Biophysical Journal, Volume 99

Supporting Material

NMR Studies on Domain Diffusion and Alignment in Modular GB1 Repeats

Joseph Walsh, Katlyn Meier, Rieko Ishima, and Angela Gronenborn

NMR Studies on Domain Diffusion and Alignment in Modular GB1 repeats

Joseph D. Walsh, Katlyn Meier, Rieko Ishima, Angela M. Gronenborn

Protein expression and purification.

Gene constructs for all proteins were designed and inserted into a modified pET-15b vector (Novagen), containing an amino-terminal 6-his tag and a TEV protease cleavage site. All proteins contained the six residue GGSGGS sequence at their N-termini and an additional serine from the TEV protease recognition sequence after cleavage. Uniform ^{15}N labeled proteins were expressed in modified minimal medium using $^{15}NH_4Cl$ (Isotec, Inc.) as the sole nitrogen source and induction by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours a 37 ºC. Cells were harvested by centrifugation and the cell pellets were suspended in buffer A (50 mM sodium phosphate buffer, pH 7.4, 0.5 M NaCl, 20 mM imidazole), homogenized and disrupted by passing twice through a microfluidizer (Microfluidics). Cell debris was removed by centrifugation at 150,000g for an hour and the supernatant was applied to a HisTrap HP (GE Healthcare) nickel column, equilibrated in buffer A. The protein was eluted using 250 mM imidazole, dialyzed overnight against 50 mM sodium phosphate buffer, 5 mM β–mercaptoethanol, pH 7.4, and subsequently digested with 1/20 w/w TEV protease at 4 ºC for one day. Passage over the HisTrap nickel column in buffer A was used to remove the tag and the 6-histidine-tagged tobacco etch virus (TEV) protease. The GB-protein containing flow-through was concentrated (Amicon) and further purified by gel filtration over a Superdex S75 26/60 column (Amersham Biosciences, GE Healthcare) in 25 mM sodium phosphate buffer pH 6.5 0.02% NaN_3. Uniform ^{15}N, ^{13}C labeled proteins of sGB1, sGB1L7I, and sGB3β1 were expressed in $^{15}NH_4Cl$ and ^{13}C-glucose containing minimal medium and purified using the same procedure. The GB1 protein (domain without any tails) was purified as described previously [1].

Selection of resonances for diffusion tensors analysis.

To determine molecular diffusion tensor from the relaxation data, the set of resonances used for analysis were selected based on the amino acid locations in the more rigid parts of the structure as determined by $^{15}N-\{^1H\}$ NOE values > 0.65, as well as the absence of conformational exchange. The NOE values were corrected for incomplete ^1H T_1 recovery.[2] Experiments to determine $^{15}N-\{^1H\}$ NOE values were performed at 600 MHz on the proteins GB1, sGB1, dGB1L7I-(6)-GB1, dGB3β_1-(6)-GB1, and dGB1L7I-(3)-GB1. Among the resonances from the double-GB1 proteins with corrected heteronuclear $^{15}N-\{^1H\}$ NOE values > 0.65, only residue 31 was identified as potentially undergoing exchange, however further relaxation analysis[3,4] did not substantiate the presence of an exchange contribution. Given the error ranges in the data and the location of residue 31 on the helix, we included the data for residue 31 in the diffusion analysis. As a consequence of these selection criteria, there were 10 resonances available for diffusion

tensor fitting for the dGB1L7I-(n)-GB1 proteins, 22 resonances available for dGB3β_1-(6)-GB1, and 44 resonances for the single-GB1 proteins.

Selection of diffusion tensor model.

Analysis assuming isotropic, axial and anisotropic molecular diffusion was carried out to derive the most appropriate model (Table S1). As is evident from the χ^2 values, neither the isotropic nor oblate models describe the rigid-body diffusion of the individual domains well enough, while both the prolate and fully anisotropic models can be used for describing the relaxation data. However, selecting the fully anisotropic model over the prolate axial model is not warranted, since the derived reduction in χ^2 is not statistically significant. Therefore, we chose the prolate model for further analysis of the domain diffusion for all our protein constructs.

TABLE S1. Diffusion tensor parameters of the single GB1domain and individual domains in the double domain protein with a 24 residue linker, for different models obtained by fitting ^{15}N relaxation data.

N [a]	x [b]	model	α [c] (°)	β [c] (°)	γ [c] (°)	D_x [d]	D_y [d]	D_z [d]	τ [e] (ns)	χ^2 [f]	F_x [g]	$P(F_x)$ [h]
sGB1												
44	1	Iso.							3.83 (0.01)	26.37		
44	4	Ob.	179 (2)	-89 (3)		4.76 (0.02)	4.76 (0.02)	3.14 (0.04)	3.95 (0.03)	15.46	11.11	<1e-6
44	4	Pro.	84 (2)	58 (4)		3.52 (0.02)	3.52 (0.02)	6.09 (0.06)	3.81 (0.04)	2.27	153.13	<1e-6
44	6	Anis.	84 (2)	59 (4)	-152 (28)	3.42 (0.04)	3.62 (0.04)	6.15 (0.07)	3.79 (0.05)	2.26	1.12	0.36
10	1	Iso.							4.04 (0.02)	43.19		
10	4	Ob.	33 (11)	-44 (6)		4.73 (0.05)	4.73 (0.05)	2.74 (0.06)	4.10 (0.08)	4.87	24.61	4.2e-5
10	4	Pro.	86 (4)	63 (7)		3.50 (0.03)	3.50 (0.03)	6.09 (0.13)	3.82 (0.08)	2.77	44.78	3.0e-6
10	6	Anis.	85 (37)	59 (36)	-136 (52)	2.75 (0.62)	4.14 (0.50)	7.39 (1.38)	3.50 (0.78)	3.52	0.36	0.83
dGB1L7I-(24)-GB1												
NTD												
10	1	Iso.							5.34 (0.02)	60.99		
10	4	Ob.	26 (9)	-42 (4)		3.50 (0.03)	3.50 (0.03)	2.00 (0.04)	5.56 (0.09)	11.00	14.63	3.5e-3
10	4	Pro.	77 (3)	61 (4)		2.55 (0.02)	2.55 (0.02)	4.54 (0.08)	5.19 (0.10)	6.65	25.51	3.7e-5
10	6	Anis.	80 (27)	57 (17)	-134 (17)	1.37 (0.28)	3.56 (0.37)	6.57 (0.61)	4.35 (0.58)	4.66	2.28	0.18
CTD												
10	1	Iso.							5.43 (0.02)	125.90		
10	4	Ob.	35 (6)	-44 (3)		3.77 (0.04)	3.77 (0.04)	1.51 (0.04)	5.53 (0.11)	7.58	47.83	3e-6
10	4	Pro.	87 (2)	64 (4)		2.38 (0.02)	2.38 (0.02)	5.18 (0.09)	5.02 (0.10)	2.38	156.70	< 1e-6
10	6	Anis.	86 (3)	60 (4)	-153 (39)	2.10 (0.15)	2.66 (0.12)	5.57 (0.41)	4.84 (0.42)	2.17	1.29	0.37

