Supplementary Materials

Efficient Long-Distance NMR-PRE and EPR-DEER Restraints for Two-Domain Protein Structure Determination

Kaiqi Wu [1], Chaowei Shi [1], Juan Li [1], Haipeng Wang [1], Pan Shi [1,2], Liu Chen [1], Fangming Wu [2], Ying Xiong[1], Changlin Tian [1,2,*]

[1] Hefei National Laboratory for Physical Science at the Microscale & School of Life Science, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China

[2] High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei, Anhui 230031, P. R. China

Materials and Methods:

Protein Expression and Purification, and Spin Labelling

The DNA sequence encoding $Rv0899_{52-326}$ was amplified from the *Mycobacterium Tuberculosis* DNA genome via PCR and cloned into pET21b (Novagen) between the *Nde* I and *Xho* I restriction sites. $Rv0899_{52-326}$ was expressed in *E. coli* strain BL21(DE3) and induced at 37°C for 5 hours. Over-expressed $Rv0899_{52-326}$ was initially purified using Ni-NTA affinity chromatography, followed by size exclusion chromatography (Superdex 75) on an ÄKTA Prime system (GE Healthcare). In the last step of protein purification, the final buffer was changed to 50 mM Na_2HPO_4/NaH_2PO_4, pH 6.5.

To prepare the PRE and DEER samples, cysteine mutations were introduced into $Rv0899_{52-326}$ using conventional site-specific mutation methods, and all mutated DNAs were verified by sequencing. Protein over-expression and purification were performed using the method described above. The purified ^{15}N-Rv0899 proteins were incubated with concentrated spin radical MTSL at a molar ratio of MTSL:protein = 10:1, in the presence of 1 mM DTT. The reactions were kept at 37°C for 3 hours, and the redundant MTSL was then removed using a size exclusion chromatography column (Superdex 75).

NMR Spectra Acquisition, Resonance Assignment and Relaxation Analysis

The two- and three-dimensional NMR spectra were recorded at 30°C on a Bruker AVANCE 600 MHz NMR spectrometer equipped with a cryo-probe. All NMR data were processed and analyzed using NMRPipe(Delaglio, et al., 1995),

NMRView(Johnson, 2004) and Sparky (T. D. Goddard and D. G. Kneller, UCSF).

Backbone ^{1}H, ^{15}N, and ^{13}C assignments were achieved by analyzing a series of TROSY-based three-dimensional data: HNCO, HN(CA)CO, HNCACB, and HN(CO)CACB experiments. Side chain ^{1}H and ^{13}C chemical shifts were assigned through several three-dimensional spectra: CCCONH, HBHACONH, HCCCONH and ^{15}N- and ^{13}C-edited NOESY experiments.

HSQC-based 2D pulse sequences were used to acquire backbone ^{15}N longitudinal relaxation (T_1), transverse relaxation (T_2), and heteronuclear ^{1}H-^{15}N steady-state NOEs of ^{15}N uniformly labelled $Rv0899_{52\text{-}326}$. The relaxation experiments were performed at 30°C on a Bruker AVANCE 500 MHz NMR spectrometer. T_1 values were determined from a series of ^{1}H-^{15}N correlation spectra with 11.7, 61.6, 142, 243, 364, 525, 757 and 1150 ms relaxation evolution delays. The T_2 values were obtained from the spectra with 0, 17.6, 35.2, 52.8, 70.4, 105.6 and 140.8 ms delays. The steady-state ^{1}H-^{15}N NOE values were determined from peak ratios observed between two spectra, one collected with a 3 s presaturation of protons and the other without proton presaturation.

Paramagnetic Relaxation Enhancement Measurements for $Rv0899_{52\text{-}326}$

For each spin labelled single-cysteine mutant, a pair of 2D ^{1}H-^{15}N TROSY-HSQC spectra were acquired for spin-labelled $Rv0899_{52\text{-}326}$: one for the spin-labelled protein in the paramagnetic form, and one in diamagnetic form that was reduced by the addition of 2−3-fold excess L-ascorbic acid. Samples were placed in the magnet at 30°C for at least 30 min before data acquisition, to ensure complete reduction of the

spin radicals. The spectra were then analyzed to measure PRE-based differences in peak intensities and line widths.

