

LETTER

Efficient long-distance NMR-PRE and EPR-DEER restraints for two-domain protein structure determination

Dear Editor,

The functional diversity of proteins is related to the cooperation of multiple domains. Independent globular domains are typically joined by a flexible length of polypeptide chain, which makes the structural analysis of multi-domain proteins difficult. Here, we describe the combined use of solution NMR (nuclear magnetic resonance) and EPR (electron paramagnetic resonance) for the structural analysis of a protein with two separate domains. The structure of each domain was determined independently using conventional NMR restraints, and the relative orientation of the two separate domains was confined using long-distance restraints obtained by NMR-PRE (paramagnetic relaxation enhancement) and EPR-DEER (double electron-electron resonance, also called PELDOR: pulsed electron double resonance).

A domain is the basic building block of the structure of a protein and an evolutionarily independent structural unit of a functional protein (Vogel et al., 2004). Certain protein domains have clearly defined functions and act as cornerstones in a variety of cellular processes. Genomic analyses have shown that over 70% of eukaryotic proteins are multi-domain proteins (Han et al., 2007). The modular nature of a multi-domain protein provides stability and new cooperative functions (Bhaskara and Srinivasan, 2011). The domains of these proteins are normally connected by flexible linkers. Therefore, determining the relative orientation of these domains is critical to our understanding of domain interactions and the functional mechanisms of

a multi-domain protein.

One major issue in the structure determination of a multi-domain protein is obtaining inter-domain restraints, especially when the domains are separated. NMR methods such as RDC (residual dipolar coupling) (Lipsitz and Tjandra, 2004), PCS (pseudocontact shift) (Schmitz et al., 2012), and PRE (paramagnetic relaxation enhancement) (Kosen, 1989) are all capable of generating orientational or long range restraints to confine the multiple domains of a protein. EPR spectroscopy also provides a variety of tools to study structures and conformational changes of biomacromolecules and complexes (Jeschke and Polyhach, 2007). DEER is an EPR method that measures the dipolar electron-electron coupling distances in the nanometre range with high precision and reliability (Jeschke and Polyhach, 2007). This technique was successfully used in the structure determination of a homodimer protein (Yang et al., 2010), the orientation determination of a membrane protein tandem domain pair (Ward et al., 2009), and to monitor conformational changes in GPCR (G protein coupled receptor) activation (Altenbach et al., 2008).

The PRE and DEER methods are complementary since they cover different distance ranges (PRE: ~14–23 Å; DEER: ~20–80 Å), and they are also good extensions to short distance restraints such as those derived from the nuclear Overhauser effect (NOE) (~6 Å). Long-distance restraints obtained using PRE and EPR are potentially valuable for studying the relative orientation of two subunits in a protein complex or two

domains in a multi-domain protein.

The potential of long distance restraints for multi-domain protein structural studies was demonstrated by combining NMR-PRE and EPR-DEER measurements to determine the three-dimensional structure of Rv0899 from *Mycobacterium tuberculosis*, which contains two domains. Rv0899 is essential for the adaptation of *M. tuberculosis* to the acidic environment of the phagosome (Raynaud et al., 2002). The 326-residue protein contains three domains, an N-terminal transmembrane (TM) domain, a B domain, and a C domain (Fig. S1) (Teriete et al., 2010). Previously, the solution structures of the B and C domains were determined individually (Li et al., 2012). Recombined Rv0899_{52–326} containing both B domain and C domain was expressed and purified as previously described (Li et al., 2012). Conventional solution NMR structural restraints (NOE, dihedral angles and hydrogen bonds) were collected for structural calculation. A total of 2054 NOE, 300 dihedral angle and 88 hydrogen bond restraints were used for structure calculation, and the results showed the converged structures of the B and C domains (Fig. S2). However, because no inter-domain NOE-restraints were assigned, the relative orientation of the two domains remained random.

Before analyzing the relative orientation of two-domain proteins, the rigid conformation of the two domains needs to be validated. Analysis of the backbone amide ¹⁵N relaxation data (T₁, T₂ and NOE) (Fig. S3) showed that the three relaxation values were at the same level along the primary sequence of