Errors (2 rmsd) in diffusion parameters are listed in parenthesis.

[a] Number of T_1/T_2 values used to fit the diffusion model.

[b] number of parameters in diffusion model.

[c] Euler angles transformed to Z-Y-Z convention from the Tensor2 output.

[d] Principal values of the diffusion tensor (10^{-7} s^{-1}).

[e] Rotational correlation time. $\tau = 1/(2\ Tr(D))$.

[f] Reduced Chi-squared, $\chi^2_{red} = \chi^2/(N-x)$.

[g] F-statistic for the axial model refers to the improvement in χ^2 over the isotropic model, F-statistic of anisotropic model refers to the improvement over the prolate axial model.

[h] Probability that the observed improvement in χ^2 could be obtained by chance. $P(F_{x-x'}) = P(F_{x-x'},x-x',N-x))$, where x refers to the number of parameters in the current model, and x' the number of parameters in the reference model(see g), $x'<x$.

Assessment of the validity of the rigid body diffusion model.

The prolate diffusion tensor fits for both, sGB1 and sGB1L7I, yielded reduced χ^2 values of 2.3. This suggests that while maintaining the same rigid body model of domain diffusion, the presence of the tails contributes an additional error, comparable to the estimated random error of the T_1 and T_2 relaxation measurements.

The prolate diffusion tensor fits for the double-GB1 protein domains exhibited reduced χ^2 similar to those of the tail-containing GB1 proteins (omitting the data of V39 in the NTD). For the CTD of the dGB1L7I-(n)-GB1 proteins, reduced χ^2 values < 2.4 were observed for all linker lengths, except the shortest, three-residue linker length protein. For the latter, a reduced $\chi^2 = 4.7$ was obtained. In contrast, for the NTD, the goodness of fit for all linker lengths was < 3.6, with no increase in the three-residue linker length protein (omitting the data from V39). Finally, for the dGB3β_1-(6)-GB1 protein, the NTD and CTD s fits had a reduced χ^2 of 3.79 and 3.97, respectively. Analysis of all the individual fits suggests that most of the error is contributed by the data from one residue, W43.

The influence of the data set size and angular sampling.

Since the number of resonances available for relaxation analysis of the dGB1L7I-(n)-GB1 proteins is small (10 T_1/T_2 values per domain), we assessed the effect of using a reduced data set on the fits of diffusion tensor parameters for the single-GB1 proteins. Equivalent prolate diffusion tensor fits were carried out with the full data set (44 T_1/T_2 values) and the minimal data sets (10 T_1/T_2 values). Comparisons of the results are provided in Fig. S1. As can be appreciated, for both data sets, τ, $D_{||}/D_{\perp}$, and Φ values all lie within a 2 rmsd. For the angle Θ, a difference as large as 10º can occur, although this difference, again, lies within a 2 rmsd. The excellent agreement between the tensor parameters derived using data from 10 resonances compared to the more complete set of 44 resonances is a gratifying result

and most likely is related to the good polar angle sampling for these particular 10 residues (Fig. S2). Furthermore, the fits are relatively robust; omission of individual data points generally does not substantially change the extracted diffusion parameters. The one exception is for the fit obtained when omitting the T_1/T_2 of V39, which results in an increase in the Φ−error and a not-significant increase in the Φ−value by ca. 10°. However, no effect upon omission of the V39 data is seen when the full data set available for sGB1, is used, indicating that the small increase in the fit Φ−value upon omission of the V39 data may be an artifact of sampling, only playing a role when fitting the tensor with the reduced data set (Fig. S3). The angular distribution of the data may be quantitatively assessed by use of the sampling tensor formalism [5]. This approach allows the determination of the directions in space sampled best, and a *generalized sampling parameter*, Ξ, is used to describe the overall angular sampling on a scale from 0 to 1, corresponding to uniform and mono-directional, respectively. For the 44 data points of single-GB1 Ξ = 0.14, whereas the sampling of the 10 data point subset is Ξ = 0.22 (increasing to 0.29 if V39 is omitted). The 10 point subset angular sampling therefore is remarkably good considering its small size; for comparison, the set of GB1 β−strand or α-helix residues yield Ξ = 0.67 and 0.83, respectively.

