the molecular size is small (τ_c=5ns, **Fig. S3a,b**). However, deleterious effects become significant when the size of the molecule increases (τ_c=30ns, **Fig. S3c,d**). The effects of proton coupling in our simulation are consistent with the results in previous studies on different types of relaxation experiments[7,11-13]. Therefore, it is essential to incorporate an effective proton-decoupling scheme in order to obtain accurate measurement of adiabatic $R_{1\rho}$ and $R_{2\rho}$ rates using the HARD experiment.

In order to remove the effects of proton coupling, the original HARD sequence was modified by including two different types of proton decoupling schemes: CW decoupling and π-pulse decoupling (**Fig. 4**). First, the effects of CW proton decoupling at different RF power levels were tested for the on-resonance condition. Density matrix simulations suggest that either low power (γB_1=1kHz) or high power (γB_1>6kHz) CW proton decoupling (on-resonance) is sufficient to remove the effects of proton coupling without introducing additional relaxation dispersion (**Fig. S4**). The simulated results are supported by experimental observations (**Fig. S5**). However, low power CW proton decoupling is sensitive to off-resonance effects (**Fig. S6a,b**), which can be easily observed in the case of large biomolecules. Although high power decoupling partially ameliorates this problem, it is not suitable for use with cryogenic NMR probes. Next, we examined the incorporation of two π pulses bracketing each composite adiabatic pulse to refocus the effects of proton coupling. Simulations predict that this π-pulse decoupling is insensitive to off-resonance effects when a hard π-pulse (γB_1=20 kHz) is used (**Fig. S6c,d**), and the prediction was supported by experiment (**Fig. S7**). Consequently, based on simulations and experimental validation, π-pulse decoupling is superior to the CW decoupling and effective across the total adiabatic spin-lock periods.

The modified pulse sequences (in Bruker format) will be available on request from the authors.

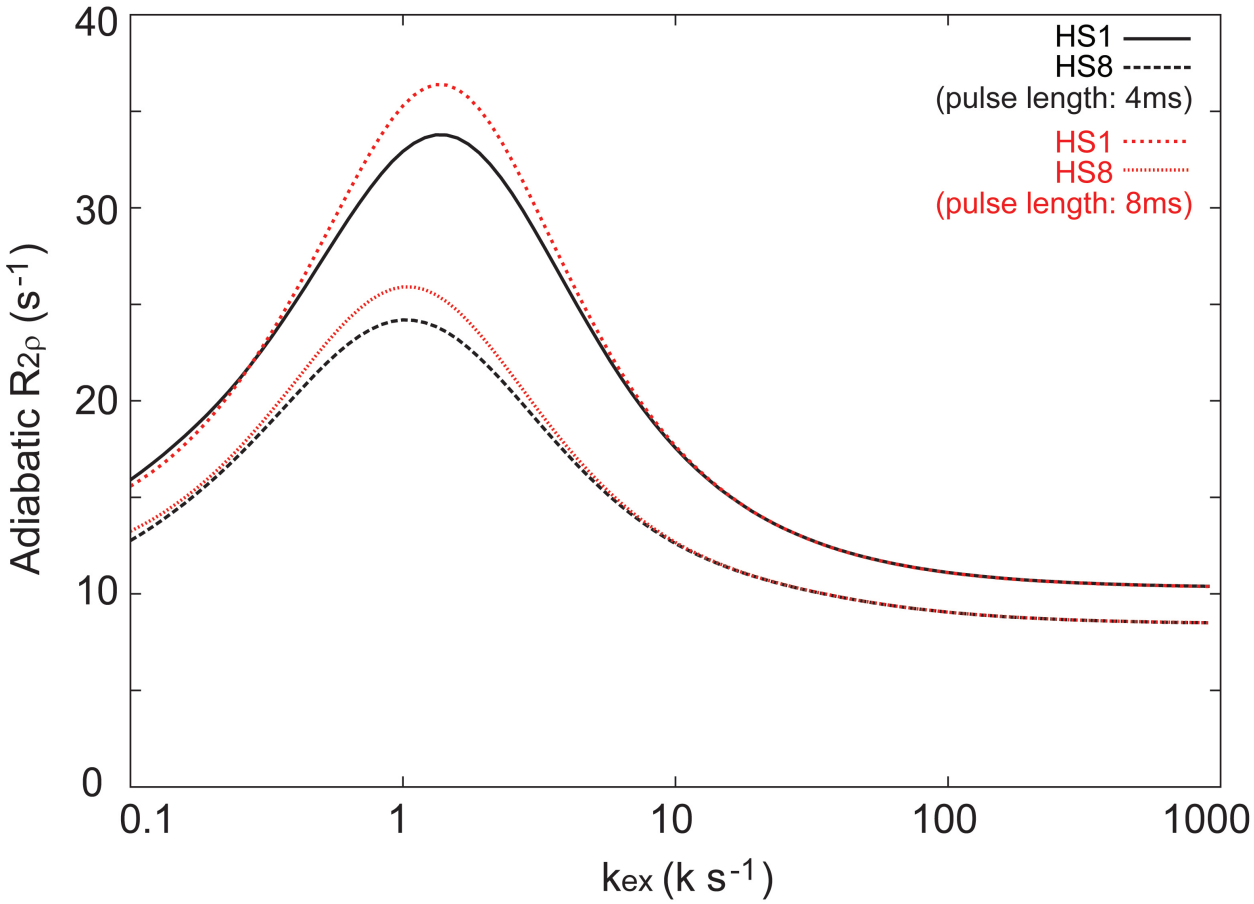

Figure S1. Simulated relaxation dispersion profiles of the adiabatic $R_{2\rho}$ experiment with two different adiabatic pulses (HS1 and HS8) and two different pulse repetition frequencies (4ms: 125Hz and 8ms: 62.5Hz).

