Published in final edited form as:

J Am Chem Soc. 2010 September 1; 132(34): 11910–11913. doi:10.1021/ja105080h.

Combining NMR and EPR Methods For Homo-Dimer Protein Structure Determination

Yunhuang Yang^{§,#}, Theresa A. Ramelot^{§,#}, Robert M. McCarrick[§], Shuisong Ni[§], Erik A. Feldmann^{§,#}, John R. Cort[§], Huang Wang^{#,+}, Colleen Ciccocanti^{#,+}, 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

[%]Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Abstract

There is a general need to develop more powerful and more robust methods for structural characterization of homo-dimers, homo-oligomers and multi-protein complexes using solution-state NMR methods. In recent years, there has been increasing emphasis on integrating distinct and complementary methodologies for structure determination of multi-protein complexes. One approach not yet widely used is to obtain intermediate and long-range distance constraints from paramagnetic relaxation enhancements (PRE) and EPR-based techniques such as Double 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 NMR methods for structure determination of Dsy0195, a homo-dimer (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

kennedm4@muohio.edu .

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 ¹⁵N-HSQC spectra of S36C, S52C mutants and *wt* protein; 2D ¹⁵N-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>.

structure determination program CYANA to build a network of inter-chain NOE constraints that can be used to accurately define both the homo-dimer interface and global homo-dimer 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 only limited by the ability to make accurate assignments of the protein amide proton and nitrogen chemical shifts.

Homo-dimer and multi-protein complex structure determination using solution NMR methods usually relies on inter-chain 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 inter-chain NOEs. Poor NMR data can result from one of several homo-dimer characteristics including 1) tight dimer association resulting in unfavorable chain exchange kinetics and a low population of mixed ¹³C/¹²C-labeled dimers, 2) weak dimer association preventing measurement of inter-chain NOEs, or 3) a homo-dimer interface whose nature prevents measurement or interpretation of inter-chain NOEs. Inability to measure and identify interfacial NOEs in isotope-filtered experiments can make it difficult or impossible to assign inter-chain NOEs in unfiltered, ¹³C and ¹⁵N-edited NOESY data, and this can prevent successful homo-dimer structure determination using conventional NMR methods alone. Programs such as CYANA⁴, which are designed to automatically assign NOESY cross-peaks, are sometimes capable, when seeded with just a few unambiguous inter-chain NOEs from isotope-filtered experiments, of assigning a sufficient network of inter-chain NOEs from unfiltered ¹³C and ¹⁵N-edited NOESY data to accurately determine a homo-dimer structure. However, CYANA and similar programs generally fail to make productive inter-chain 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 only exacerbated as the protein systems become larger and more complex. In general, more powerful and robust methods are needed for structural characterization of homo-dimers, homo-oligomers and protein complexes using solution-state NMR methods. This challenge is of great biological significance; for example, more than 80% of *E. coli* proteins are predicted to exist as homo-dimers 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 multi-protein complexes is also growing concern for *Structural Genomics* efforts. In particular, multi-protein 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 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 multi-protein complexes. For example, traditional NMR data have been combined with residual dipolar couplings (RDCs),⁶ chemical shift (CS)-Rosetta^{7, 8} and small angle scattering (SAXS)⁹ data in order to solve homo-dimeric protein structures.

One approach not yet widely used for multi-protein 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 EPR-based techniques such as Double Electron Electron Resonance (DEER),¹²⁻¹⁵ which, when used together, can provide additional distance constraints spanning to 10-70Å. These

supplemental distance measurements can be used for structure validation, filtering candidate oligomer or multi-protein structures, or as additional constraints for *de novo* structure calculations. In this letter, we describe integration of PRE and DEER data with conventional solution-state NMR methods for structure determination of the homo-dimer Dsy0195 protein (NESG target name Dhr8C) from *Desulfitobacterium hafniense*.

Dsy0195 belongs to 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 homo-dimer of 62 residues per monomer. 1D ¹⁵N T_1 / T_2 measurements ($\tau_c = 12.2 \pm 0.7$ 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 inter-chain ¹³C-edited/¹²C-filtered NOEs, which were insufficient for homo-dimer 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. MTSL [(1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate] 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.