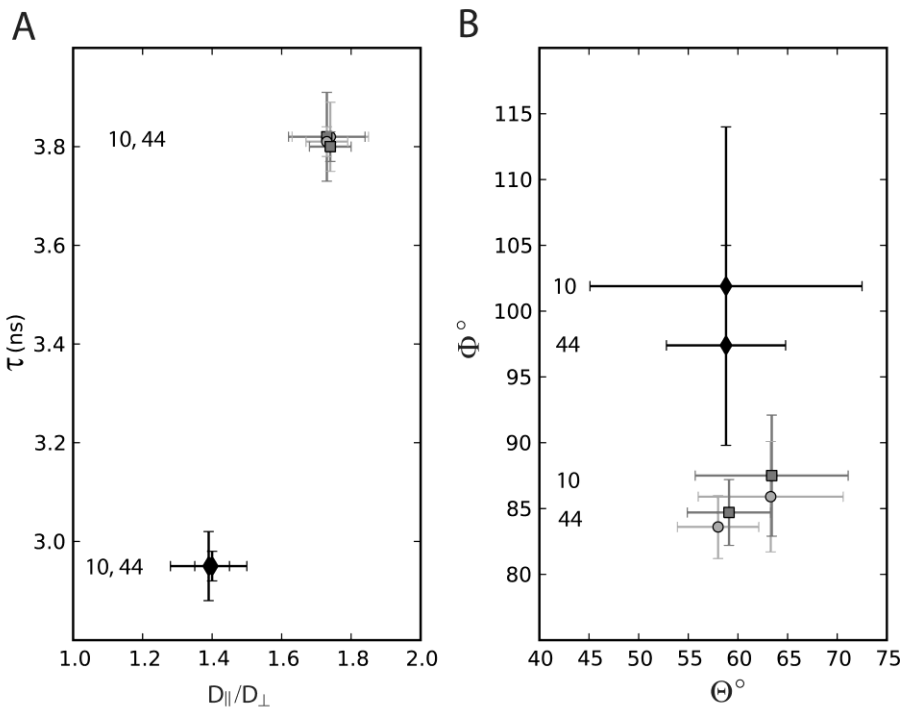

Figure S1. The effect of adding flexible N- and C-terminal tails to the domain on the diffusion tensors. The parameters obtained for tail-less GB1 (black diamonds) are compared with those of the tailed sGB1L7I (squares) and sGB1 (circles). Fitting was carried out using either 44 T_1/T_2 data points or a

subset of 10 data points, equivalent to those available for the dGB1L7I-(n)-GB1 proteins. **A.** Rotational correlation times, τ, plotted versus the anisotropy factor, $D_{\parallel}/D_{\perp}$ **B.** Polar (Θ) and azimuthal (Φ) angles describing the orientations of the principal axes of diffusion. All error bars indicate a range of 2 rmsd.

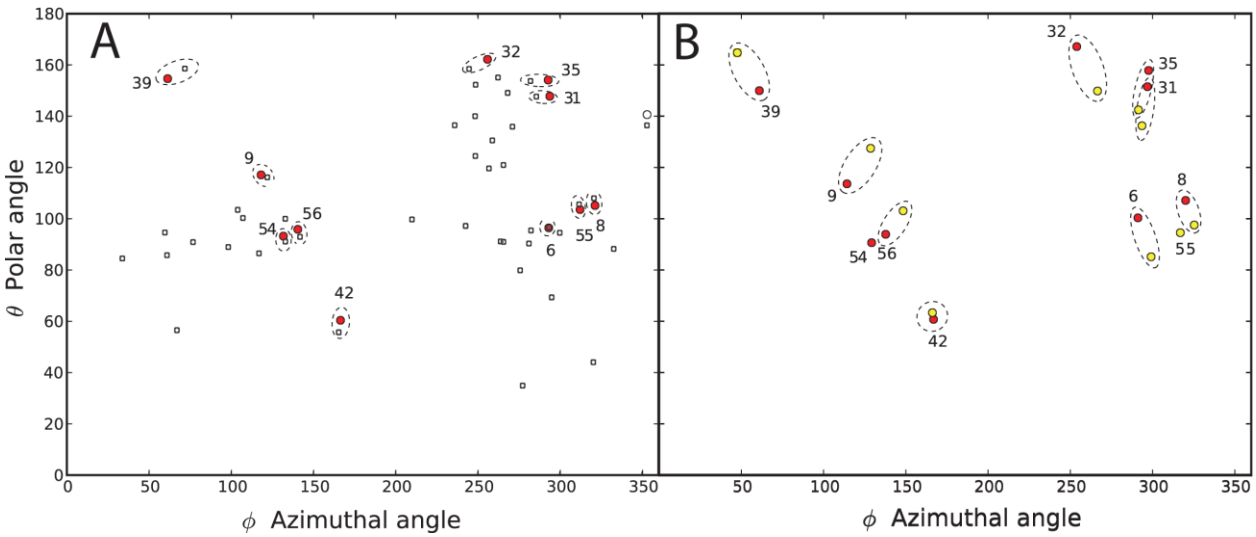

Figure S2. Distribution of the HN vectors in the diffusion frame orientation of the axial diffusion tensor. **A.** The sGB1 HN vector distribution using resonances associated with the rigid regions in the structure. (open circles). For comparison, the HN vector distribution obtained with the 10 resonance subset that is equivalent to the set used in the double-domain fits (filled red circles). **B.** Distribution of orientations of the HN vectors for the NTD (yellow) and CTD (red) diffusion tensor of dGB1L7I-(3)-GB1.

Robustness of the diffusion tensor fit for single GB1.

In order to further test the robustness of the diffusion tensor fits, we carried out relaxation data analysis, systematically omitting one of the T^j_1/T^j_2 values for each residue *j*. This so-called jack-knife procedure was applied to the data for GB1, sGB1 and sGB1L7I. We first fit the single-domain diffusion tensors using the same subset of 10 T_1/T_2 data values that were available for the dGB1L7I-(n)-GB1 systems. No significant differences (within the error of the fit) for the diffusion tensor parameters fits were noted upon omitting any one data point. However, omission of the T_1/T_2 data of residue V39 leads to a significant increase in the uncertainty in the Φ–value. In fact the uncertainty was twice the uncertainty obtained by omitting any of the other data points. In addition, for the tail-containing proteins sGB1 and sGB1L7I, omission of the V39 T_1/T_2 data resulted in an increase in the Φ–value by ca. 10° (although , this did not translate to a significant difference in the fit, due to the increased error). This increased Φ–value and concomitant increased error was not observed when using the full data set of 44 T_1/T_2 values. This indicates that the V39 T_1/T_2 data is only an important contributor to being able to determine a well defined minimum for the diffusion tensor parameters for cases where the angular sampling by the amide groups is limited. Furthermore V39 T_1/T_2 does not constitute an outlier, as demonstrated by its small contribution to the error and insignificant change in the reduced χ^2 when it is omitted from the fit (Table 1, Fig. S3). Indeed, χ^2 is very similar for all fits when a single data point is omitted. The largest effect is seen for residue N8 in sGB1L7I and residue E56 for sGB1. These, when omitted, lower the reduced χ^2 by ~ 2. The contribution of these two residues to the error in the diffusion tensor fit may be related to slight dynamic and structural differences of the backbone at residue N8 due to the neighboring mutation site L7I. Since the carbonyl oxygen of N8 is within hydrogen bonding distance to the E56 amide, it is possible that any structural variations near the mutation site effects the backbone amide orientation of E56 and the degree of agreement of E56 T_1/T_2 value with the diffusion tensor fit.