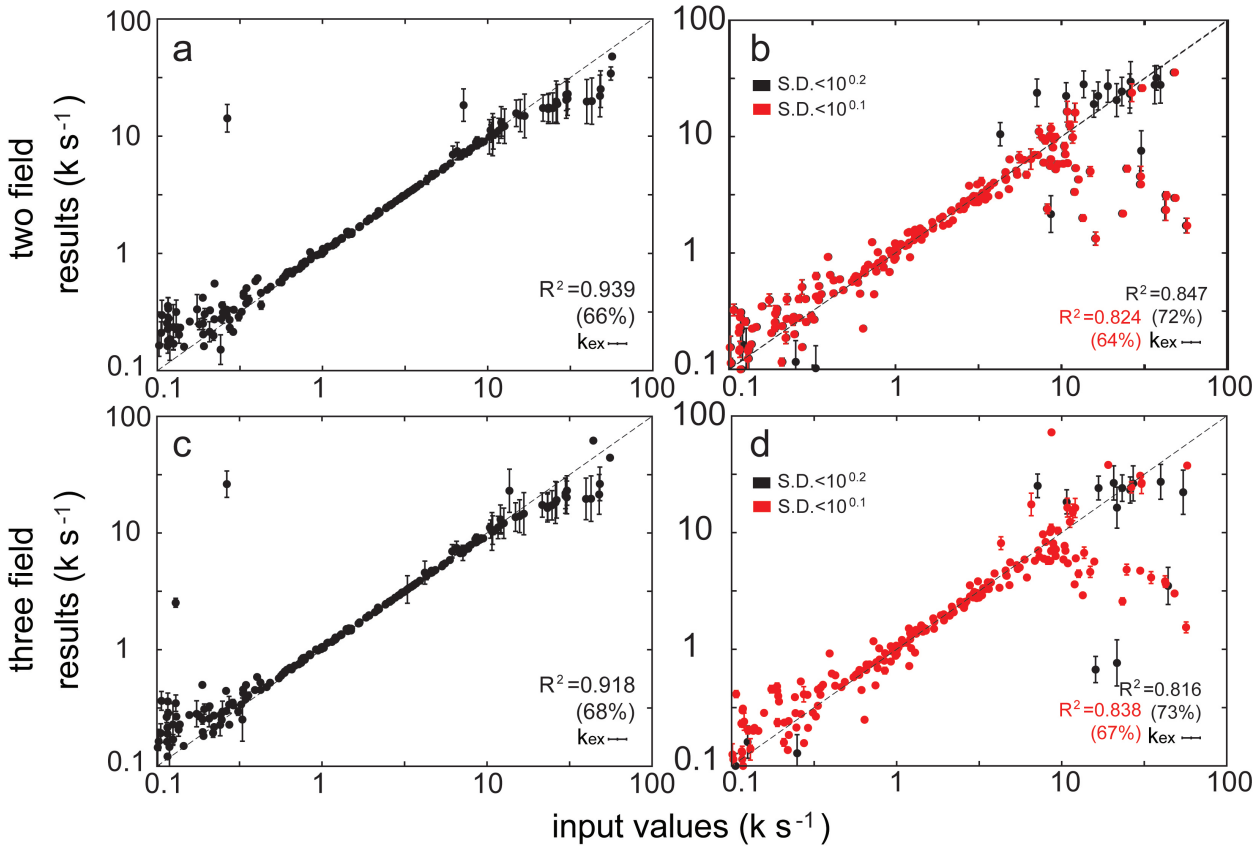

Figure S2. Fitting the simulated relaxation data of constant-time CPMG experiments without (a,c) or with (b,d) up to 5% random errors at two (14.1T, and 18.8T) or three (14.1T, 16.5T, and 18.8T) magnetic fields using approximate analytic solutions by NESSY[23]. The data of different CPMG frequencies (vCPMG = 25, 50, 100, 125, 250, 500, 1000 Hz) were simulated using Bloch-McConnell equation with the same set of dynamic parameters in **Fig.2**. The fit results are plotted against the input values, and those with large standard deviations (S.D. of $k_{ex} > 10^{0.2}$) during Monte Carlo sampling are not shown. The red data points are those with smaller standard deviations (S.D. of $k_{ex} < 10^{0.1}$). The coefficient of determination (R^2) is calculated for each dynamic parameter. The numbers in the parentheses are the percentages of the data remained after filtering out the results with large standard deviations.

