Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

NMR structure of hypothetical protein MG354 from Mycoplasma genitalium

Permalink

https://escholarship.org/uc/item/3rp262rc

Authors

Pelton, Jeffrey G. Shi, Jianxia Yokotoa, Hisao <u>et al.</u>

Publication Date

2005-04-12

Peer reviewed

Prot-00430-2004-R1

NMR structure of hypothetical protein

MG354 from Mycoplasma genitalium

Jeffrey G. Pelton¹, Jianxia Shi², Hisao Yokota¹, Rosalind Kim¹, and David E. Wemmer^{1,3*}

¹Berkeley Structural Genomics Center, Physical Biosciences Division of the Lawrence Berkeley National Laboratory, Berkeley, CA. 94720.

²Amgen Inc., 1120 Veterans Boulevard, South San Francisco, CA 94080

³Department of Chemistry, University of California, Berkeley, CA. 94720.

Grant sponsor: National Institutes of Health; Grant number: GM62412

*Correspondence to David E. Wemmer, Department of Chemistry, University of California, Berkeley, CA. 94720. E-mail: DEWemmer@lbl.gov.

Introduction. *Mycoplasma genitalium* (Mg) and *M. pneumoniae* (Mp) are human pathogens with two of the smallest genomes sequenced to date (~ 480 and 680 genes, respectively). The Berkeley Structural Genomics Center is determining representative structures for gene products in these organisms, helping to understand the set of protein folds needed to sustain this minimal organism. The protein coded by gene MG354 (gi3844938) from *M. genitalium* has a relatively unique sequence, related only to MPN530 from *M. pneumoniae* (68% identity, coverage 99%) and MGA_0870 from the avian pathogen *M. gallisepticum* (23% identity, coverage 94%), has no homologue with a determined structure, and no functional annotations.

Results. The overall structure of MG354 (137 residues) was determined with a backbone root-mean-square-deviation (rmsd) of 0.6 Å based on 850 structurally significant NOE restraints, 69 phi dihedral restraints, and 86 hydrogen bond restraints (Table 1). A total of 40 N-H^N residual dipolar couplings were used to validate the model (see materials and methods). The structure consists of seven helices that fold into a single domain (Figure 1). The most striking feature is the long central helix (H6), surrounded by the remaining six helices to produce a hydrophobic core. Within the CATH protein fold database¹, MG354 falls into the mainly alpha class and most closely resembles proteins with an orthogonal-bundle architecture.

A search for similar folds was conducted with CE², Dali³, and VAST⁴, recently shown to be among the best at matching protein structures.⁵ Three energy-minimized coordinate sets from the 26-member family of NMR structures were used as templates. The three servers returned different sets of hits, each of which showed some similarity to MG354. For example, CE matched four contiguous helices (86 of 252 residues) in 1F5Q (chain B, γ -herpesvirus cyclin) to the last five helices in MG354 (H3 to H7) with an rmsd of 4.9 Å and a Z-score of 4.6. One helix in 1F5Q was paired with both H5 and H6. Combined, the match covered 62% and 8%, respectively, of the residues in domains 1 and 2 of the cyclin, as defined in the CATH database¹. Similarly, Dali matched four contiguous helices (75 of 508 residues) in 1GW5 (chain A, clathrin adaptor protein) to H1, H4, H6, and H7 of MG354 with an rmsd of 3.2 Å, and a Z-score was 3.9. However, the first helix of 1GW5 was oriented perpendicular to H1. Compared to these, VAST matched 5 helices (50 of 411 residues) in 1UX5 (formin homology-2 domain) to H1, H2, part of H3, H4, and H6 with an rmsd of 2.4 Å. In this case the topology was somewhat different, since 1UX5 has an extra 6-residue helix in between those that matched H2 and H3. Thus, all three methods were able to identify protein fragments that had some similarity to MG354, but none of the proteins contained all seven essential helices, and the matches were not strong enough to suggest similar functions.

In cases of low sequence and structural similarity, clues to function can sometimes be obtained by matching local structural patterns with those of known ligand binding sites and enzyme active sites. PROCAT⁶ classified MG354 as an acetylglucosamidase (E.C. 3.2.1.96) based on the locations of Glu-23 and Asp-26 and RIGOR⁷ classified MG354 as a lipase (E.C. 3.1.1.3) based on the positions of Glu-23, Asp-26, and Asp-41. The sidechains for all three residues are clustered together on the surface. Glu-23 and Asp-26 are at the edge of a 21 Å³ solvent accessible cleft identified by the CASTp server⁸. The cleft is defined by residues Gln-22, Leu-25, Asp-26, Leu-134, Asn-135, and Asn-137. On the other hand, PINTS⁹, which uses a statistical score to filter out insignificant hits, returned no functional classification. The hydrolase and lipase predictions are intriguing, but the lack of consensus among the programs leaves the function of MG354 in doubt. Thus, because of its limited sequence and structural homology, the structure of MG354 represents a challenge for the further development of programs to predict function based on local structure comparisons.