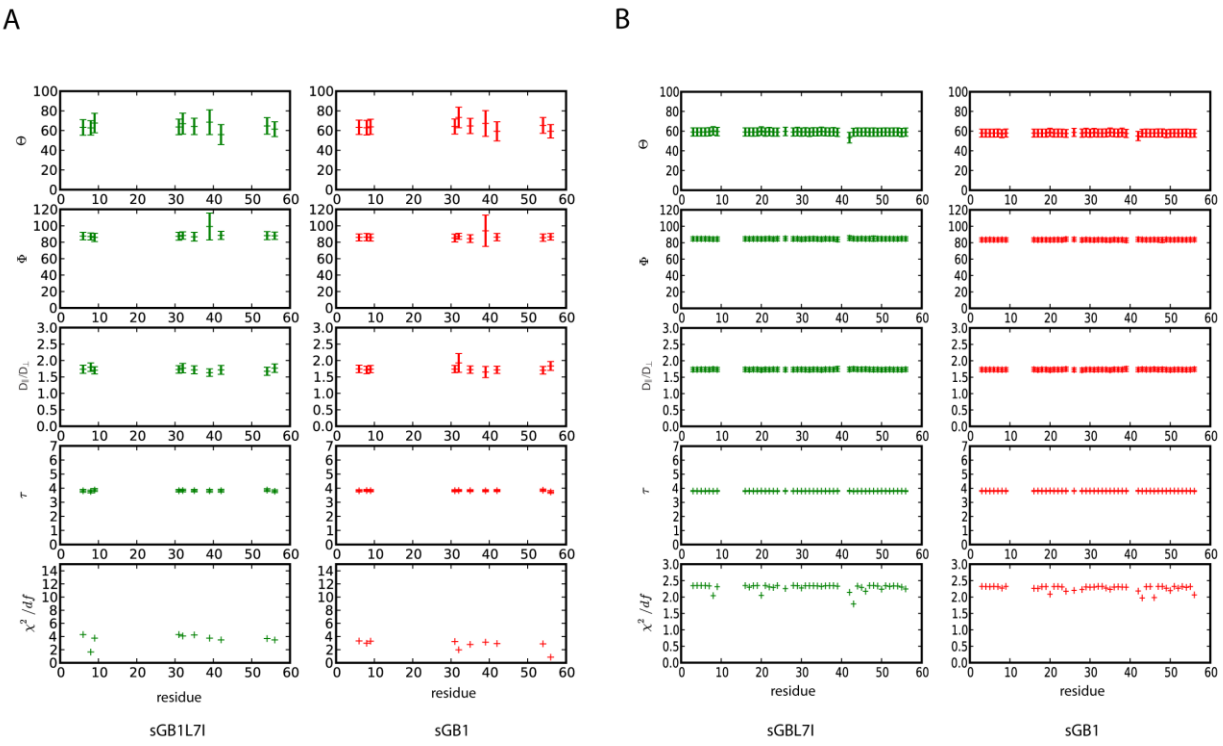

Figure S3. Prolate diffusion tensor fits for sGB1L7I and sGB1 using data from all rigid residues or a subset of 10 data points available for dGB1L7I-(n)-GB1 excluding one data point in turn. The fit obtained in this manner is then plotted at the position of the residue excluded from the fit. **A.** 10 point data set. **B.** rigid residues data set. All error bars indicate a range of 2 rmsd

Robustness of the diffusion tensor fit for the double dGB1L7I-(n)-GB1 domains.

No significant changes in the diffusion tensor parameters of the CTD of the dGB1L7I-(n)-GB1 proteins were observed upon omission of any one data point. The one change of note occurred upon omission of the data of V39. This caused a change in the fit Φ–value by ca. 10°, however due to the accompanying large increase in the error in Φ, this change was not deemed significant. These results for the CTD parallel those obtained for the tailed single-GB1 proteins when using the same small subset of ten data points.

In contrast, the tensor fit for the NTD was highly dependent on the presence or absence of data for two residues: for all linker lengths, omission of the data for residues 39 or 42 caused a significant drop in the reduced χ^2 (Table 1, Fig. S4). Omission of the data for only one of these residues led to reduced χ^2 values similar the effect seen for the CTD fits. Leaving out the data of residue 42 reduced Θ by ca. 15° for all linker lengths, without any other associated changes. Leaving the data for residue 39 out caused a larger effect. Values of Θ and Φ increased by ca. 10° and ca. 25°, respectively, associated with a large increase in the uncertainty in both angles. Thus, the difference in Φ became larger than 2 rmsd, while Θ increased, but remained within 2 rmsd. Furthermore, leaving out the data for residue 39 decreased the anisotropy of the NTD by ca. 0.4 for the protein with a 3 residue linker, with progressively smaller changes for proteins with longer linker lengths. For the double domain protein with a 24 residue linker, the decrease in anisotropy was ca. 0.2.