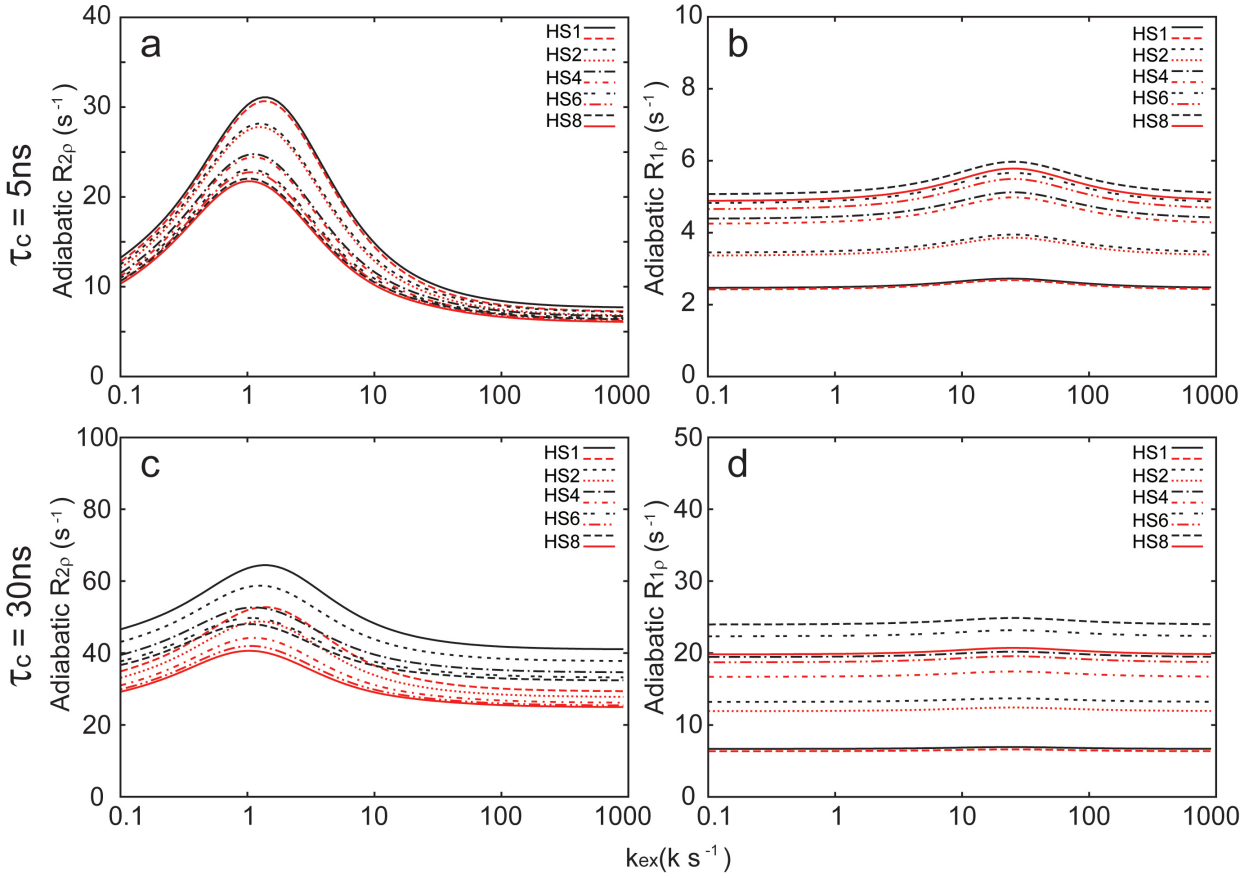

Figure S3. Effects of proton coupling on the ^{15}N spin relaxation during adiabatic pulses. The relaxation rates during adiabatic pulses (HS1, HS2, HS4, HS6, HS8) without proton coupling are plotted in black, and the rates with proton coupling are shown in red. The results were simulated using the expanded density matrix with two different rotational correlation times (τ_c=5ns: a, b; τ_c=30ns: c, d). The dynamic parameters in the chemical exchange were set as $\Delta\omega$ = 240 Hz and p_a = 0.96.

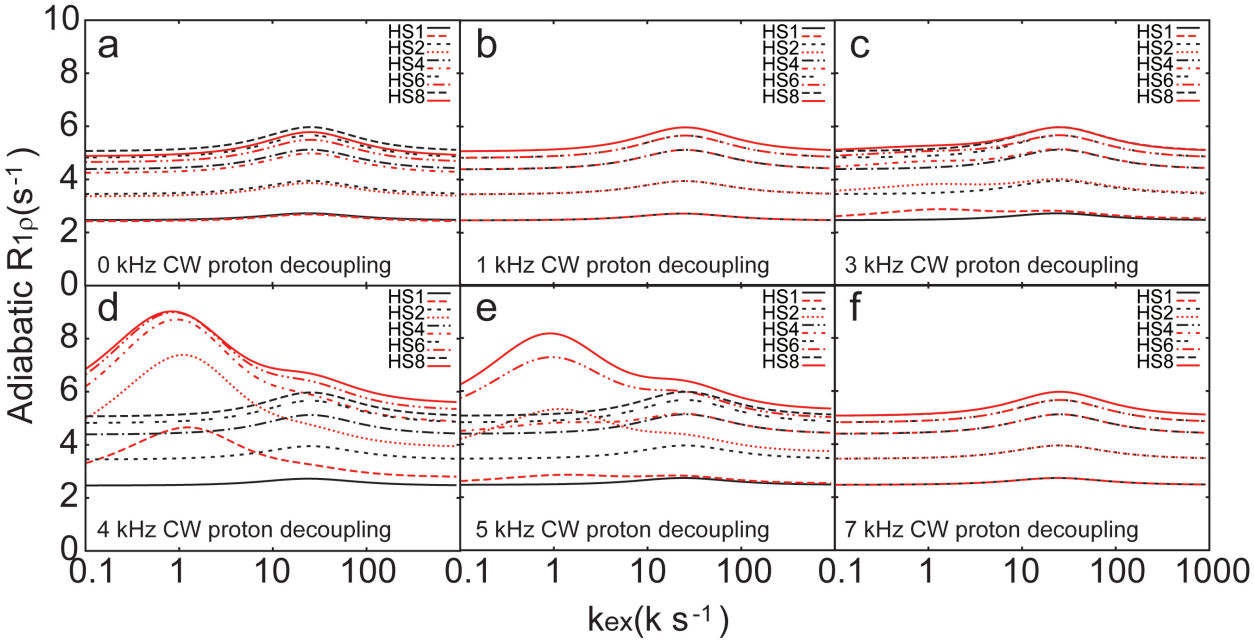

Figure S4. Effects of power levels of the on-resonance CW proton decoupling on adiabatic relaxation dispersion experiments. The adiabatic $R_{1\rho}$ relaxation rates without proton coupling are plotted in black, and the rates with proton coupling and the CW decoupling scheme are shown in red. Different power levels of CW decoupling scheme were tested, and the results were simulated using the expanded density matrix. The dynamic parameters were set as $\Delta\omega$ = 240 Hz, p_a = 0.96, and τ_c = 5ns.