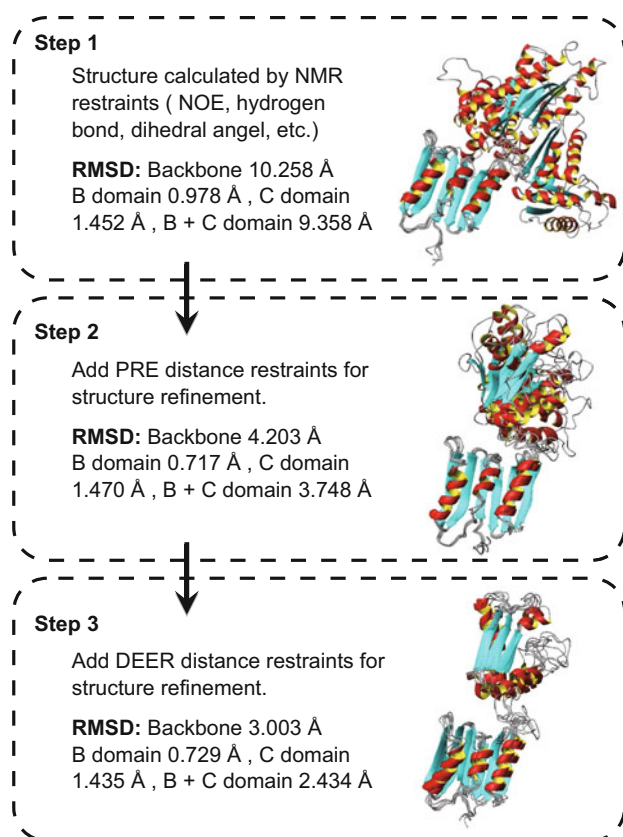


Figure 1. Flow chart of the 3-step procedure used to determine the structure of Rv0899_{52–326}, showing the gradual improvement in the quality of the resultant structures. Step 1, only conventional NMR restraints (NOE, dihedral angles, and H-bonds) were used in the structure calculation; Step 2, 428 PRE restraints were included in the structure calculation; Step 3, three DEER restraints as well as 428 PRE restraints were incorporated into the structure calculation.

Rv0899_{52–326}, which strongly indicated that the two domains have similar diffusion properties and may be in close contact or even packed together. Furthermore, Rv0899_{52–326} samples were prepared for small angle X-ray scattering (SAXS) data acquisition (Fig. S4). The Guinier plot showed the pattern of a monodispersed system, indicating the uniform size and shape of the Rv0899 protein molecules in solvent, which could only be satisfied only if two domains of Rv0899 had confined relative orientation to each other. From the Pair Distance Distribution Function (PDDF) of Rv0899, a typical elongated shape of the protein molecule was present rather than a globular or flattened shape (Koch et al., 2003). According to the overall shape derived from SAXS data, Rv0899_{52–326}

had a substantially rigid conformation in the aqueous buffer.

Verification of the rigidity or inflexibility of the two domains provided a basis for further structural studies of Rv0899_{52–326}, and two approaches (NMR-PRE and EPR-DEER) were applied to collect long-range restraints between these two domains.

In NMR-PRE experiments, three residues (G117, S136 and T256) were respectively mutated to cysteine and then labelled with the spin radical compound MTSL (Fig. S5) through disulfide bond formation between MTSL and cysteine residues (positions shown in Fig. S6). HSQC spectra of MTSL-labelled Rv0899_{52–326} were acquired to verify that the protein did not undergo significant conformational changes after spin label-

ling. The intensity attenuation of HSQC amide proton resonance was analysed for proteins with MTSL labelling at different sites under oxidative or reduced conditions (Battiste and Wagner, 2000), and the intensity attenuation was converted to distance restraints following previously described methods (detail procedures in Supplementary Material) (Battiste and Wagner, 2000). The distance restraints calculated from the collected signal intensity attenuations (Fig. S7A and S7B) were classified into different categories with different uncertainties for further structural calculation: for peaks with a paramagnetic/diamagnetic sample intensity ratio of <0.85, the distance restraints were calculated (from equation (2) in Supplementary Material) and were assigned uncertainties of ± 4 Å; for peaks that were broadened beyond detection in the paramagnetic spectrum, distances were set to be <14 Å; for peaks with intensity ratios in the range of 0.85–1.0 the distance restraints were set as >23 Å.

A total of 428 PRE distance restraints were derived and then used for structure calculations. The two domains were well converged, and appeared in two separate groups. The overall backbone root-mean-square-deviation (RMSD) of Rv0899 was significantly reduced from 10.258 Å to 4.203 Å (Fig. 1, step 1 \rightarrow step 2), and minor improvements were made in the backbone RMSD in the B and C domains. This strongly suggests that the PRE restraints are very valuable in determining the relative orientation of two domains.