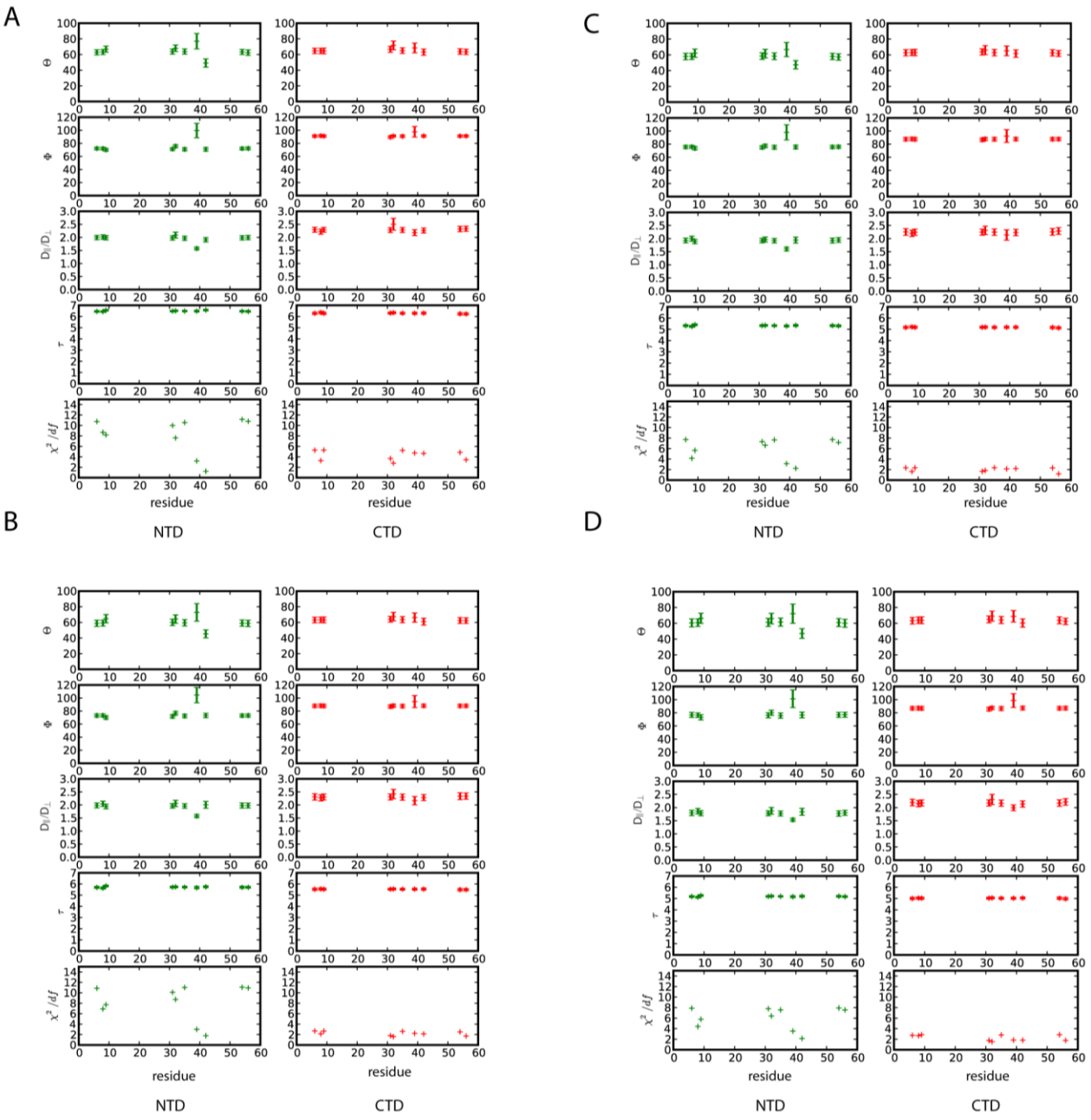

Figure S4. Prolate diffusion tensor fit for dGB1L7I-(n)-GB1 (n=3, 6, 12, 24) using the10 data point set, omitting one data point in turn. **A.** n=3. **B.** n=6. **C.** n=12. **D.** n=24. All error bars indicate a range of 2 rmsd

Robustness of the diffusion tensor fit for the double dGB3β_1-(6)-GB1 domains.

The robustness of the fits was tested by leaving out the data for one residue at a time from either the full set of 22 T_1/T_2 values, or the subset of 10 T_1/T_2 values available for the dGB1L7I-(n)-GB1 constructs. Omission of the T_1/T_2 value for residue V39, using the subset of 10 data points, resulted in an increase in the NTD Θ and Φ values, and a significant decrease in the anisotropy with a large increase in the uncertainty, similar to our observation for the dGB1L7I-(n)-GB1 proteins. When all 22 available data points with NOE values > 0.65 were used, a similar effect was noted, although the uncertainties in Θ and Φ were notably less than in the 10 point subset (Fig. S5). Again, only omission of the data for residue V39 resulted in a changed fit, although leaving out the data for either 39 or 42 reduced χ^2 by ~5 for the 10 data point set, and ~2.5 for the 22 data point set. For both double domain proteins, omission of the data for either residue 39 or 42 brought the reduced χ^2 of the NTD into the range obtained for the CTD.