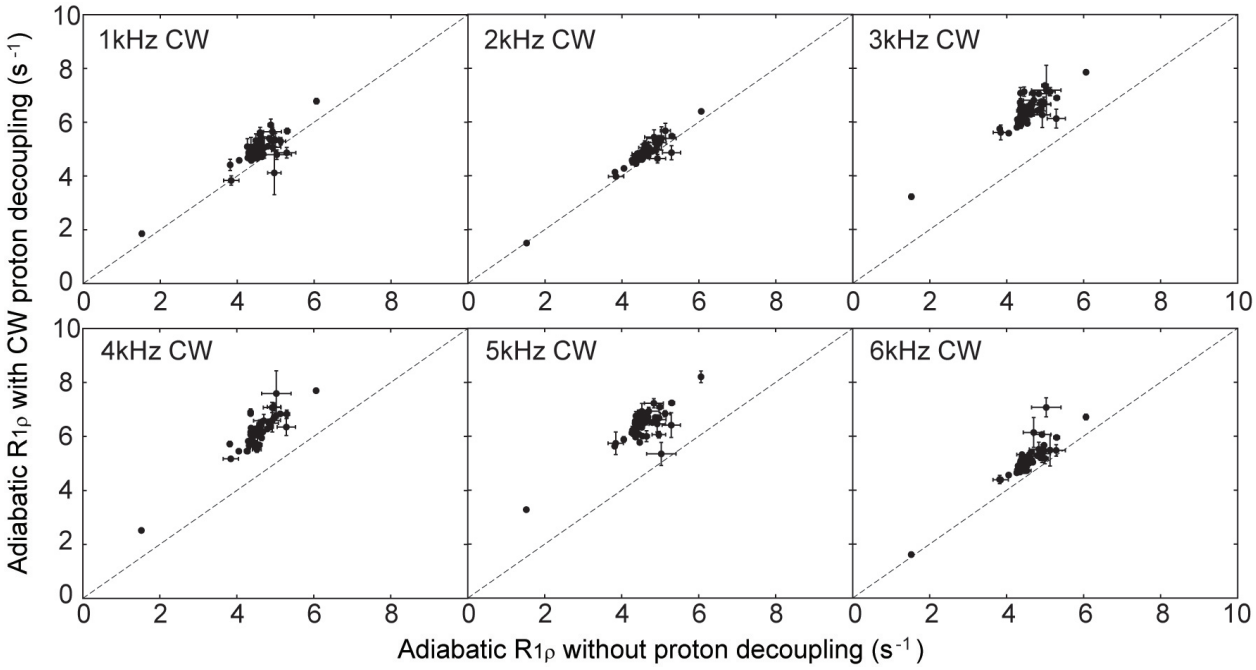

Figure S5. Experimental tests of different power levels of the CW proton decoupling scheme. The ^{15}N, ^{2}H labeled Ub14 sample (25 ºC) was used for the tests, and the $R_{1\rho}$ relaxation rates obtained using the adiabatic pulse HS8 were measured at 600MHz. The relaxation rates under different power levels of the CW decoupling scheme were compared with the ones without proton decoupling.

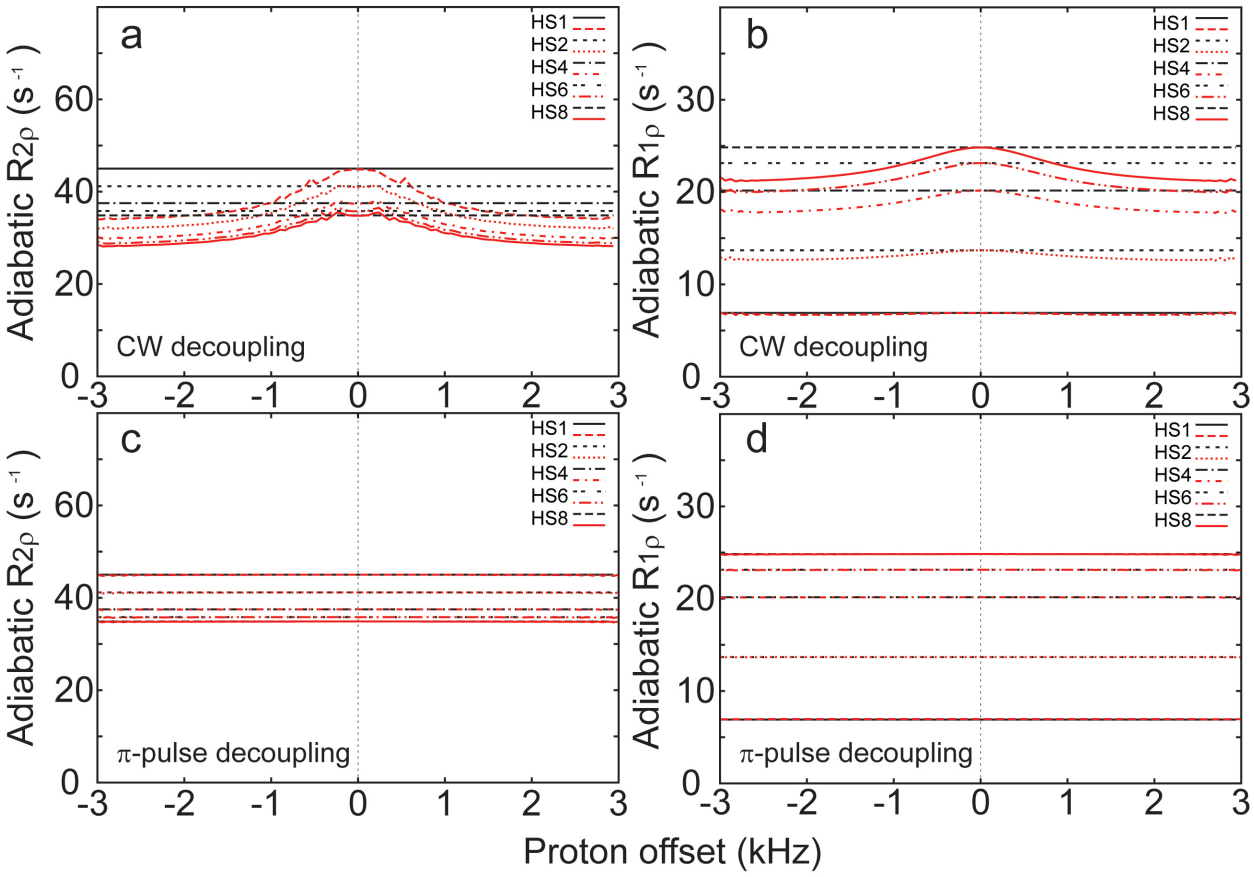

Figure S6. Off-resonance effects for different decoupling schemes. The CW decoupling field was set to be 1 kHz, and the π-pulse corresponded to a γB_1 field of 20 kHz. The adiabatic relaxation rates without proton coupling are plotted in black, and the rates with proton coupling and a given decoupling scheme are shown in red. The off-resonance effects on the spin inversion by the π-pulse were also considered. The results were simulated by the expanded density matrix, and the dynamic parameters were set as $\Delta\omega$ = 240 Hz, kex = 3 kHz, p_a = 0.96, and τ_c = 30ns.