In the EPR-DEER experiments, MTSL was used to label three cysteine pairs in Rv0899_{52–326} (117–302, 117–266 and 136–256, positions shown in Fig. S8). The DEER spectrum for the MTSL labels at 136–256 was collected and Fourier transformed to derive distance distribution fitted by DeerAnalysis 2006 (Jeschke, et al., 2006) using Tikhonov regularisation with justified broadening factor (Fig. S9G and S9H). A large distance of up to 57 Å was measured between the MTSL radical pair (Fig. S9I). In total, three sets of DEER-derived distance restraints were col-

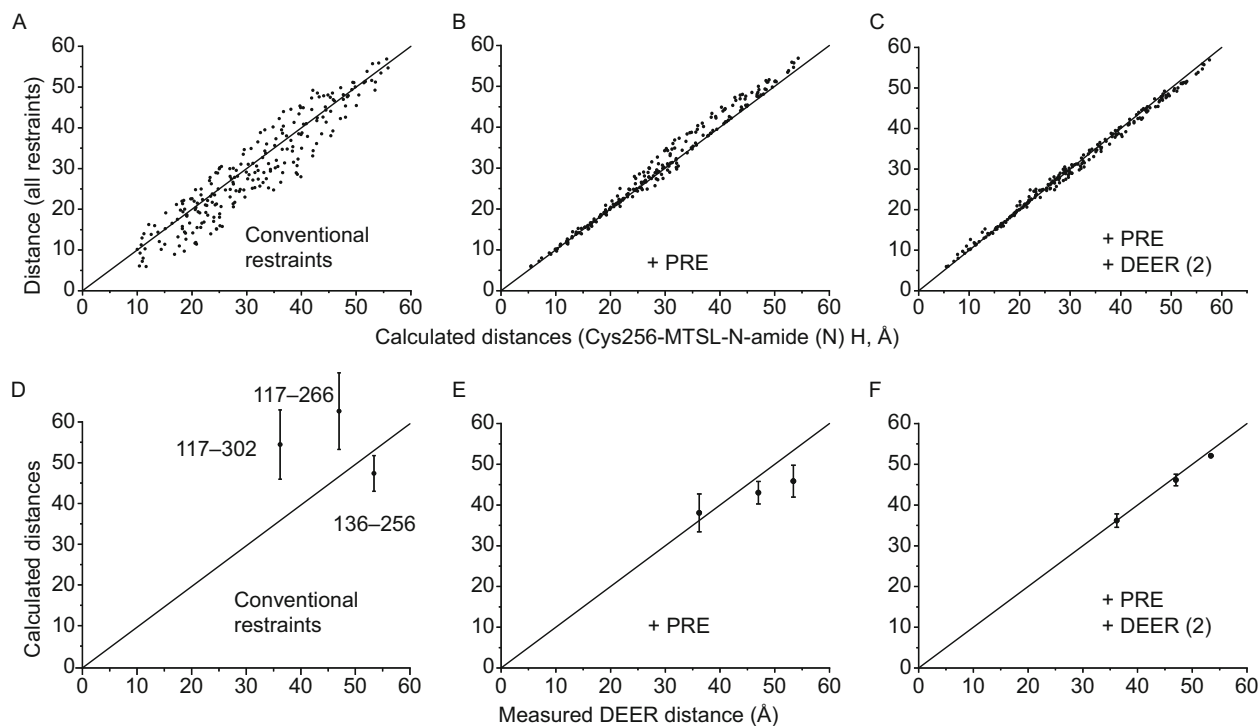


Figure 2. The correlation of distances from the spin radical nitrogen atom at residue 136 to the amide protons and the correlation of distances between nitrogen atoms of the spin radical pair from models and the values from the DEER experiment.

(A) Distances from the models by conventional restraints versus distances from the models by all restraints. (B) Distances from the models by conventional restraints together with all PRE restraints versus distances from the models by all restraints. (C) Distances from the models by conventional restraints together with all PRE restraints and two DEER restraints versus distances from the models by all restraints. (D) Distances from the models by conventional restraints versus the DEER experimental values for the three radical pairs. (E) Calculated distances from the models by conventional restraints together with all PRE restraints versus the DEER experimental values for the three radical pairs. (F) Calculated distances from the models by conventional restraints together with all PRE restraints and two DEER restraints versus the DEER experimental values for the three radical pairs.

lected and applied for further structure computation (Fig. S9). The DEER distance restraints were applied together with the uncertainties of its line-width as follows: ± 2.5 Å for the spin radical pair 117–302, ± 3.5 Å for the spin radical pair 117–266 and ± 2.0 Å for the spin radical pair 136–256. A converged relative orientation between these two domains was obtained when the DEER distance restraints were applied together with NOE and PRE restraints during structure calculation. The backbone RMSD of the final 10 structures were significantly reduced from 4.203 Å to 3.003 Å (Fig. 1, step 2 \rightarrow step 3) when only 3 DEER restraints were included in the structure calculation, indicating the efficiency of long-range restraints obtained from DEER in multi-domain structure determination.