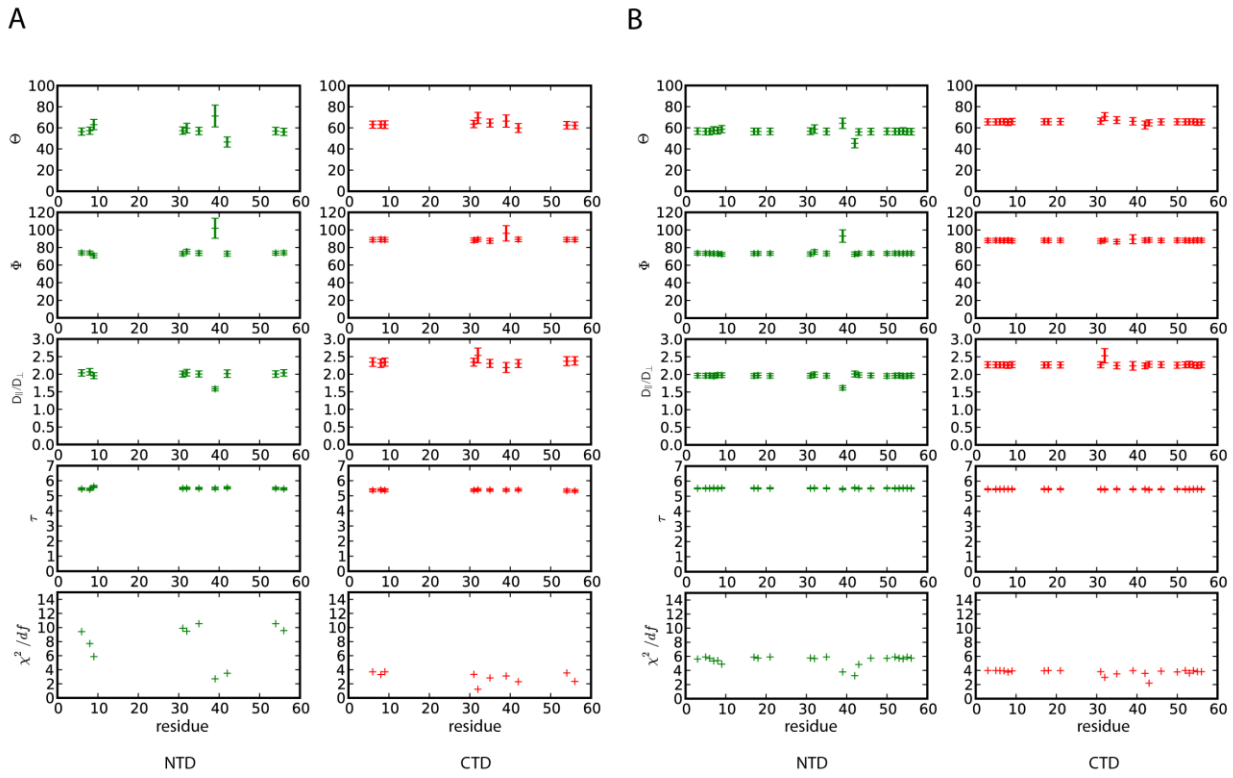

Figure S5. Prolate diffusion tensor fit for dGB3β_1-(6)-GB1 using either the 10 or 22 data point set, omitting one data point in turn. **A.** Fits based on the subset of 10 data points available for dGB1L7I-(n)-

GB1. **B.** Fits based on the full dataset of 22 data points available for dGB3β_1-(6)-GB1. All error bars indicate a range of 2 rmsd.

Relaxation Data.

The T_1, T_2 relaxation rate data belonging to rigid portions of the domains of dGB1L7I-(n)-GB1 (n=3, 6, 12, 24) and dGB3β_1-(6)-GB1 are shown in Fig. S6. Additionally, the ratio, T_1/T_2 used to fit the domain diffusion tensors is shown in Fig. S7. The $^{15}N-\{^1H\}$ NOE values for all resolved residues of dGB1L7I-(3)-GB1 and dGB1L7I-(6)-GB1 are shown in Fig. S8.

