J|A|C|S COMMUNICATIONS

Published on Web 08/10/2010

Combining NMR and EPR Methods for Homodimer Protein Structure Determination

Yunhuang Yang,[†,‡] Theresa A. Ramelot,[†,‡] Robert M. McCarrick,[†] Shuisong Ni,[†] Erik A. Feldmann,[†,‡] John R. Cort,[‡,§] Huang Wang,[‡,||] Colleen Ciccosanti,[‡,||] Mei Jiang,[‡,||] Haleema Janjua,[‡,||] Thomas B. Acton,[‡,||] Rong Xiao,[‡,||] John K. Everett,[‡,||] Gaetano T. Montelione,[‡,||,⊥] and Michael A. Kennedy*[,†,‡]

Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056, Northeast Structural Genomics Consortium, Piscataway, New Jersey 08854, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, and Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Received June 24, 2010; E-mail: kennedm4@muohio.edu

Abstract: There is a general need to develop more powerful and more robust methods for structural characterization of homodimers, homo-oligomers, and multiprotein complexes using solution-state NMR methods. In recent years, there has been increasing emphasis on integrating distinct and complementary methodologies for structure determination of multiprotein complexes. One approach not yet widely used is to obtain intermediate and long-range distance constraints from paramagnetic relaxation enhancements (PRE) and electron paramagnetic resonance (EPR)-based techniques such as double electron electron resonance (DEER), which, when used together, can provide supplemental distance constraints spanning to 10−70 Å. In this Communication, we describe integration of PRE and DEER data with conventional solution-state nuclear magnetic resonance (NMR) methods for structure determination of Dsy0195, a homodimer (62 amino acids per monomer) from *Desulfitobacterium hafniense*. Our results indicate that combination of conventional NMR restraints with only one or a few DEER distance constraints and a small number of PRE constraints is sufficient for the automatic NMR-based structure determination program CYANA to build a network of interchain nuclear Overhauser effect constraints that can be used to accurately define both the homodimer interface and the global homodimer structure. The use of DEER distances as a source of supplemental constraints as described here has virtually no upper molecular weight limit, and utilization of the PRE constraints is limited only by the ability to make accurate assignments of the protein amide proton and nitrogen chemical shifts.

Homodimer and multiprotein complex structure determination using solution NMR methods usually relies on interchain nuclear Overhauser effects (NOEs) detected and assigned in isotope-filtered NOE experiments.[1−3] These techniques, which require preparation of equal mixtures of isotopically labeled and unlabeled samples, are often insensitive, can be difficult to interpret due to imperfect artifact suppression, and can fail to provide unambiguous interchain NOEs. Poor NMR data can result from one of several homodimer characteristics, including (1) tight dimer association resulting in unfavorable chain exchange kinetics and a low population of mixed $^{13}C/^{12}C$-labeled dimers, (2) weak dimer association preventing measurement of interchain NOEs, or (3) a homodimer interface whose nature prevents measurement or interpretation of interchain NOEs. Inability to measure and identify interfacial NOEs in isotope-filtered experiments can make it difficult or impossible to assign interchain NOEs in unfiltered, ^{13}C- and ^{15}N-edited NOESY data, and this can prevent successful homodimer structure determination using conventional NMR methods alone. Programs such as CYANA,[4] which are designed to automatically assign NOESY cross-peaks, are sometimes capable, when seeded with just a few unambiguous interchain NOEs from isotope-filtered experiments, of assigning a sufficient network of interchain NOEs from unfiltered ^{13}C- and ^{15}N-edited NOESY data to accurately determine a homodimer structure. However, CYANA and similar programs generally fail to make productive interchain NOE assignments from unfiltered edited NOESY data if the rotational and translational solution space of one monomer is unrestricted relative to the other.

Problems associated with reliance on edited-filtered NOESY data for NMR-based structure determination of homo-oligomers and protein complexes are exacerbated as the protein systems become larger and more complex. In general, more powerful and robust methods are needed for structural characterization of homodimers, homo-oligomers, and protein complexes using solution-state NMR methods. This challenge is of great biological significance; for example, more than 80% of *Escherichia coli* proteins are predicted to exist as homodimers or higher-order oligomers, and the function of these protein complexes is predicted to be associated with their structural symmetry.[5] Solution-state NMR-based structure determination of homo-oligomers and multiprotein complexes is also growing concern for structural genomics efforts. In particular, multiprotein complexes represent an increasing fraction of targets of the National Institutes of Health's Protein Structure Initiative (www.nigms.nih.gov/Initiatives/PSI/).