Methods. Protein expression. The gene for MG354 was cloned into the expression plasmid pSKB3 (a gift from Dr. Stephen Burley) and transformed into *E. coli* strain BL21(DE3)pSJS1244 as described.¹⁰ Cells were harvested by centrifugation and lysed by sonication in 50 mM Tris-HCl (pH 8.0) and 300 mM NaCl. The supernatant was applied to a Ni-NTA column (Qiagen; Valencia, CA.), washed with the same buffer containing 15 mM imidazole, and eluted with a gradient up to 1 M imidazole. The His tag was cleaved with the Tobacco Etch Virus (TEV) protease, and the protein was reapplied to the Ni-NTA column. The flow through was dialyzed against 20 mM potassium phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, and 1 mM DTT and concentrated for NMR experiments. Final protein concentrations ranged from 0.75 mM to 1.2 mM.

NMR spectroscopy. NMR spectra were recorded at 25 °C on a Bruker DRX 500 MHz spectrometer equipped with a triple-resonance cryoprobe and a Bruker DRX 600 MHz spectrometer equipped with a traditional triple-resonance probe. One 3D ¹³C-edited NOESY-HSQC spectrum was recorded at 800 MHz on a Varian INOVA spectrometer. Backbone ¹HN, ¹⁵N, and ¹³C resonances were assigned using HNCO, HN(CA)CO, HNCA, CBCA(CO)NH, HNCACB, and ¹⁵N-separated NOESY-HSQC spectra recorded

as described.¹¹⁻¹³ 1 H α and sidechain 1 H and 13 C resonances were assigned using HCANH, HBHA(CO)NH, H(C)CONH, (H)C(CO)NH, and HCCH-TOCSY experiments as described.^{11,12} For a review of the experiments used, see Cavanagh *et al.*¹⁴ NMR spectra were processed using NMRPipe¹⁵ and analyzed using NMRView.¹⁶ A total of 92% of the resonances were assigned, including all amide protons except Lys-115. **Structure calculations.** NOEs identified in 3D ¹⁵N-edited and ¹³C-edited NOESY-HSQC spectra and 4D ¹⁵N/¹³C-edited and ¹³C/¹³C edited HMQC-NOESY-HSOC spectra were classified as strong (1.8 - 2.9 Å), medium (1.8 - 3.3 Å) or weak (1.8 - 5.0 Å) as described.^{11,12} Phi angle restraints were derived from ${}^{3}J_{HNH\alpha}$ coupling constants obtained from an HNHA spectrum.¹⁷ Angles were constrained to $-60 + -30^{\circ}$ for ${}^{3}J_{HNH\alpha}$ of less than 5 Hz and $-170 \pm 70^{\circ}$ for ${}^{3}J_{HNH\alpha}$ of greater than 8 Hz. DYANA¹⁸ was used for structure calculations. Manually assigned NOEs and torsion restraints were used to initially determine the fold. Subsequently, in-house software was used to assign NOEs by matching peak chemical shifts and filtering with distances in representative structures. Once the backbone rmsd fell below 1.5 Å, hydrogen bond restraints were added for those amide protons with slow exchange (protection factors greater than 80) and short HN-O distances. Structures were viewed and analyzed with MOLMOL.¹⁹

To validate the structural model, a total of 40 N-H^N residual dipolar couplings were obtained from analysis of IPAP experiments²⁰ recorded on isotropic and partially aligned (Pf1 phage) samples. At least two rdc values were measured for each helix. Using CNS^{21} , the best 20 of 200 structures had no distance or dihedral restraint violations greater than 0.5 Å and 5 deg, respectively, and the rmsd for N-H^N rdcs fell from 4.6 +/-0.4 Hz (Qrdc²² = 0.48 +/- 0.04) to 0.6 +/- 0.2 Hz (Qrdc = 0.06 +/- 0.01) with the additional restraints. The maximal residual coupling was 21 Hz. The mean structures calculated with and without rdcs superimposed with a coordinate rmsd of 0.4 Å, indicating good agreement between the NOE, dihedral, hydrogen bond, and rdc data.

Acknowledgements. We thank Dr. Corey Liu for recording a 3D ¹³C-edited NOESY-HSQC spectrum at 800 MHz at the Stanford Magnetic Resonance Facility and Andrew Marshall of the J. Puglisi laboratory (Stanford University) for providing Pf1 phage. We are also grateful to Barbara Gold for cloning, Marlene Hernandez and Bruno Martinez for expression studies, and John-Marc Chandonia for bioinformatics assistance. Coordinates and structure restraints have been deposited in the PDB under accession number 1TM9 and the chemical shift assignments have been deposited in the BMRB under accession number 6244.