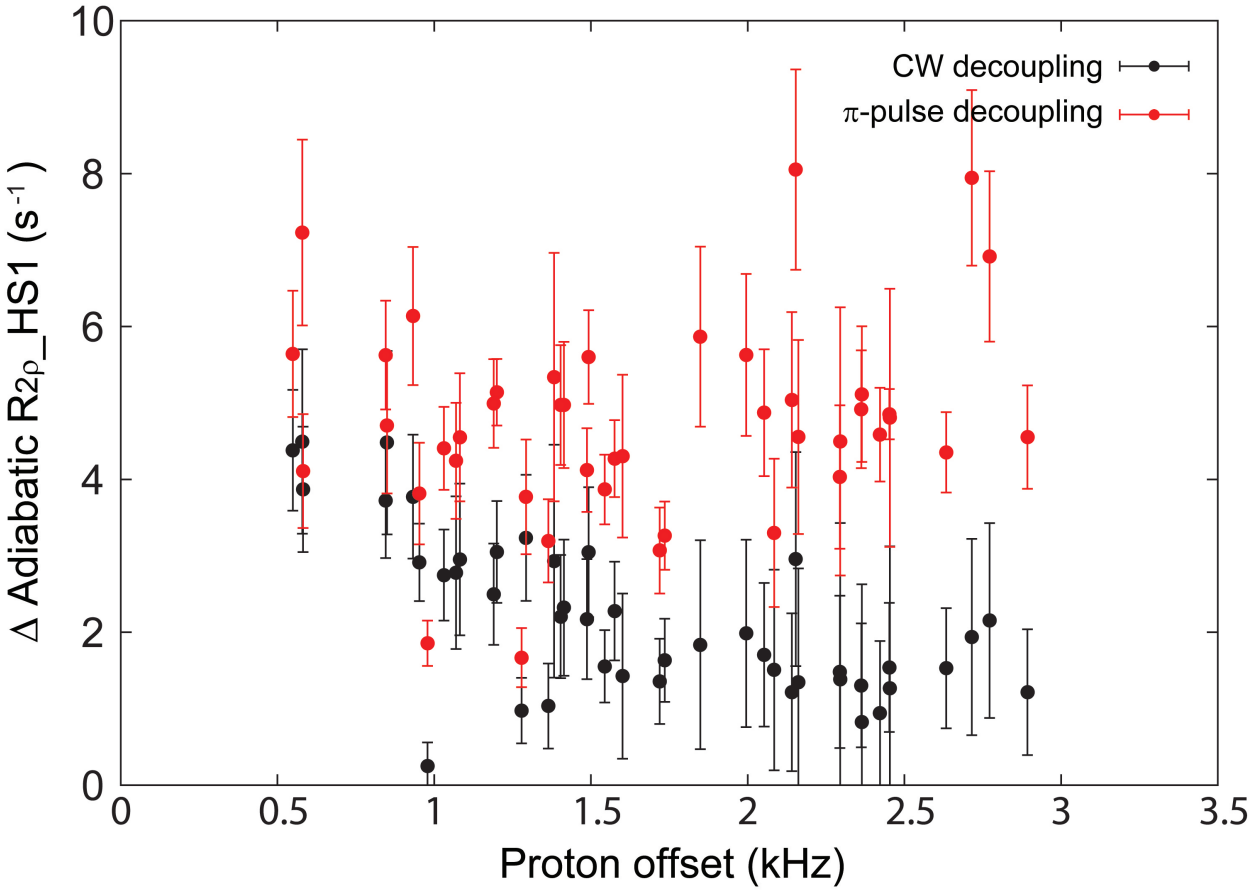

Figure S7. Experimental tests of off-resonance effects on different decoupling schemes. The ^{15}N, ^{2}H labeled Ube2g2 (15 ºC) was used for the tests, and the $R_{2\rho}$ relaxation rates were measured with the adiabatic pulse HS1 at 800MHz. The center of the decoupling RF field was placed at 6 ppm, the power level of CW proton decoupling was 1 kHz, and the power level for π-pulse decoupling was 22 kHz. The differences between the relaxation rates with a given proton decoupling scheme and the ones without proton decoupling scheme are plotted against the proton offset.