Intensity ratios of correlated peaks from the oxidized (I_{ox}, after radical labeling in a non-reducing condition) and reduced (I_{red}, with the excessive presence of reducing reagent L-ascorbic acid) spectra were converted into paramagnetic relaxation rate enhancements (PREs, R_2^{sp}) by estimating the additional transverse relaxation needed to reduce peak intensity relative to diamagnetic conditions by the observed intensity ratio using the following previously developed formula:

$$I_{ox} / I_{red} = R_2 \exp(-R_2^{sp} \tau) / (R_2 + R_2^{sp}) \quad \text{.....(1)}$$

Distances between the amide proton and nitrogen of the spin radical were calculated using the following equation:

$$r = [K/R_2^{sp} (4\tau_c + 3\tau_c / (1 + \omega_h^2 \tau_c^2))]^{1/6} \quad \text{.....(2)}$$

with the protein correlation time τ_c obtained from the backbone relaxation analysis of the $Rv0899_{52\text{-}326}$ protein.

Solution DEER Spectra Acquisition and Analysis

Double spin-labelled proteins were concentrated to approximately 0.25 mM for further DEER data collection using an ELEXSYS E580 pulsed EPR spectrometer operating at X-band with a dielectric ring resonator. All measurements were performed with an over-coupled resonator at 50 K, which was maintained using liquid helium. The four pulse DEER sequence was used, with the pump pulse frequency positioned at the center of the nitroxide spectrum, while the frequency of the observer pulses was increased by 80 MHz. The observe sequence used a 32 ns π-pulse, and the

pump π-pulse was typically 28 ns. The experiment repetition time was 4 μs.

Distance distribution was fitted by DeerAnalysis2006 (Jeschke, et al., 2006) using Tikhonov regularization with a broadening factor of 1 for radical pairs 117/302 and 117/266 (0.1 for 136/256), which was justified by the L-curves. The zero time was set to 80 ns, the background was selected from 0−674 ns for 117/302 (0−393 ns for 117/266, 0−528 ns for 136/256), and the cutoff time for the spectrum was set at 2152 ns for 117/302 (3252 ns for 117/266, 3600 ns for 136/256). The dipolar coupling evolution data were corrected for background echo decay using homogeneous three-dimensional spin distribution.

Protein Structure Calculation with NOE, PRE and DEER Restraints

Structural calculations were carried out using Xplor-NIH software (Schwieters, et al., 2003) starting with an extended $Rv0899_{52-326}$ template with randomized backbone torsion angles followed by 50 cycles of Powell energy minimization. Simulated annealing was carried out using 50,000 steps spanning 100 psec at 3,000 K followed by 100,000 × 0.003 psec steps with gradual cooling to 100 K over a span of 300 psec in steps of 50 K. During the final 50 ps of cooling, increased weight was applied to both covalent geometry and NOE-derived distance restraints. The van der Waals energy function was represented by a simple repel function and was increased during cooling by varying the force constant of the repel function from 0.003 to 4 kcal/(mol·$Å^4$) while decreasing the van der Waals radii from 0.9 to 0.75 Å. Finally, structures were energy-minimized using 250 steps of Powell energy minimization. The refinement protocol consisted of slow cooling from 1,000 K to 100 K spanning

45 ps. A force constant of 200 kcal/(mol·rad) was used for the dihedral angle restraints. Hydrogen bond restraints were enforced using flat-well harmonic potentials, with the force constant fixed at 20 kcal/(mol·Å^2). Flat-well harmonic potentials were applied for NOE, PRE and DEER restraints. A force constant of 50 kcal/(mol·Å^2) was applied for NOE, while 25 kcal/(mol·Å^2) was applied for PRE and DEER, only if they were included in the list of distance restraints.

Overall, 200 structures were generated by Xplor-NIH. The 20 lowest energy structures were subjected to further calculations. At this stage, the spin-labelled cysteine residues associated with the PRE restraints were changed back to their wild-type amino acids and Powell energy minimization was performed using Xplor-NIH. The final ten Rv0899$_{52\text{-}326}$ structures were analyzed with PROCHECK-NMR (Laskowski, et al., 1993) and visualized using MOLMOL (Koradi, et al., 1996). The coordinates of the ten conformers of Rv0899$_{75\text{-}326}$ have been deposited in the Protein Data Bank (Accession number: 2L26).