Figures

Figure 1. Structure of MG354. A) Stereoview of the 26 final conformers. Residues 5-137 are shown. B) Stereoview of a ribbon diagram of a representative structure generated with the program Molscript.²³ Although not included in the calculations, the structures suggest that a disulfide bond forms between Cys-18 and Cys-99.
 Table 1. Experimental NMR data and structural statistics for MG354.

Restraints	
NOE upper distance limits	
Total	850
Intra-residue	278
Sequential $(i-j = 1)$	238
Medium Range ($1 < i-j < 5$)	213
Long range (spanning 5 or more	
residues	121
Hydrogen bond restraints (two per H-bond)	86
Phi torsion angle restraints	69
Distance restraint violations ^a	
Residual DYANA target function	1.4 +/- 0.3
Number of violations > 0.2 (Å)	3 +/- 1
Average maximum (Å)	0.3 +/- 0.1
Torsion angle restraint violations	
Number $> 2.5^{\circ}$	0
Average maximum (degrees)	0.02 +/- 0.02
Van der Waals violations	
Number > 0.3 Å	1 +/- 1
Average maximum violation (Å)	0.23 +/- 0.08
Coordinate precision (Å)	
Backbone atoms in helices	0.6 +/- 0.1
(residues 5-15,22-32,46-54,	
57-74,86-90,94-105,119-135)	
Backbone (residues 5-135)	1.1 +/- 0.2
Heavy atoms (residues 5-135)	1.7 +/- 0.2
Procheck statistics ^b	
Residues in most favored regions (%)	64.5
Residues in additional allowed regions (%)	28.7
Residues in generously allowed regions (%)	5.8
Residues in disallowed regions (%)	1.1

^aThere were no distance restraint violations greater than 0.2 Å that occurred in 8 or more structures. The maximum violation considering all structures and restraints was 0.6 Å.

^bStructures were not energy minimized.

References

- Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH
 A hierarchic classification of protein domain structure. Structure 1997;5:1093-1108.
- Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739-747.
- Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123-138.
- Gibrat JF, Madej T, Bryant SH. Surprising similarities in structure comparison. Curr Opin Struct Biol 1996;6:377-385.
- Novotny M, Madsen D, Kleywegt GJ. Evaluation of protein fold comparison servers. Proteins 2004;54:260-270.
- Wallace AC, Laskowski RA, Thornton JM. Derivation of 3D coordinate templates for searching structural databases: Application to the Ser-His-Asp catalytic triads of the serine proteinases and lipases. Protein Sci 1996;5:1001-1013.
- Kleywegt GJ. Recognition of spatial motifs in protein structures. J Mol Biol 1999;285(4):1887-1897.
- Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. Protein Sci 1998;7:1884-1897.
- 9. Stark A, Russel RB. Annotation in three dimensions. PINTS: Patterns in nonhomologous tertiary structures. Nucleic Acids Res 2003;31(13):3341-3344.

- 10. Kim R, Sandler SJ, Goldman S, Yokota H, Clark AJ, Kim S-H. Overexpression of archaeal proteins in *Escherichia coli*. Biotechnology Let 1998;20:207-210.
- Pelton JG, Kustu S, Wemmer DE. Solution Structure of the DNA-binding domain of NtrC with three alanine substitutions. J Mol Biol 1999;292:1095-1110.
- 12. Rubin SM, Pelton JG, Yokota H, Kim R, Wemmer DE. Solution Structure of a putative ribosome binding protein from *Mycoplasma pneumoniae* and comparison to a distant homolog. J Struct Funct Genomics 2004;4(4):235-243.
- Shi J, Pelton JG, Cho HS, Wemmer DE. Protein signal assignments using specific labeling and cell-free synthesis. J Biomol NMR 2004;28:235-247.
- Cavanagh J, Fairbrother WJ, Palmer AGI, Skelton NJ. Protein NMR
 Spectroscopy: Principles and Practice. New York: Academic Press; 1995. 587 p.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277-293.
- Johnson BA, Blevins RA. NMRView: a computer program for the visualization and analysis of NMR data. J Biomol NMR 1994;4:603-614.
- Kuboniwa H, S. G, Delaglio F, Bax A. Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. J Biomol NMR 1994;4:871-878.
- Guntert P, Mumenthaler C, Wuethrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283-298.

- 19. Koradi R, Billeter M, Wuethrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14(1):51-55.
- 20. Ottiger M, Delaglio F, Bax A. Measurement of J and Dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson 1998;131:373-378.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR System. Acta Cryst 1998;D54:905-921.
- 22. Ottiger M, Bax A. Bicelle-based liquid crystals for NMR measurement of dipolar couplings at acidic and basic pH values. J Biomol NMR 1999;13:187-191.
- Kraulis PJ. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Applied Cryst 1991;24:946-950.