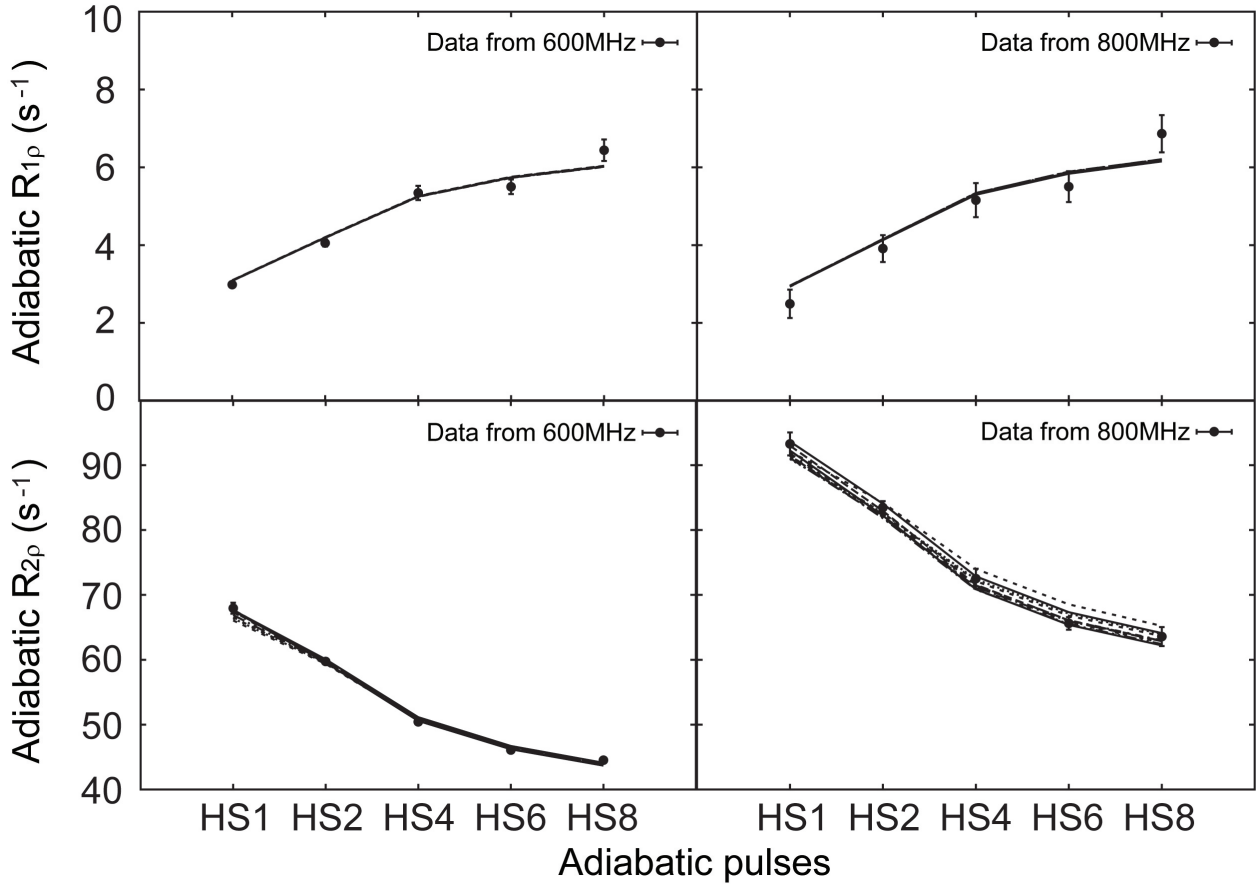

Figure S8. The curve fitting of relaxation data of residue 70 (k_{ex} = ~1.8 k s^{-1}) in Ub14 using the geometric approximation method. The 10 fitting curves with best scores are shown in the plots.

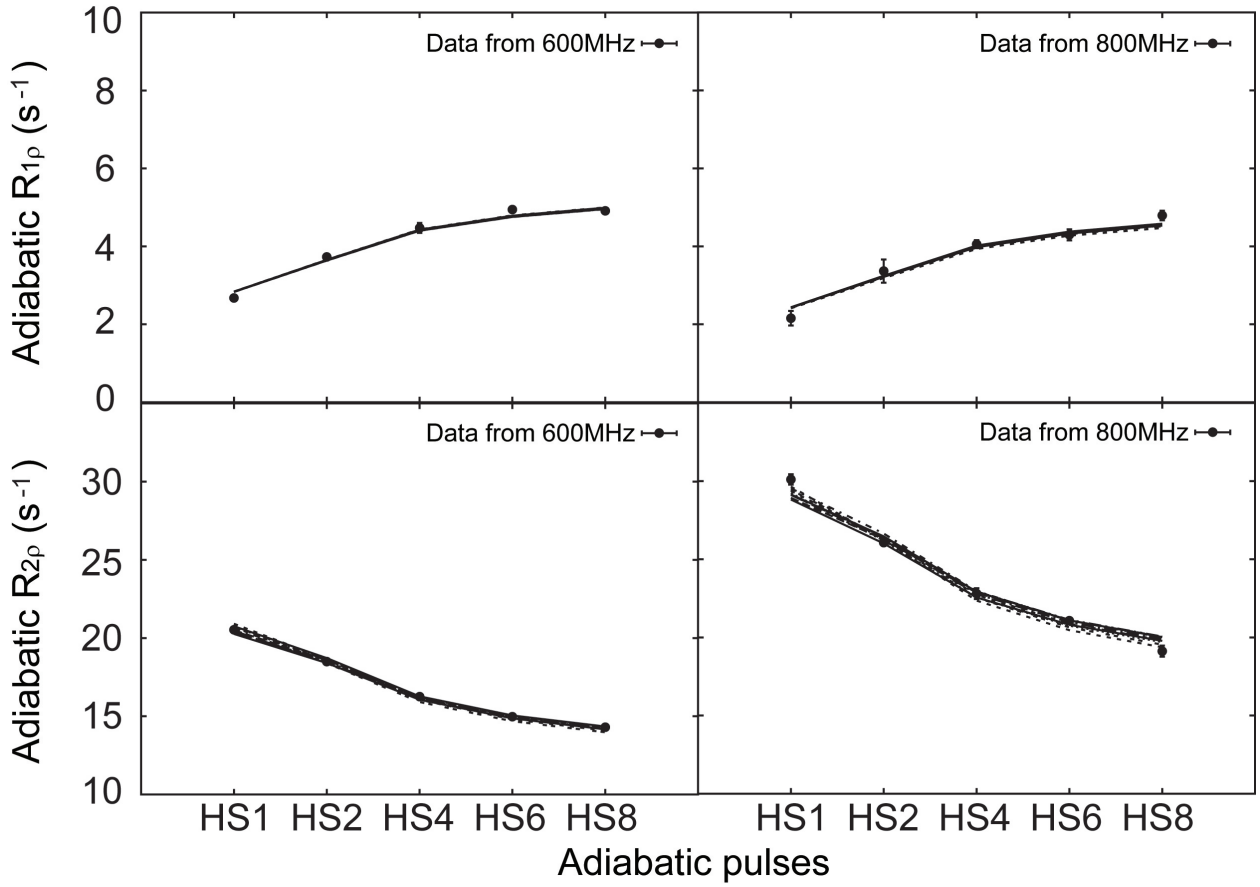

Figure S9. The curve fitting of relaxation data of residue 6 (k_{ex} = ~0.6 k s^{-1}) in Ub14 using the geometric approximation method. The 10 fitting curves with best scores are shown in the plots.