In recent years, there has been an increasing emphasis on integrating distinct and complementary methodologies to obtain supplemental constraints that can be used in combination with conventional solution-state NMR methods for structure determination of multiprotein complexes. For example, traditional NMR data have been combined with residual dipolar couplings (RDCs),[6] chemical shift (CS) Rosetta,[7,8] and small-angle X-ray scattering (SAXS)[9] data in order to solve homodimeric protein structures.

† Miami University.
‡ Northeast Structural Genomics Consortium.
§ Pacific Northwest National Laboratory.
|| Rutgers, The State University of New Jersey.
⊥ University of Medicine and Dentistry of New Jersey.

 10.1021/ja105080h

One approach not yet widely used for multiprotein structure determination is to combine solution-state NMR-derived NOESY constraints with intermediate and long-range distance constraints obtained from both paramagnetic relaxation enhancements (PRE)[10,11] and electron paramagnetic resonance (EPR)-based techniques such as double electron electron resonance (DEER),[12−15] which, when used together, can provide additional distance constraints spanning to 10−70 Å. These supplemental distance measurements can be used for structure validation, for filtering candidate oligomer or multiprotein structures, or as additional constraints for *de novo* structure calculations. In this Communication, we describe integration of PRE and DEER data with conventional solution-state NMR methods for structure determination of the homodimer Dsy0195 protein (NESG target name DhR8C) from *Desulfitobacterium hafniense*.

Dsy0195 belongs to the YabP family (PF07873), which is involved in spore coat assembly during the process of sporulation.[16] The crystal structure of this protein is now available (PDB ID: 3IPF).[17] In solution, it is a homodimer of 62 residues per monomer. 1D ^{15}N T_1/T_2 measurements (τ_c = 12.2 ± 0.7 ns at 20 °C) and gel-filtration chromatography with mass analysis by static light scattering confirmed the dimeric state of Dsy0195 in solution (Figure S1, Supporting Information). It exhibited only a few unambiguous interchain ^{13}C-edited/^{12}C-filtered NOEs, which were insufficient for homodimer structure determination using the CYANA procedure described below. The wild-type (wt) Dsy0195 protein sequence contains no cysteines. Therefore, spin labels were introduced following site-directed mutagenesis. Amino acids targeted for site-directed spin labeling (SDSL) were identified from preliminary NMR structures of the monomer. Two serine residues (S36 and S52) occurring in structured β-sheet regions of the protein were selected for replacement by cysteine in two different mutant constructs. The crystal structure was not available at the time residues were selected for SDSL. (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate (MTSL) was used to spin-label free cysteine in the S36C and S52C mutants. Excess and/or free MTSL was removed by size exclusion chromatography prior to PRE or DEER measurements.

Continuous-wave (cw) EPR spectra were measured to confirm that MTSL was covalently bound to the cysteines in the S36C and S52C mutants (Figure S2, Supporting Information). Comparison of two-dimensional (2D) ^{15}N-HSQC spectra of ^{15}N-uniformly labeled wt, S36C, and S52C mutants indicated that the Dsy0195 structure was not perturbed by the S36C and S52C mutations (Figure S3, Supporting Information). 2D ^{15}N-HSQC spectra of the mutants covalently bound to MTSL demonstrated that resonances either were broadened beyond detection for residues near the S36C or S52C mutation sites or experienced substantial chemical shift perturbations or significant broadening for residues more remote from the mutation sites (Figure S4, Supporting Information).

Electron transverse relaxation times of covalently bound MTSL in the S36C and S52C mutants were measured to be 1.77 and 1.93 μs, respectively (Figure S5, Supporting Information), which were long enough to enable collection of high signal-to-noise DEER data at Q-band (34 GHz). Distance distributions were obtained from Tikhonov regularization simulations using the program DeerAnalysis2009.[18] The intensity of the DEER modulation as a function of the time T for S36C is shown in Figure 1A, and the corresponding dipolar spectrum is shown in Figure 1B. The best fit of the distance distribution between nitroxide spin-labels in the S36C mutant homodimer had a peak at 34 Å after Tikhonov regularization (Figure 1C), consistent with the distance measured from crystal structure (33.5 Å between C_β of S36 in the dimer A and B chains). While