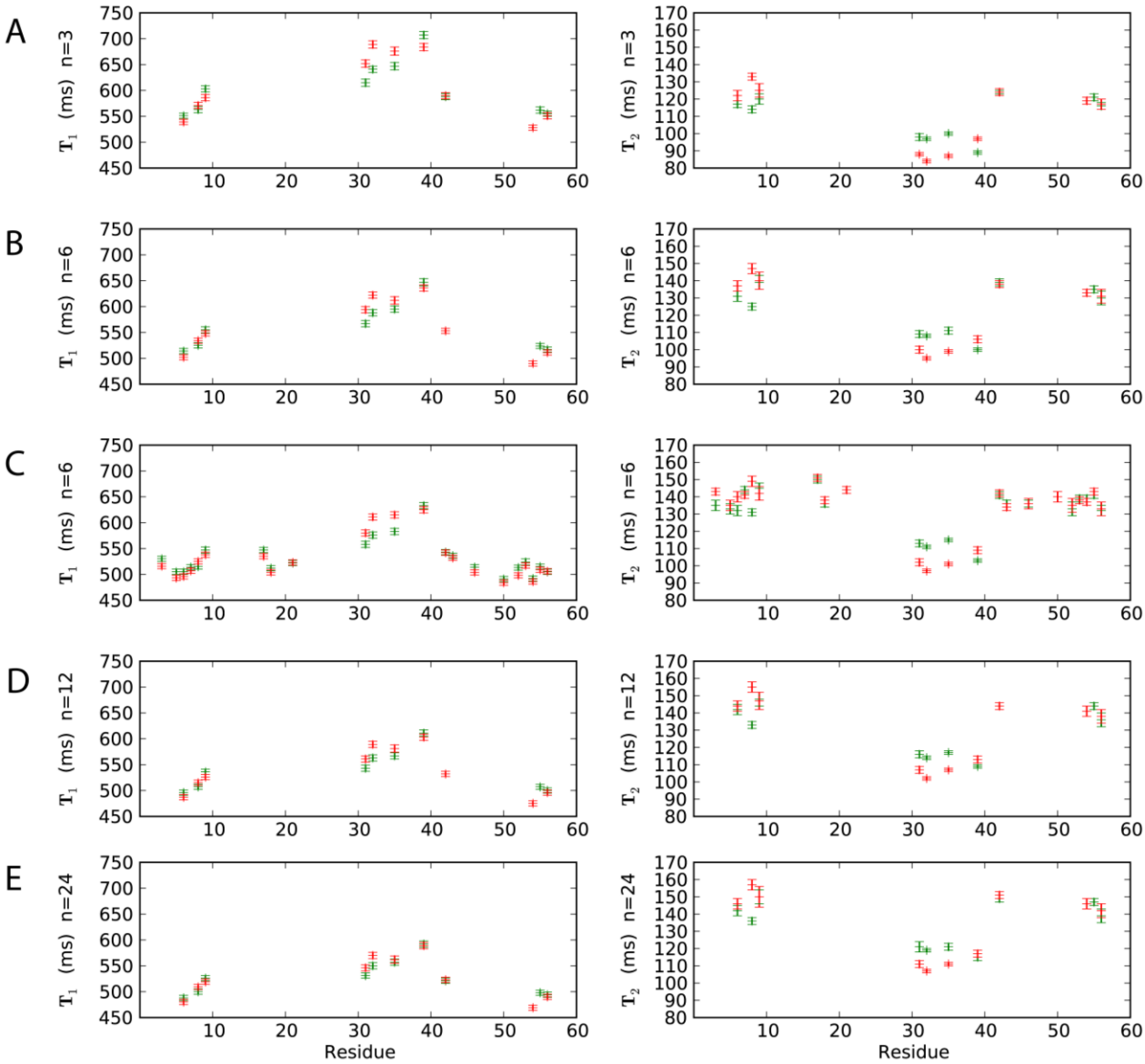

Figure S6. T_1 and T_2 relaxation rates for double-GB1 proteins of various inter-domain linker lengths, measured at 600 MHz. **A.** dGB1L7I-(3)-GB1 **B.** dGB1L7I-(6)-GB1 **C.** dGB3β_1-(6)-GB1 **D.** dGB1L7I-(12)-GB1 **E.** dGB1L7I-(24)-GB1.

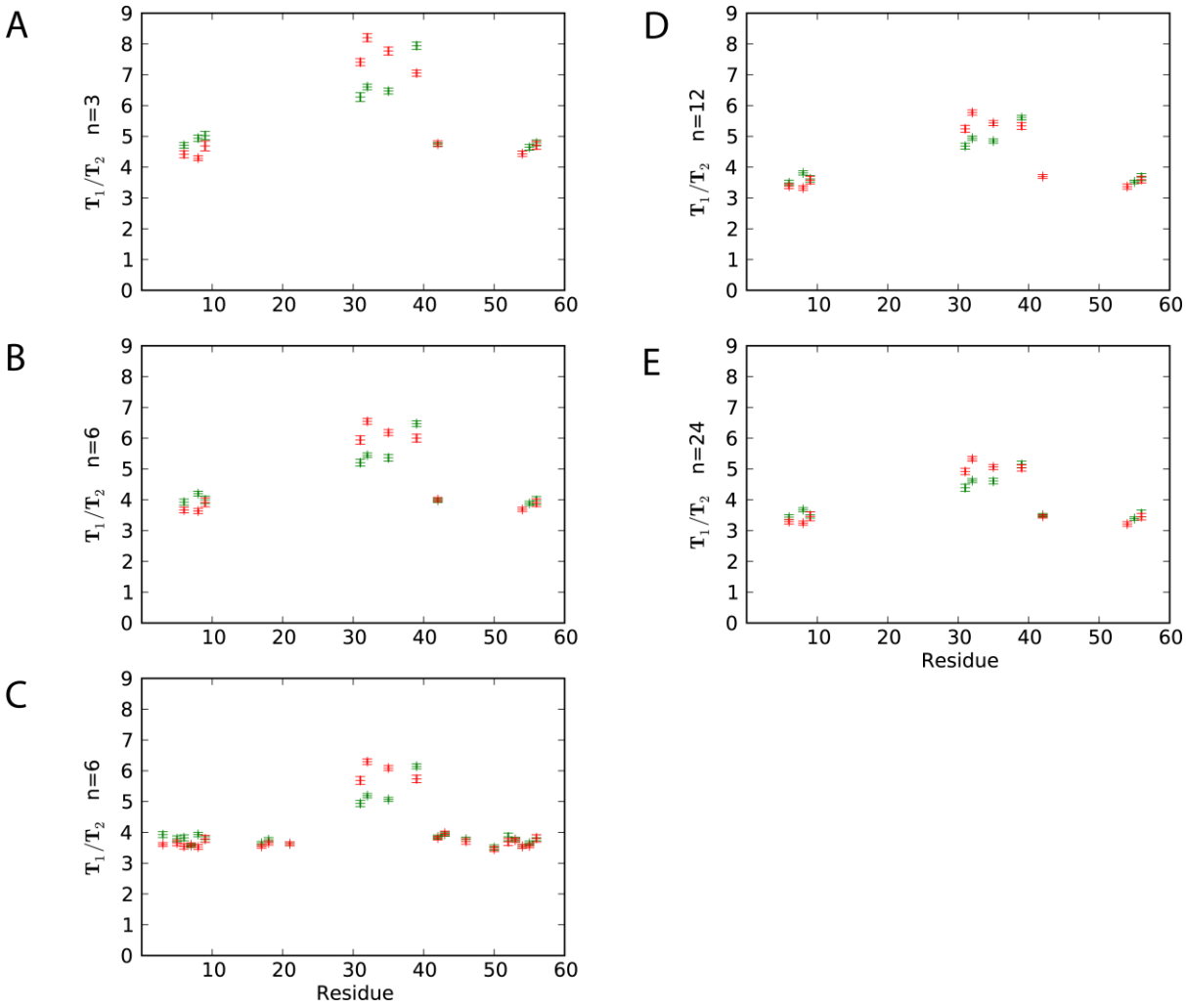

Figure S7. Ratio T_1/T_2 relaxation rates from rigid portions of double-GB1 proteins of various inter-domain linker lengths (600 MHz). **A.** dGB1L7I-(3)-GB1 **B.** dGB1L7I-(6)-GB1 **C.** dGB3β_1-(6)-GB1 **D.** dGB1L7I-(12)-GB1 **E.** dGB1L7I-(24)-GB1.