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 ¹⁵N-HSQC spectra of ¹⁵N-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). Two-dimensional ¹⁵N-HSQC spectra of the mutants covalently bound to MTSL demonstrated that resonances were either broadened beyond detection for residues nearby 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 μ s and 1.93 μ s, respectively (Figure S5, Supporting Information), which were long enough to enable collection of high signal-to-noise DEER data at Q-band (34GHz). 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 1(A) and the corresponding dipolar spectrum is shown in Figure 1(B). The best fit of the distance distribution between nitroxide spin-labels in the S36C mutant homo-dimer had a peak at 34Å after Tikhonov regularization (Figure 1(C)), consistent with the distance measured from crystal structure (33.5Å between C β of S36 in the dimer A and B chains). While 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, 20} 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.^{19, 20} If the distance 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.²¹

Following the PRE analysis procedure proposed by Rumpel *et al.*,¹¹ 67 lower-bound and 35 upper-bound inter-chain constraints were derived from a 1:1 mixture of ¹⁵N-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 than backbone atoms, were neglected. After culling PREs that likely contained errors for reasons just discussed, a final set of 29 lower-bound and 8 upper-bound inter-chain 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 cycles.⁴ Here, we tested the power of seeding CYANA calculations with inter-chain constraints derived from both DEER and PRE experiments. Three steps were taken for homo-dimer structure calculations. First, the monomer structure of Dsy0195 was calculated with conventional NMR methods. All NOEs were automatically assigned by CYANA. Some erroneous intra-chain assignments of inter-chain NOEs caused bad quality of the monomeric structure with a high MolProbity Clashescore. The Protein Structure Validation Suite (PSVS ver 1.4) program,²² was used to help identify incorrectly assigned ambiguous NOEs, which were removed, and the final well-folded monomeric structure was determined. Second, final intra-chain NOEs, hydrogen bond restraints, and dihedral angle constraints obtained from the monomeric structure calculations were used to seed the homo-dimer structure calculation. CYANA calculations in the absence of seeding with inter-chain constraints resulted in two separate, well-folded monomer sub-units with no dimer interface. Third, different numbers of intermediate and long-range distance inter-chain 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 anywhere from 118-130 inter-chain NOE constraints and to reproduce the correct dimer interface. Resulting structures were checked to determine numbers and accuracy of final inter-chain NOE assignments by comparing the average structure to the crystal structure. In general, seeding with more long-range-distance inter-chain constraints either from DEER or PRE increased the number of inter-chain NOEs correctly assigned by CYANA and improved the backbone 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 inter-chain NOEs, which were combined with intra-chain 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 ambiguous assigned inter-chain NOEs were manually removed for reasons described above. For the final calculation cycle, 125 inter-chain NOEs remained. The final ensemble of 20

structures with lowest energy has deposited in the PDB (2KYI). The quality of final structure was assessed using PSVS ver 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 2A and 2B, respectively. The backbone RMSD of average NMR-PRE-DEER structure against crystal structure was 1.13 Å, which is significantly better than the structure obtained using NMR data alone (1.51 Å), PDB (2KS0).

In conclusion, combining intermediate and long-range distance constraints obtained from PRE and DEER techniques with conventional NMR intra-chain NOESY-based distance constraints can provide a powerful hybrid approach for solving the three-dimensional structure of protein homo-dimers. 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 inter-chain NOEs obtained by conventional ¹³C-edited/¹²C-filtered NOESY experiments, can be sufficient for CYANA to build a network of inter-chain NOE constraints that can be used to accurately define both the homo-dimer interface and global homo-dimer 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 only limited by the ability to make accurate assignments of the protein amide proton and nitrogen chemical shifts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported by the National Institute of General Medical Sciences, Grant Number: U54-GM074958; National Science Foundation, Grant number CHE-0645709; Bruker Biospin, Miami University and Ohio Board of Regents. 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.