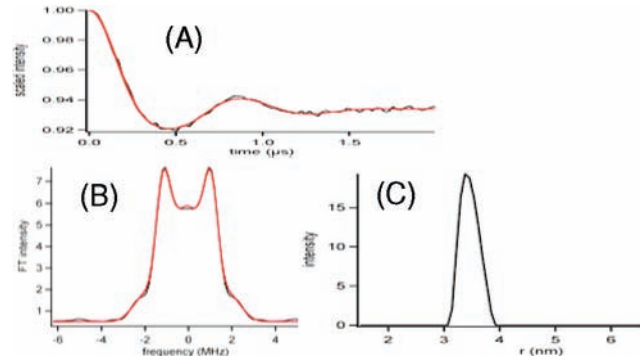

Figure 1. (A) Time-domain DEER signal of 0.25 mM S36C-MTSL at 20 mM ammonium acetate, 200 mM NaCl, and 5 mM $CaCl_2$ (pH 4.5) diluted to a final concentration of 30% (w/w) glycerol. The black line represents the three-dimensional homogeneous background-subtracted experimental data, and the red line is the time-domain simulation of the data. (B) Dipolar spectrum (black line) obtained from the Fourier transform of the DEER signal in (A). The red line is the Fourier transform of the time-domain simulation of the data. (C) Best fit of distance distribution obtained from Tikhonov regularization. The DEER experiment was carried on a Bruker ELEXSYS E580 pulsed EPR instrument at 80 K.

the agreement between the DEER distance and the distance measured from the crystal structure is striking, the apparent precision of the agreement is coincidental since multiple rotamer states can be populated for each MTSL-modified cysteine.[19] The distance measured between MTSL spin labels from DEER data for the S52C mutant was 20 Å (Figure S6, Supporting Information), or about 5 Å longer than the distance between C_β atoms of S52 measured from the crystal structure of the native protein. While cysteine side-chain flexibility causes uncertainty in the distance between tethered MTSL spin labels, the uncertainty in the distance interpretation becomes relatively less important as the distance between spin labels becomes longer.[20,21] If the distances between C_β atoms of the cysteines modified by MTSL are used directly as distance constraints in structure calculations, use of multiple DEER distance measurements should compensate for individual uncertainties. It should also be pointed out that, while the lower limit for distance determination using DEER is around 1.6−1.9 nm, cw EPR can be used as an alternative method for label-to-label distance measurement, which has an upper limit of around 1.6−1.9 nm.[22]

Following the PRE analysis procedure proposed by Rumpel et al.,[11] 67 lower-bound and 35 upper-bound interchain constraints were derived from a 1:1 mixture of ^{15}N-labeled/wt-Dsy0195 and unlabeled/MTSL-spin-labeled S52C mutant protein. Constraints derived from PRE data (with 4 Å bounds) were consistent with the crystal structure (Figure S7, Supporting Information). Useful distance constraints derived from PRE experiments spanned a lower limit of ∼10 Å (resonances for atoms closer than ∼10 Å to the spin label were broadened beyond detection) and an upper limit of ∼18 Å (beyond which no PRE could be detected). Due to a number of potential sources of error for interpreting PRE distances,[10,11] constraints between MTSL and side chains or residues in loop regions, which were more flexible, were neglected. After culling PREs that likely contained errors for reasons just discussed, a final set of 29 lower-bound and 8 upper-bound interchain constraints were included in the calculations.

The CYANA program was used to combine NMR and EPR data for structure calculations. Automated CYANA structure calculations typically involve about seven cycles. In the initial cycle, network anchoring and distance constraint combinations enable an efficient and reliable search for the correct fold. Using 3D structure-based filters, additional NOEs are automatically assigned in the following

Table 1. Correlation between Number of Final Interchain NOEs Assigned and Backbone Root-Mean-Square Deviation to the Crystal Structure of the Average Structure Calculated Using CYANA 2.1 Seeded with Different Interchain Constraints

constraint seeds	no. assigned[a]	rmsd[b]
DEER (S36C)	130	1.45
DEER (S52C)	118	1.90
DEER (both)	120	1.53
PRE (10%)[c]	100	1.96
PRE (50%)[c]	146	1.52
PRE (100%)[c]	154	1.43
PRE (10%)[c] and DEER (both)	152	1.52
PRE (50%)[c] and DEER (both)	146	1.52
PRE (100%)[c] and DEER (both)	156	1.28