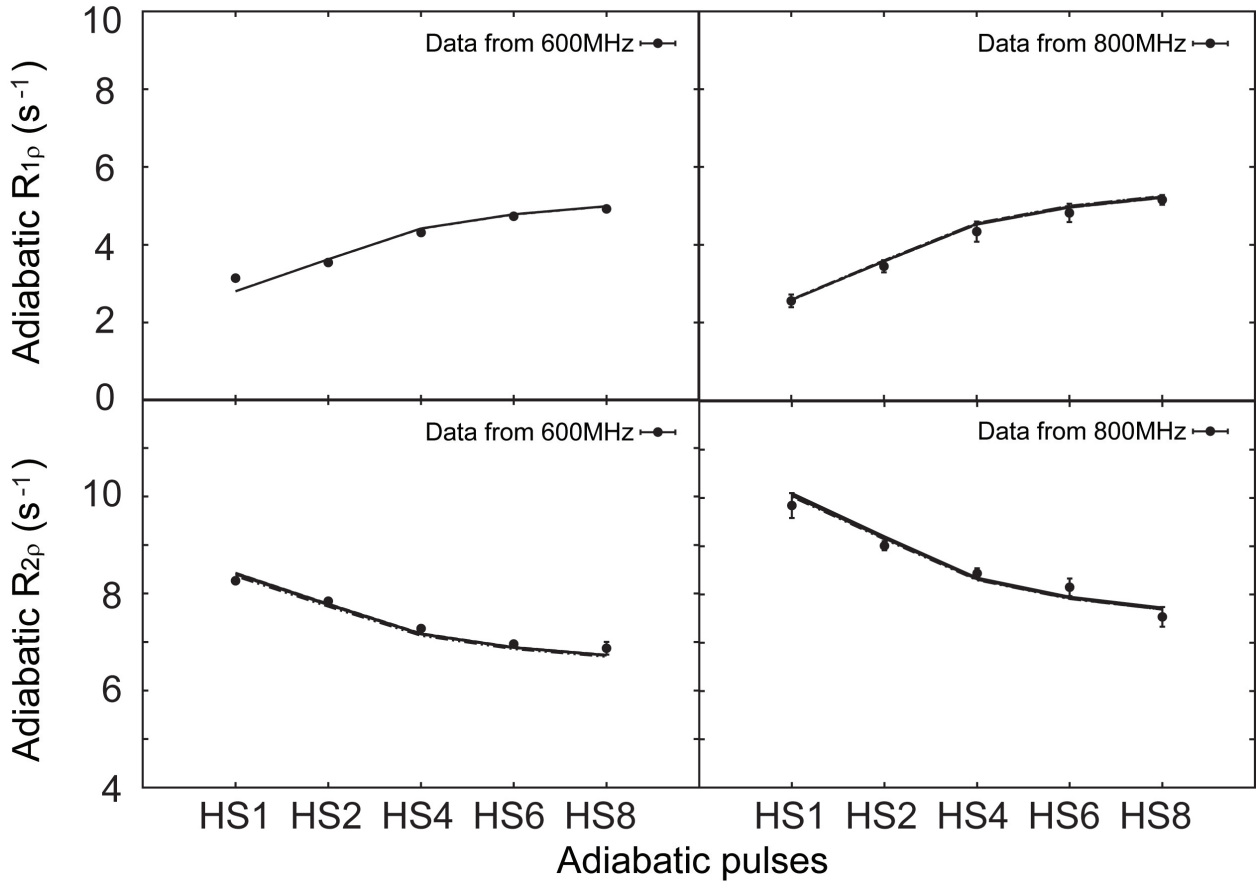

Figure S10. The curve fitting of relaxation data of residue 18 (k_{ex} = ~37 k s^{-1}) in Ub14 using the geometric approximation method. The 10 fitting curves with best scores are shown in the plots.

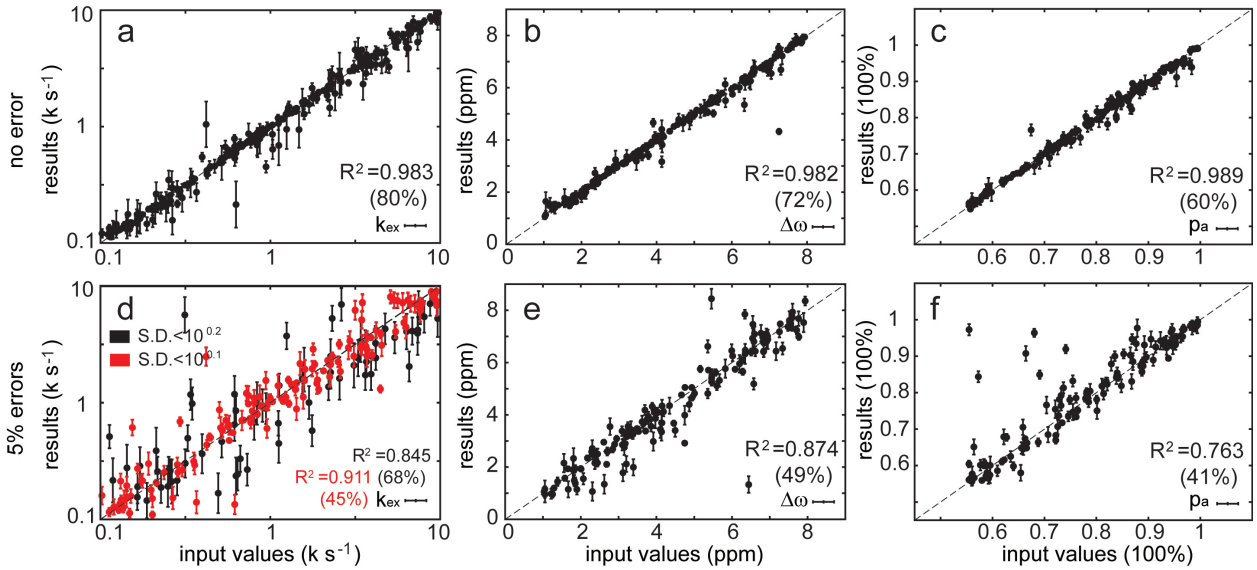

Figure S11. Data analysis of adiabatic $R_{1\rho}$, adiabatic $R_{2\rho}$, and R_1 at two (14.1T and 18.8T) magnetic fields in the slow exchange regime using geometric approximation. 300 relaxation data sets without (a, b, c) or with up to 5% errors (d, e, f) at two different magnetic fields were simulated using Bloch-McConnell equation with random dynamic parameters. The fit results are plotted against the input values, and those with large standard deviations (S.D. of $k_{ex} > 10^{0.2}$, S.D. of $\Delta\omega > 0.4$ppm, S.D. of $p_a > 2.5\%$) during Monte Carlo sampling are not shown. The red data points are those with smaller standard deviations (S.D. of $k_{ex} < 10^{0.1}$). The coefficient of determination (R^2) is calculated for each dynamic parameter. The numbers in the parentheses are the percentages of the data remained after filtering out the results with large standard deviations.