References

1. Ikura M, Bax A. Journal of the American Chemical Society 1992;114:2433–2440.
2. Folkers PJM, Folmer RHA, Konings RNH, Hilbers CW. Journal of the American Chemical Society 1993;115:3798–3799.
3. Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE. Journal of the American Chemical Society 1997;119:6711–6721.
4. Guntert P. Methods Mol Biol 2004;278:353–78. [PubMed: 15318003]
5. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. Plos Computational Biology 2006;2:1395–1406.
6. Wang X, Bansal S, Jiang M, Prestegard JH. Protein Science 2008;17:899–907. [PubMed: 18436958]
7. Beuck C, Szymczyzna BR, Kerkow DE, Carmel AB, Columbus L, Stanfield RL, Williamson JR. Structure 2010;18:377–89. [PubMed: 20223220]
8. Raman S, Huang YJ, Mao B, Rossi P, Aramini JM, Liu G, Montelione GT, Baker D. J Am Chem Soc 2010;132:202–7. [PubMed: 20000319]
9. Grishaev A, Wu J, Trewella J, Bax A. Journal of the American Chemical Society 2005;127:16621–16628. [PubMed: 16305251]
10. Battiste JL, Wagner G. Biochemistry 2000;39:5355–5365. [PubMed: 10820006]
11. Rumpel S, Becker S, Zweckstetter M. Journal of Biomolecular Nmr 2008;40:1–13. [PubMed: 18026911]

12. Banham JE, Timmel CR, Lea SM, Abbott RJM, Jeschke G. *Angewandte Chemie-International Edition* 2006;45:1058–1061.
13. Finiguerra MG, Prudencio M, Ubbink M, Huber M. *Magnetic Resonance in Chemistry* 2008;46:1096–1101. [PubMed: 18932181]
14. Martin RE, Pannier M, Diederich F, Gramlich V, Hubrich M, Spiess HW. *Angewandte Chemie-International Edition* 1998;37:2834–2837.
15. Pannier M, Veit S, Godt A, Jeschke G, Spiess HW. *Journal of Magnetic Resonance* 2000;142:331–340. [PubMed: 10648151]
16. van Ooij C, Eichenberger P, Losick R. *J Bacteriol* 2004;186:4441–8. [PubMed: 15231775]
17. Vorobiev SM, Chen Y, Seetharaman J, Janjua H, Xiao R, Ciccocanti C, Wang H, Everett JK, Nair R, Acton TB, Rost B, Andre I, Rossi P, Kennedy M, Montelione GT, Hunt JF, Tong L, NESG. 2009
18. Jeschke G, Chechik V, Ionita P, Godt A, Zimmermann H, Banham J, Timmel CR, Hilger D, Jung H. *Applied Magnetic Resonance* 2006;30:473–498.
19. Borbat PP, Mchaourab HS, Freed JH. *Journal of the American Chemical Society* 2002;124:5304–5314. [PubMed: 11996571]
20. Sale K, Song LK, Liu YS, Perozo E, Fajer P. *Journal of the American Chemical Society* 2005;127:9334–9335. [PubMed: 15984837]
21. Banham JE, Baker CM, Ceola S, Day IJ, Grant GH, Groenen EJJ, Rodgers CT, Jeschke G, Timmel CR. *Journal of Magnetic Resonance* 2008;191:202–218. [PubMed: 18280189]
22. Bhattacharya A, Tejero R, Montelione GT. *Proteins* 2007;66:778–95. [PubMed: 17186527]

Figure 1.

(A) Time-domain DEER signal of 0.25mM S36C-MTSL at 20mM ammonium acetate, 200mM NaCl and 5 mM CaCl₂ (pH 4.5) diluted to a final concentration of 30% (% w/w) glycerol; The black line in (A) is 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. and (C) best fit of distance distribution obtained from Tikhonov regularization. The DEER experiment was carried on a Bruker ELEXSYS E580 pulsed EPR instrument at 80K.

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

Table 1

The correlation of numbers of final inter-chain NOES assigned and the backbone RMSD to the crystal structure of the average structure calculated using CYANA 2.1 seeded with different inter-chains constraints.

Constraint seeds	No. of Assign. ^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 inter-chain 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 inter-chain constraints from PRE data used for structure calculation.