[a] Number of assigned interchain NOEs using CYANA 2.1. [b] Backbone rmsd to the crystal structure of average structure calculated by use of CYANA 2.1 against crystal structure. [c] Percentages of total interchain constraints from PRE data used for structure calculation.

cycles.[4] Here, we tested the power of seeding CYANA calculations with interchain constraints derived from both DEER and PRE experiments. Three steps were taken for homodimer structure calculations. First, the monomer structure of Dsy0195 was calculated with conventional NMR methods. All NOEs were automatically assigned by CYANA. Some erroneous intrachain assignments of interchain NOEs caused bad quality of the monomeric structure with a high MolProbity Clashscore. The Protein Structure Validation Suite (PSVS version 1.4) program[23] was used to help identify incorrectly assigned ambiguous NOEs, which were removed, and the final well-folded monomeric structure was determined. Second, final intrachain NOEs, hydrogen bond restraints, and dihedral angle constraints obtained from the monomeric structure calculations were used to seed the homodimer structure calculation. CYANA calculations in the absence of seeding with interchain constraints resulted in two separate, well-folded monomer subunits with no dimer interface. Third, different numbers of intermediate and long-range interchain constraints derived from DEER and PRE experiments were used to seed CYANA calculations (Table 1). As shown in Table 1, seeding with just one long-range DEER distance enabled CYANA to generate between 118 and 130 interchain NOE constraints and to reproduce the correct dimer interface. The resulting structures were checked to determine numbers and accuracy of final interchain NOE assignments by comparing the average structure to the crystal structure. In general, seeding with more long-range interchain constraints from either DEER or PRE increased the number of interchain NOEs correctly assigned by CYANA and improved the backbone root-mean-square deviation (rmsd) with the crystal structure.

The best structures resulted from seeding CYANA with both experimentally derived DEER distances and all available PRE constraints. In this case, CYANA assigned 156 interchain NOEs, which were combined with intrachain NOEs, dihedral angle constraints, and hydrogen bond constraints for structure calculations using Xplor-NIH, followed by refinement with CNS-water. During the calculation and refinement cycles, about 31 ambiguously assigned interchain NOEs were manually removed for reasons described above. For the final calculation cycle, 125 interchain NOEs remained. The final ensemble of 20 structures with lowest energy has deposited in the Protein Data Bank (PDB 2KYI). The quality of final structure was assessed using PSVS version 1.4 (Table S1, Supporting Information). Structural details of Dsy0195 will be discussed elsewhere. The final NMR-PRE-DEER average structure and crystal structure are shown in Figure 2. The backbone rmsd of average NMR-PRE-DEER structure against crystal structure was 1.13 Å, which is significantly better than for the structure obtained using NMR data alone (1.51 Å, PDB 2KS0).

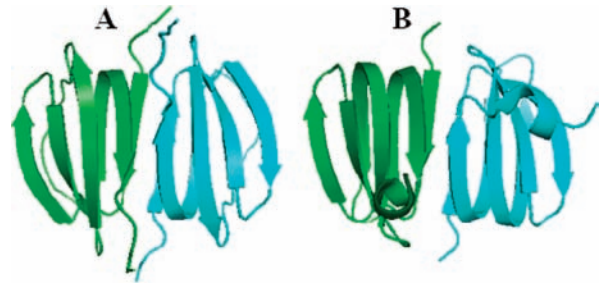

Figure 2. Ribbon drawings of (A) average structure of Dsy0195 determined by combined NMR, PRE, and DEER constraints and (B) crystal structure.

In conclusion, combining intermediate and long-rang distance constraints obtained from PRE and DEER techniques with conventional NMR intrachain NOESY-based distance constraints can provide a powerful hybrid approach for solving the 3D structure of protein homodimers. Our results indicate that only one or a few DEER distance constraints, combined with a small number of PRE constraints, in the complete absence of seeding with any experimentally determined interchain NOEs obtained by conventional ^{13}C-edited/^{12}C-filtered NOESY experiments, can be sufficient for CYANA to build a network of interchain NOE constraints that can be used to accurately define both the homodimer interface and the global homodimer structure. The use of DEER distances as a source of supplemental constraints, as described here, has virtually no upper molecular weight limit, and utilization of the PRE constraints is limited only by the ability to make accurate assignments of the protein amide proton and nitrogen chemical shifts.