Procedure for Correlation of Distances

For further studies of the efficiency and effect of incorporating long-distance NMR-PRE and EPR-DEER restraints, three additional series of calculations were performed. At first, the structures of protein were acquired with only conventional solution NMR structural restraints (NOE, dihedral angles and hydrogen bonds), without PRE and DEER restraints. Then, a set of calcualtions were carried out to include the conventional solution NMR structural restraints and the total 428 PRE distance restraints. Afterwards, the third series of calculations were divided into three

sets and each was accomplished with the conventional solution NMR structural restraints, the total PRE restraints and two DEER restraint (calculations with 2DEER(117-266&136-266), 2DEER(117-302 & 136-256) or 2DEER(117-302 & 117-266)). All calculations followed similar procedures as mentioned in the above section.

Average distances from nitrogen of the spin radicals of residue G117, S136 and T256 to the amide proton were computed for the final 10 structures for each series of calculations mentioned above and compared to the average distances obtained from the final 10 models by all restraints. Distance between nitroxide radicals of two labels were also acquired for the final 10 structures of each series of calculation and compared to the values from the DEER experiment for correlation verification.

Figure S1. Domain organization of Rv0899 (Teriete, et al., 2010)**.**

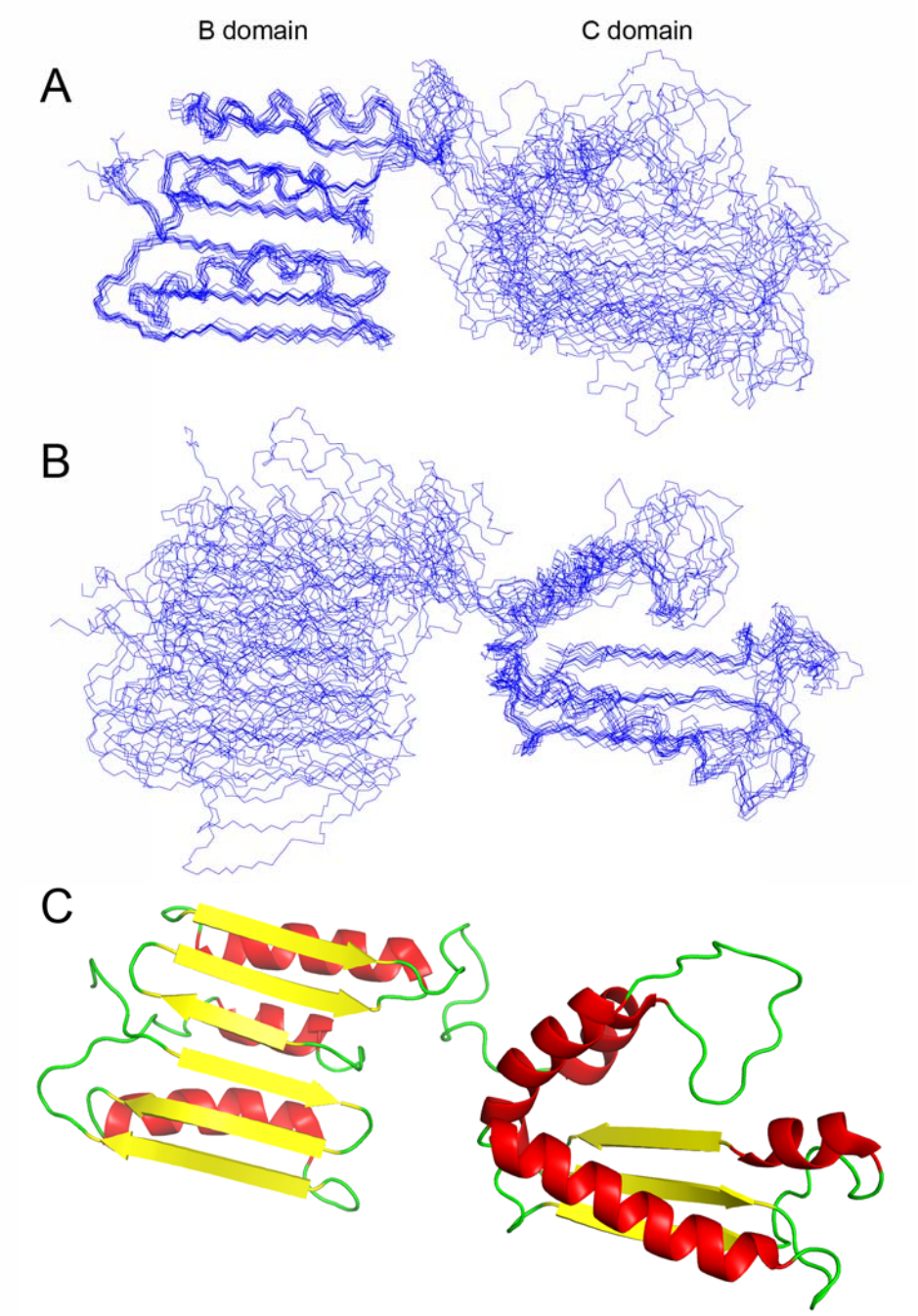

Figure S2. Structure of Rv0899$_{52-326}$ constrained by NOE, dihedral angle and hydrogen bond restraints. This structure is converged in the B domain (A) and C domain (B). (C) Cartoon representation of one conformer.

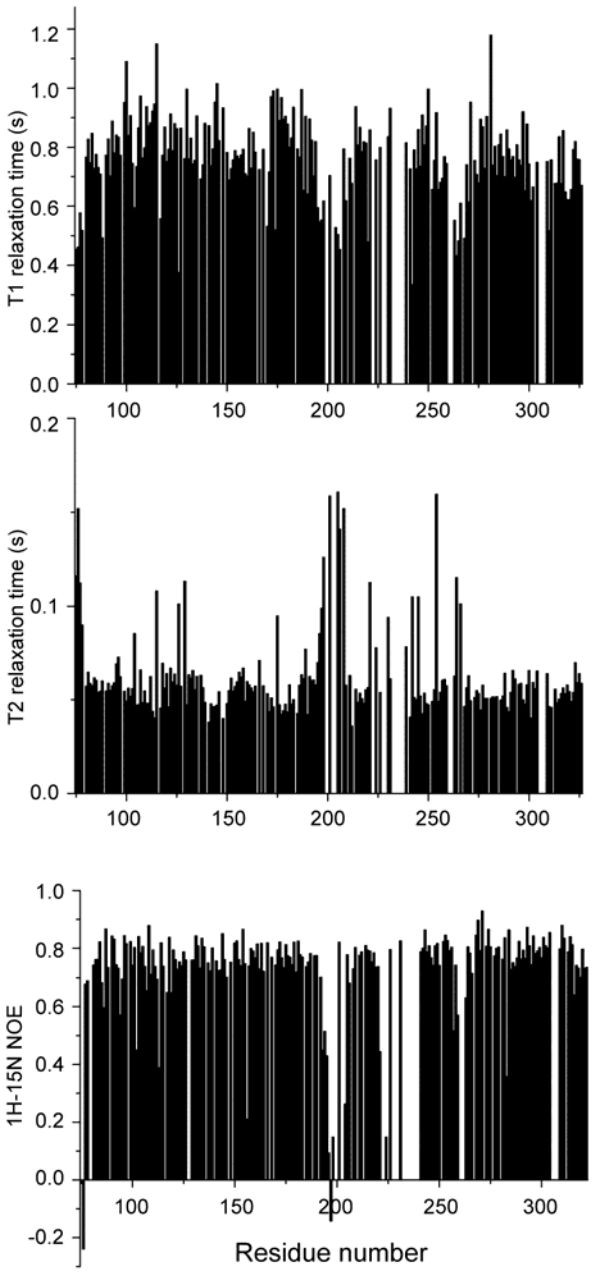

Figure S3. Backbone amide ^{15}N relaxation analysis of Rv0899$_{52-326}$ in a buffer of 50 mM Na_2HPO_4/NaH_2PO_4, pH 6.5 at 30°C. A quite uniform level of T_1, T_2 and NOE relaxation data were observed along the primary sequence of Rv0899$_{52-326}$, especially in the two domains, indicating similar global and local correlation times between the two domains.