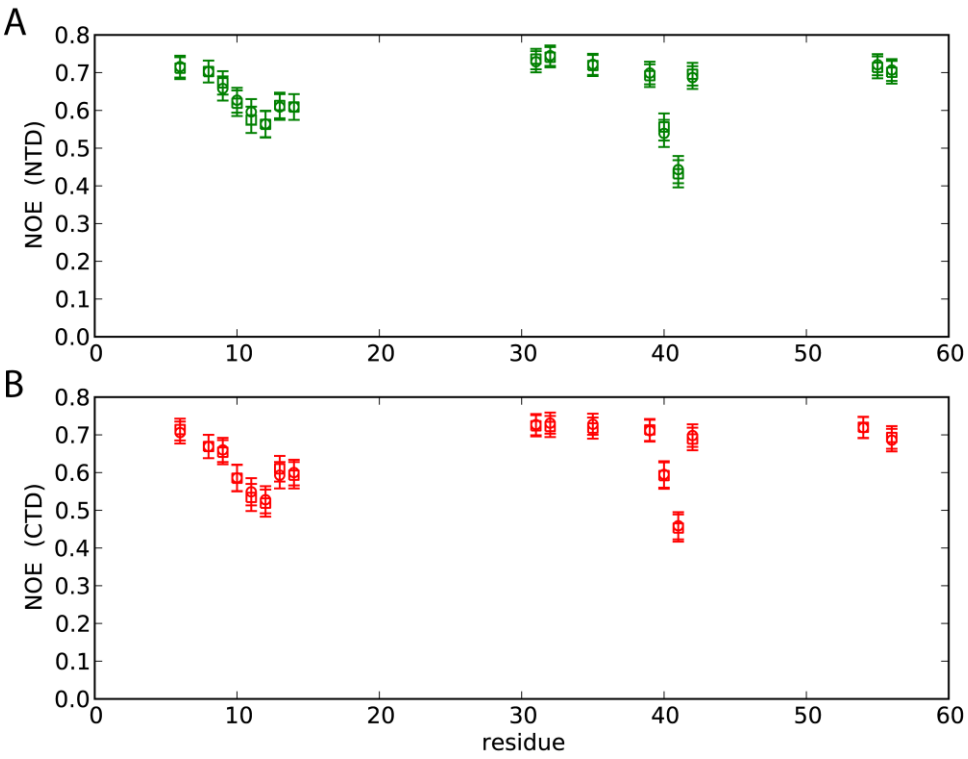

Figure S8. A. The $^{15}N-\{^{1}H\}$ NOE values for the NTD of dGB1L7I-(3)-GB1 (squares) and dGB1L7I-(6)-GB1 (circles). **B.** The $^{15}N-\{^{1}H\}$ NOE values for the CTD of dGB1L7I-(3)-GB1 (squares) and dGB1L7I-(6)-GB1 (circles).

Alignment tensor fits

The alignment tensor parameters obtained for GB1, sGB1 and the NTD and CTD of dGB1L7I-(3)-GB1 and dGB1L7I-(18)-GB1 are shown in Table S2. The values for Da and R are quoted to three and two significant figures, respectively, and the angles are quoted to the nearest degree. The errors in the fit parameters were estimated by a Monte Carlo simulation of 500 runs using the normally distributed error superimposed on the experimental data. The standard deviation used in the Monte Carlo simulation was the mean difference between the experimental and back-calculated RDCs. We emphasize that these are the estimated errors for the alignment tensor parameters when using a particular sized data set = N. The error resulting from the use different sized data sets may be judged by inspection of the parameters obtained for GB1 and sGB1, where larger data sets were available for comparison.

TABLE S2. Alignment tensor parameters extracted from RDC data measured in C12E5 alignment media.

n[a]	N[b]	Da[c] (Hz)		R[c]		α[d]		β[d]		γ[d]		rms[e] (Hz)	χ^{2}[f]
						GB1							
-	30	3.32	(0.10)	0.37	(0.05)	145	(3)	74	(2)	87	(1)	0.25	0.29
-	7	3.62	(0.15)	0.40	(0.02)	157	(4)	78	(2)	89	(1)	0.09	0.1
						sGB1							
-	38	6.91	(0.14)	0.17	(0.02)	123	(3)	65	(1)	101	(1)	0.43	0.21
-	8	7.15	(0.09)	0.17	(0.01)	148	(8)	67	(1)	99	(1)	0.06	0.03
						dGB1L7I-(n)-GB1							
						NTD							
3	8	10.79	(0.35)	0.26	(0.07)	162	(5)	65	(2)	110	(1)	0.4	1.66
18	8	7.74	(0.19)	0.20	(0.05)	164	(5)	63	(1)	105	(1)	0.24	0.61
						CTD							
3	8	12.96	(0.10)	0.19	(0.02)	172	(1)	70	(1)	96	(1)	0.09	0.08
18	8	9.72	(0.16)	0.14	(0.04)	175	(3)	66	(1)	98	(1)	0.15	0.25

Errors (2 rmsd) in alignment parameters are listed in parenthesis.

[a] Number of residues in the inter-domain linker.

[b] Number of RDC values used in the SVD fit of the alignment tensor, using PALES[6].

[c] Magnitude (Da) and rhombicity (R) of the alignment tensor.

[d] Euler angles using the right-handed Z-Y-Z convention.

[e] Root-mean-square deviation between experimental RDCs and RDC back-calculated with the fit alignment tensor.

[f] Reduced Chi-squared, $\chi^2_{red} = \chi^2/(N-5)$.

References:

(1) Jee, J.; Byeon, I. J. L.; Louis, J. M.; Gronenborn, A. M. *Proteins-Structure Function and Bioinf.* **2008**, *71*, 1420.
(2) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 12593.
(3) Loria, J. P.; Rance, M.; Palmer, A. G. *J. Am.Chem.Soc.* **1999**, *121*, 2331.
(4) Kneller, J. M.; Lu, M.; Bracken, C. *J. Am. Chemi.Soc.* **2002**, *124*, 1852.
(5) Fushman, D.; Ghose, R.; Cowburn, D. *J. Am.Chem. Soc.* **2000**, *122*, 10640.
(6) Zweckstetter, M. *Nat. Protoc* . **2008**, *3*, 679.