Acknowledgment. This project was supported by the National Institute of General Medical Sciences, grant no. U54-GM074958; National Science Foundation, grant no. NSF (MRI-0722403); Bruker Biospin, Miami University, and Ohio Board of Reagents. A portion of the NMR experiments were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

Supporting Information Available: Analytical gel filtration with static light scattering detection for Dsy0195; cw EPR spectra of MTSL binding to S36C and S52C mutants; 2D ^{15}N-HSQC spectra of S36C, S52C mutants, and wt protein; 2D ^{15}N-HSQC spectra of S52C with and without MTSL; electron transverse relaxation time measurement of S36C and S52C mutants; DEER experiment of S52C-MTSL; correlation plot of PRE distance constraints with crystal structure; NMR structure quality assessment table. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ikura, M.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 2433–2440.
(2) Folkers, P. J. M.; Folmer, R. H. A.; Konings, R. N. H.; Hilbers, C. W. *J. Am. Chem. Soc.* **1993**, *115*, 3798–3799.
(3) Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.
(4) Guntert, P. *Methods Mol. Biol.* **2004**, *278*, 353–378.
(5) Levy, E. D.; Pereira-Leal, J. B.; Chothia, C.; Teichmann, S. A. *PLOS Comput. Biol.* **2006**, *2*, 1395–1406.
(6) Wang, X.; Bansal, S.; Jiang, M.; Prestegard, J. H. *Protein Sci.* **2008**, *17*, 899–907.
(7) Beuck, C.; Szymczyna, B. R.; Kerkow, D. E.; Carmel, A. B.; Columbus, L.; Stanfield, R. L.; Williamson, J. R. *Structure* **2010**, *18*, 377–389.

(8) Raman, S.; Huang, Y. J.; Mao, B.; Rossi, P.; Aramini, J. M.; Liu, G.; Montelione, G. T.; Baker, D. *J. Am. Chem. Soc.* **2010**, *132*, 202–207.
(9) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621–16628.
(10) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355–5365.
(11) Rumpel, S.; Becker, S.; Zweckstetter, M. *J. Biomol. NMR* **2008**, *40*, 1–13.
(12) Banham, J. E.; Timmel, C. R.; Lea, S. M.; Abbott, R. J. M.; Jeschke, G. *Angew. Chem., Int. Ed.* **2006**, *45*, 1058–1061.
(13) Finiguerra, M. G.; Prudencio, M.; Ubbink, M.; Huber, M. *Magn. Reson. Chem.* **2008**, *46*, 1096–1101.
(14) Martin, R. E.; Pannier, M.; Diederich, F.; Gramlich, V.; Hubrich, M.; Spiess, H. W. *Angew. Chem., Int. Ed.* **1998**, *37*, 2834–2837.
(15) Pannier, M.; Veit, S.; Godt, A.; Jeschke, G.; Spiess, H. W. *J. Magn. Reson.* **2000**, *142*, 331–340.
(16) van Ooij, C.; Eichenberger, P.; Losick, R. *J. Bacteriol.* **2004**, *186*, 4441–4448.
(17) Vorobiev, S. M.; Chen, Y.; Seetharaman, J.; Janjua, H.; Xiao, R.; Ciccosanti, C.; Wang, H.; Everett, J. K.; Nair, R.; Acton, T. B.; Rost, B.; Andre, I.; Rossi, P.; Kennedy, M.; Montelione, G. T.; Hunt, J. F.; Tong, L., Northeast Structural Genomics Consortium (NESG), 2009.
(18) Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H. *Appl. Magn. Reson.* **2006**, *30*, 473–498.
(19) Hilger, D.; Polyhach, Y.; Padan, E.; Jung, H.; Jeschke, G. *Biophys. J.* **2007**, *93*, 3675–3683.
(20) Borbat, P. P.; Mchaourab, H. S.; Freed, J. H. *J. Am. Chem. Soc.* **2002**, *124*, 5304–5314.
(21) Sale, K.; Song, L. K.; Liu, Y. S.; Perozo, E.; Fajer, P. *J. Am. Chem. Soc.* **2005**, *127*, 9334–9335.
(22) Banham, J. E.; Baker, C. M.; Ceola, S.; Day, I. J.; Grant, G. H.; Groenen, E. J. J.; Rodgers, C. T.; Jeschke, G.; Timmel, C. R. *J. Magn. Reson.* **2008**, *191*, 202–218.
(23) Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins* **2007**, *66*, 778–795.

JA105080H