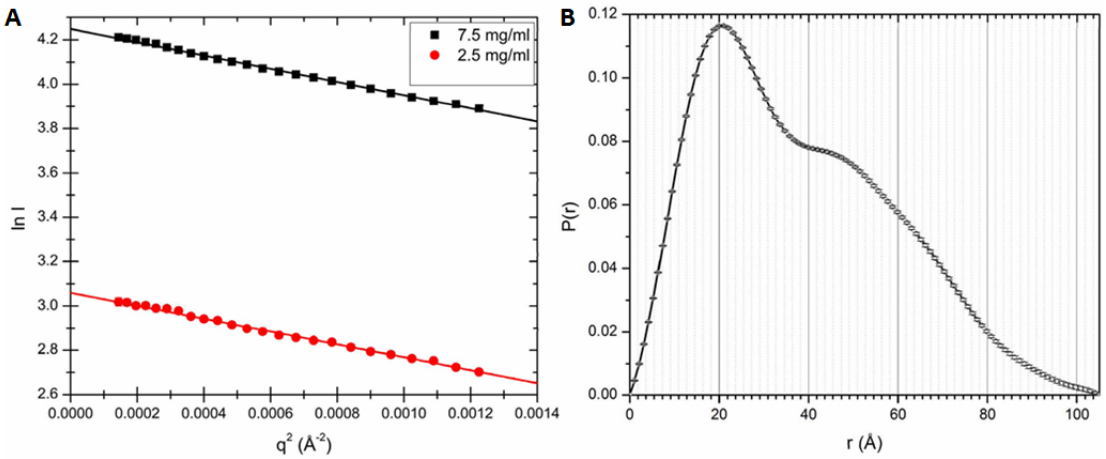

Figure S4. SAXS data analysis of Rv0899$_{52-326}$. (A) Guinier plot result for Rv0899 at different concentrations. **(B)** Pair distance distribution function (PDDF) fitting result for Rv0899.

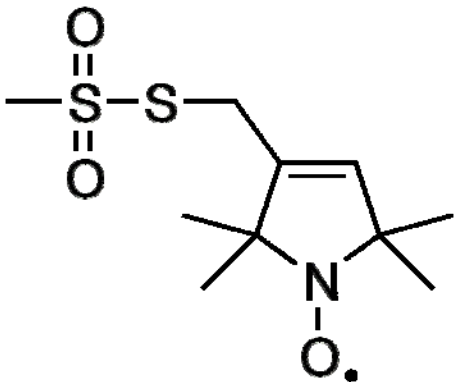

Figure S5. Chemical structure of spin radical label MTSL (S-(2,2,5,5-tetramethyl-2,5- dihydro-1H-pyrrol-3-yl) methyl methane sulfonothioate).

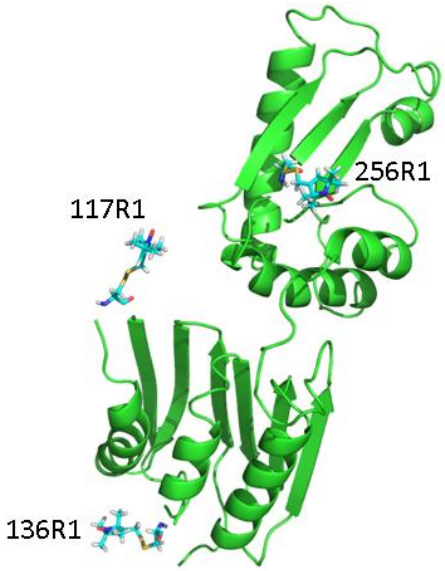

Figure S6. Positions of spin labels for the residues of 117, 136 and 256. The spin labels R1 were shown in sticks.

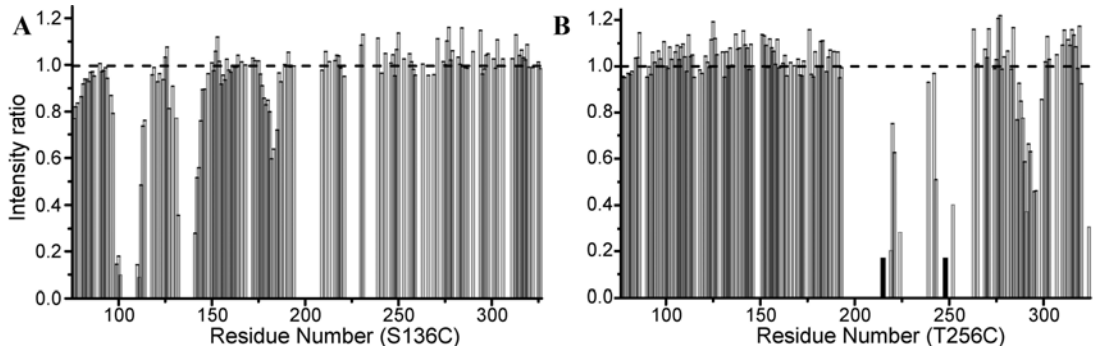

Figure S7. Histogram of intensity ratios of ^{1}H-^{15}N TROSY-HSQC cross peaks against Rv0899$_{52-326}$ residue number at two spin label positions. (A) S136C and (B) T256C. The absence of a bar at any position indicates overlap or a lack of assignment that prevented the quantification of the intensity ratio.