Further more, the correlation analysis was conducted for distances from nitrogen of the spin radicals at residue 117, 136 and 256 to the backbone amide protons of every residue of Rv0899 (Figs. 2A, 2B, 2C and S10). As an example at residue 256, the correlation showed gradual improvement with addition of long-range restraints. In Fig. 2A, the correlation was rather dispersed for the results of conventional restraints due to the large overall backbone RMSD. After incorporation of PRE restraints (Fig. 2B), the correlation improved significantly, especially for the short distances corresponding to the range where PRE took effects. The addition of two extra DEER restraints further improved the long-range distances correlation as a result of the confined orientations between two domains (Fig. 2C). Correlations of dis-

tances from the experimental DEER and calculated results by different constraints were also studied for the three radical pairs of 117–302, 117–266 and 136–256 (Fig. 2D–F). The Fig. 2D shows low correlation between the calculated distances and that of experimental values with large deviation and error bar. Further improvement of correlation could be observed after the addition of PRE constraints as indicated in Fig. 2E. After addition of two additional DEER long distance restraints, the average distances of three radical pairs fitted well to the DEER experimental values with small error bars (Fig. 2F). The analysis of correlation of distances further verified the efficiency of adding long-range restraints for improvement structures.

In summary, the combined application of NMR-PRE and EPR-DEER long

distance measurements together with conventional solution NMR restraints showed significant potential for structural studies of large proteins, including protein complexes and multi-domain proteins. This method is valuable not only for the determination of three-dimensional structure, but also for the analysis of conformational changes in different functional states. Our findings suggest that combinational methods can be applied to answer important biological questions involving multi-domain proteins.

FOOTNOTES

The authors are grateful for help provided by Mr. Ya Wang and Prof. Jiangfeng Du from the University of Science and Technology of China by allowing access to their E-580 pulse EPR instrument. We also thank Prof. John M. Flanagan in Penn. State University for SAXS data acquisition and analysis. This research was supported by the National Basic Research Program (973 Program) (Nos. 2011CB911104 and 2013CB910200) and the Chinese Natural Science Foundation of China (Grant No. 31100563) to Y. Xiong and (Grant No. 31170817) to C. Tian.

Kaiqi Wu, Chaowei Shi, Juan Li, Haipeng

Wang, Pan Shi, Liu Chen, Fangming Wu, Ying Xiong, and Changlin Tian declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

Kaiqi Wu¹, Chaowei Shi¹, Juan Li¹, Haipeng Wang¹, Pan Shi^{1,2}, Liu Chen¹, Fangming Wu², Ying Xiong¹, 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 230026, China

² High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei 230031, China

✉ Correspondence: cctian@ustc.edu.cn

REFERENCES

Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., and Teichmann, S. A. (2004). *Curr Opin Struct Biol* 14, 208–216.

Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007). *Nat Rev Mol Cell Biol* 8, 319–330.

Bhaskara, R. M., and Srinivasan, N. (2011). *Sci Rep* 1, 40.

Lipsitz, R. S., and Tjandra, N. (2004). *Annu Rev Biophys Biomol Struct* 33, 387–413.

Schmitz, C., Vernon, R., Otting, G., Baker, D., and Huber, T. (2012). *J Mol Biol* 416, 668–677.

Kosen, P. A. (1989). *Method Enzymol* 177, 86–121.

Jeschke, G., and Polyhach, Y. (2007). *Phys Chem Chem Phys* 9, 1895–1910.

Yang, Y., Ramelot, T. A., McCarrick, R. M., Ni, S., Feldmann, E. A., et al. (2010). *J Am Chem Soc* 132, 11910–11913.

Ward, R., Zoltner, M., Beer, L., El Mkami, H., Henderson, I. R., et al. (2009). *Structure* 17, 1187–1194.

Altenbach, C., Kusnetzow, A. K., Ernst, O. P., Hofmann, K. P. and Hubbell, W. L. (2008). *Proc Natl Acad Sci U S A* 105, 7439–7444.

Raynaud, C., Papavinasasundaram, K. G., Speight, R. A., Springer, B., Sander, P., et al. (2002). *Mol Microbiol* 46, 191–201.

Li, J., Shi, C., Gao, Y., Wu, K., Shi, P., et al. (2012). *J Mol Biol* 415, 382–392.

Koch, M. H. J., Vachette, P., and Svergun, D. I. (2003). *Q Rev Biophys* 36, 147–227.

Battiste, J. L., and Wagner, G. (2000). *Biochemistry* 39, 5355–5365.

Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., et al. (2006). *Appl Magn Reson* 30, 473–498.