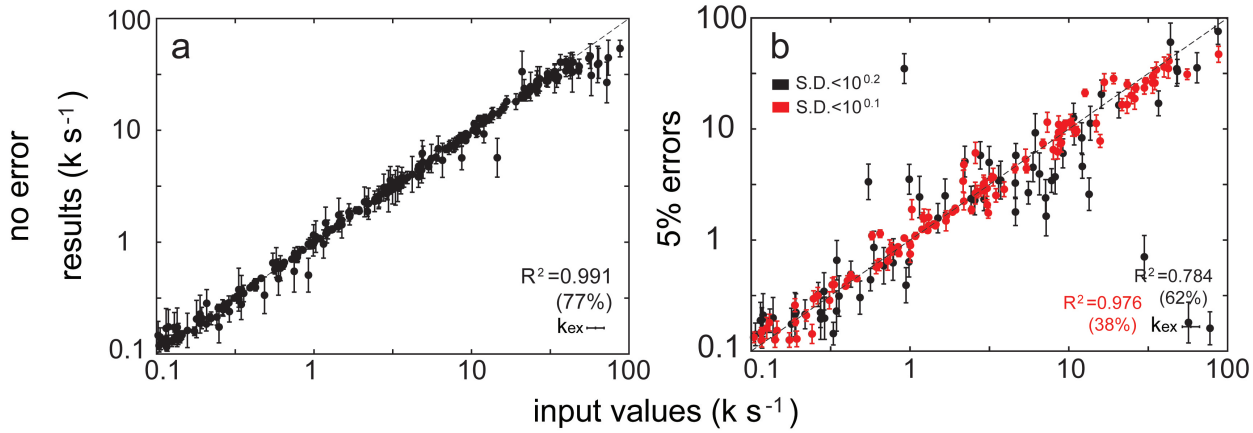

Figure S12. Data analysis of adiabatic $R_{1\rho}$, adiabatic $R_{2\rho}$, and R_1 at two (14.1T and 18.8T) magnetic fields in the mixed slow and fast exchange regimes using geometric approximation. 300 relaxation data sets without (a) or with up to 5% errors (b) at two different magnetic fields were simulated using Bloch-McConnell equation with random dynamic parameters. After a model is selected for each result, the fit results are plotted against the input values, and those with large standard deviations (S.D. of $k_{ex} > 10^{0.2}$) during Monte Carlo sampling are not shown. The red data points are those with smaller standard deviations (S.D. of $k_{ex} < 10^{0.1}$). The coefficient of determination (R^2) is calculated for each dynamic parameter. The numbers in the parentheses are the percentages of the data remained after filtering out the results with large standard deviations.

Residue number	Model	k_{ex} (k s^{-1})	S.D. (k s^{-1})	k_{ex} (10^x k s^{-1})	S.D. (10^x k s^{-1})
4	Slow	0.283489	0.237777	-0.54746	0.331211
6	Slow	0.6273	0.383077	-0.20252	0.251004
13	Slow	0.192178	0.104802	-0.7163	0.226438
15	Slow	0.548582	0.626487	-0.26076	0.424876
18	Fast	37.00794	5.373404	1.568295	0.062838
20	Fast	61.53278	43.68095	1.789107	0.286957
32	Slow	0.283921	0.153106	-0.5468	0.224115
34	Slow	0.278541	0.389108	-0.55511	0.493449
35	Slow	3.921751	1.845212	0.59348	0.197464
36	Fast	3.134942	1.11583	0.496229	0.151489
42	Fast	1.811874	0.845072	0.258128	0.195852
43	Slow	2.314314	1.530191	0.364422	0.269514
44	Slow	0.32999	0.164353	-0.4815	0.208232
49	Fast	38.36913	3.647764	1.583982	0.041227
50	Slow	0.160814	0.063731	-0.79368	0.167898
57	Slow	0.381549	0.511766	-0.41845	0.479189
59	Fast	55.18542	28.93867	1.741824	0.218415
61	Fast	>100	…	>2	…
65	Fast	99.32569	36.12866	1.997062	0.154679
67	Slow	0.263689	0.267118	-0.57891	0.386756
69	Slow	1.491089	0.890492	0.173504	0.245998
70	Fast	1.776879	0.305477	0.249658	0.0743
71	Fast	1.853067	0.737783	0.267891	0.16864
73	Fast	3.172482	1.240512	0.501399	0.165765
74	Fast	7.358664	2.230237	0.866799	0.129688
76	Fast	34.11158	5.27262	1.532902	0.066864

Table S1. The conformational exchange rates of residues in Ub14 characterized by adiabatic relaxation dispersion experiments and geometric approximation method. The k_{ex} values are also represented on the logarithmic scale on the right two columns.

Reference

(1) Carver, J. P.; Richards, R. E. *J Magn Reson* **1972**, *6*, 89.

(2) Luz, Z.; Meiboom, S. *J Chem Phys* **1963**, *39*, 366.

(3) Trott, O.; Palmer, A. G. *J Magn Reson* **2002**, *154*, 157.

(4) Trott, O.; Abergel, D.; Palmer, A. G. *Mol Phys* **2003**, *101*, 753.

(5) Traaseth, N. J.; Chao, F. A.; Masterson, L. R.; Mangia, S.; Garwood, M.; Michaeli, S.; Seelig, B.; Veglia, G. *J Magn Reson* **2012**, *219*, 75.

(6) Metropolis, N.; Ulam, S. *J Am Stat Assoc* **1949**, *44*, 335.

(7) Korzhnev, D. M.; Skrynnikov, N. R.; Millet, O.; Torchia, D. A.; Kay, L. E. *J Am Chem Soc* **2002**, *124*, 10743.

(8) Mcconnell, H. M. *J Chem Phys* **1958**, *28*, 430.

(9) Abragam, A. *Principles of Nuclear Magnetism*; Oxford Univ Press (Clarendon), Oxford, 1961.

(10) Redfield, A. G. *Ibm J Res Dev* **1957**, *1*, 19.

(11) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. A. *J Magn Reson* **1992**, *97*, 359.

(12) Massi, F.; Johnson, E.; Wang, C. Y.; Rance, M.; Palmer, A. G. *J Am Chem Soc* **2004**, *126*, 2247.

(13) Auer, R.; Tollinger, M.; Kuprov, I.; Konrat, R.; Kloiber, K. *J Biomol Nmr* **2011**, *51*, 35.