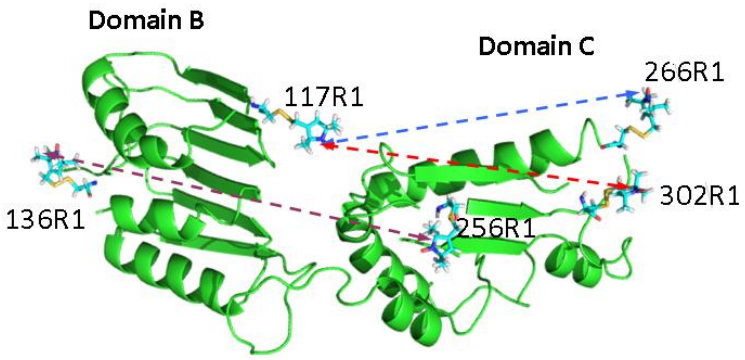

Figure S8. Pairs of spin labels for the DEER experiment. All the pairs are inter-domain as indicated. The spin labels R1 were shown in sticks.

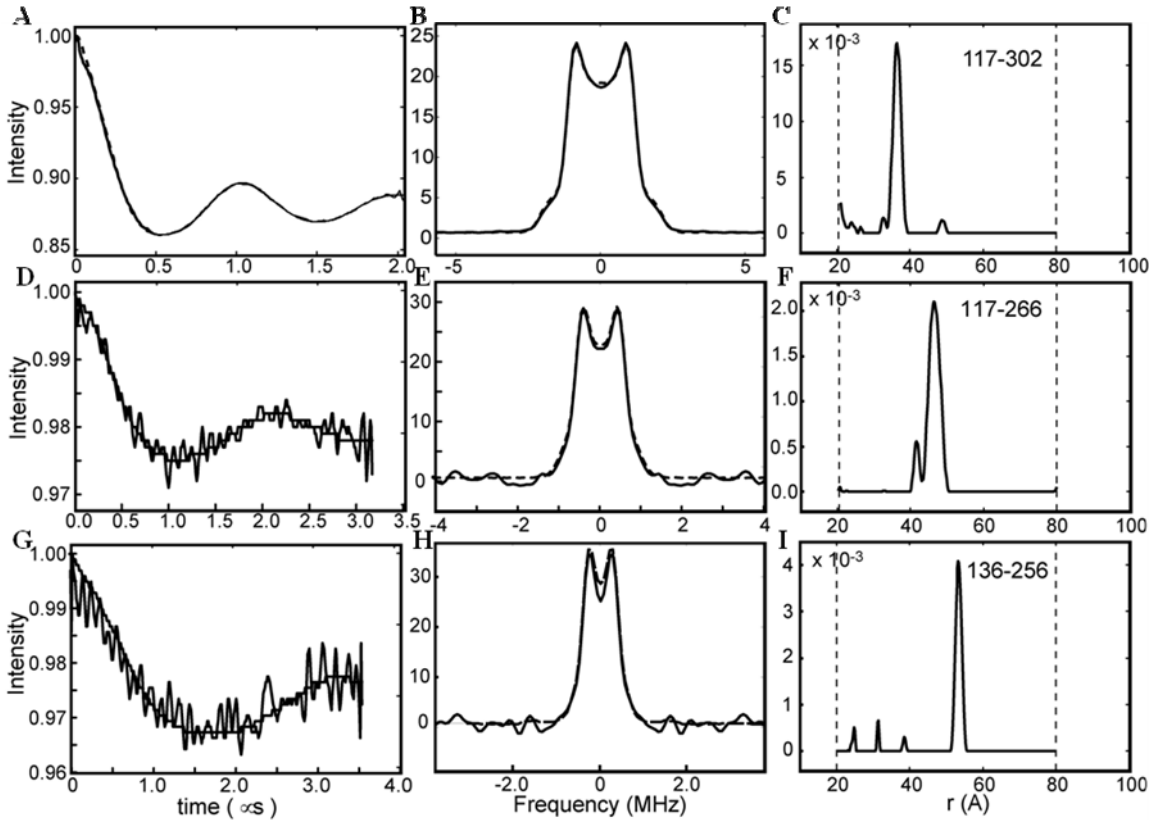

Figure S9. EPR-DEER data acquisition in time-domain, Fourier transformed data in frequency-domain, and derived distance distribution. Time-domain DEER signal for Rv0899 117-302 (A), 117-266 (D) and 136-256 (G) at 50 mM NaH_2PO_4/Na_2HPO_4, pH 7.2, in a final concentration of 30% glycerol. Frequency-domain dipolar spectrum of 117-302 (B), 117-266 (E) and 136-256 (H) (solid line) obtained from the Fourier transform of the DEER time-domain signal. The dashed line is the Fourier transform of the time-domain simulation of the data. Best fit of distance distribution for 117-302 (C), 117-266 (F) and 136-256 (I).

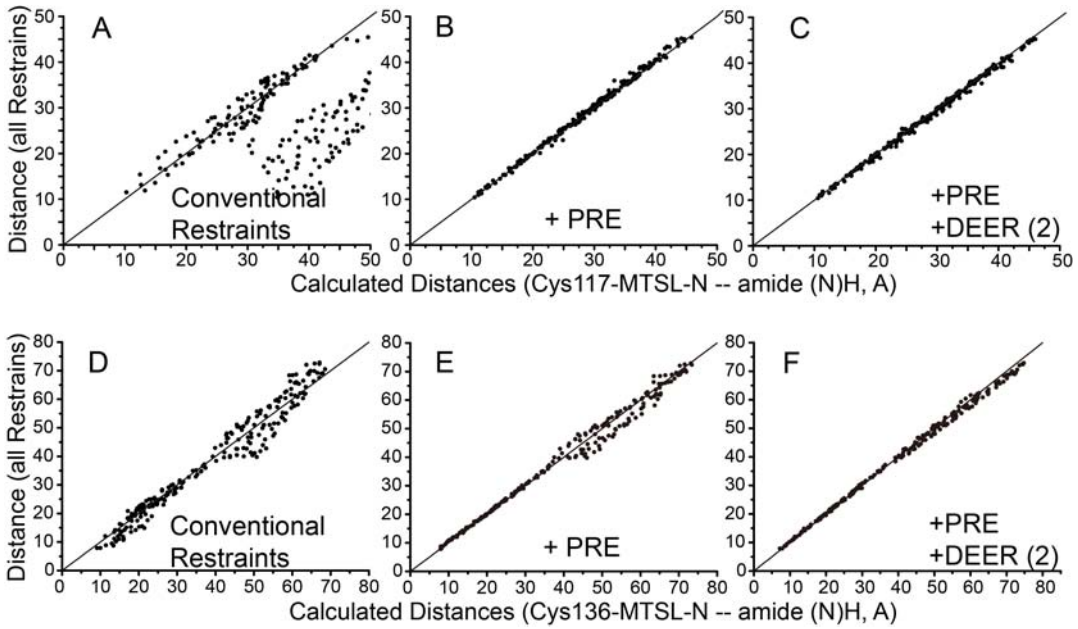

Figure S10. The correlation of distances from nitrogen of the spin radicals of residue 117 and 136 to the amide proton. Distances from the models by conventional restraints versus distances from the models by all restraints for residue 117 (A) and residue 136 (D). Distances from the models by conventional restraints together with all PRE restraints versus distances from the models by all restraints for residue 117 (B) and residue 136 (E). Distances from the models by conventional restraints together with all PRE restraints and two DEER restraints versus distances from the models by all restraints for residue 117 (C) and residue 136 (F).

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 6 (3): 277-293.

Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods in Molecular Biology. 278: 313-352.

Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J., Timmel, C., Hilger, D. and Jung, H. (2006). DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. Applied Magnetic Resonance. 30 (3): 473-498.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N. and Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. J Magn Reson. 160 (1): 65-73.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 26: 283-291.

Koradi, R., Billeter, M. and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. Journal of Molecular Graphics. 14 (1): 51-55, 29-32.

Teriete, P., Yao, Y., Kolodzik, A., Yu, J., Song, H., Niederweis, M. and Marassi, F. M. (2010). Mycobacterium tuberculosis Rv0899 adopts a mixed alpha/beta-structure and does not form a transmembrane beta-barrel. Biochemistry. 49 (13): 2768-2777.