PROTEINS: Structure, Function, and Bioinformatics 62:288–291 (2006)

STRUCTURE NOTE

NMR Structure of Protein yqbG from *Bacillus subtilis* Reveals a Novel α-Helical Protein Fold

Gaohua Liu,[1,4] Yang Shen,[1,4] Rong Xiao,[2,4] Tom Acton,[2,4] Li Chung Ma,[2,4] Andrzej Joachimiak,[3,5] Gaetano T. Montelione,[2,4] and Thomas Szyperski[1,4]*

[1]*Department of Chemistry, University at Buffalo, the State University of New York, Buffalo, New York*
[2]*Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey*
[3]*Structural Biology Center, Argonne National Laboratory, Argonne, Illinois*
[4]*Northeast Structural Genomics Consortium*
[5]*Midwest Center for Structural Genomics*

Introduction. 131-residue protein yqbG (gi|16079665, SwissProt/TrEMBL ID YQBG-BACSU, access number P45923) encoded by gene YQBG from *Bacillus subtilis* has no significant sequence similarity with any protein with known three-dimensional structure and thus was selected as a target (ID APC1551) by the Midwest Center of Structural Genomics (MCSG; http://www.mcsg.anl.gov) for X-ray crystal structure determination. Because protein yqbG is highly soluble and monodispersed while no diffraction quality crystals could be obtained, it was transferred to NMR division of the Northeast Structural Genomics consortium (NESG; http://www.nesg.org) for structure determination (NESG target ID SR215). Notably, yqbG belongs to temporarily created Protein Family (Pfam[1]) PB071099, which contains only protein lin1273 (45% sequence identity with yqbG) as a second member. An iterative PSI-Blast[2] search of the nonredundant Genbank reveals that yqbG exhibits significant sequence identity to a third protein, yqbG (gi|52785100) from *Bacillus licheniformis* (55% identity). All three proteins belong to PIRSF family SF007500.[3] So far, no functional annotations are available. Here we report the high-quality NMR solution structure of protein yqbG,, which exhibits a novel α-helical fold.

Materials and Methods. Uniformly (*U*) ^{13}C,^{15}N-labeled yqbG was cloned, expressed, and purified by following standard protocols. Briefly, the full-length gene from *Bacillus subtilis* was cloned into a pET21 (Novagen) derivative, yielding the plasmid pSR215-21.3. The resulting construct contains eight nonnative residues at the C-terminus (LEHHHHHH) that facilitate protein purification. *Escherichia coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pSRF215-21.3 and cultured in MJ9 minimal medium containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources.[4] *U*-^{13}C,^{15}N yqbG was purified by using a two-step protocol consisting of Ni-NTA affinity (Qiagen) and gel filtration (HiLoad 26/60 Superdex 75; Amersham Biosciences) chromatography. The final yield of purified *U*-^{13}C,^{15}N yqbG (>98% homogenous by SDS-PAGE; 16.6 kDa by MALDI-TOF mass spectrometry) was about 13.5 mg/L. The final samples of *U*-^{13}C,^{15}N labeled yqbG were prepared at a concentration of about 0.9 mM in 95% H_2O/5% D_2O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, and 0.02% NaN_3 at pH 6.5.

All NMR spectra were recorded at 25°C. Five G-matrix Fourier transform[5] (GFT) NMR experiments[5–7] were performed for resonance assignment on a Varian INOVA 600 spectrometer equipped with a cryogenic probe (total measurement time: 104 h), and a simultaneous $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-resolved NOESY spectrum[7] was acquired on a Varian INOVA 750 spectrometer equipped with a conventional probe (measurement time: 23 h) to determine ^1H–^1H distance constraints. Spectra were processed and analyzed with the programs PROSA[8] and XEASY[9], respectively. Sequence specific backbone (H^N, H^α, N, C^α) and H^β/C^β resonance assignments were obtained by using (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}/\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN[6] and (4,3)D $\underline{H}^{\alpha\beta}\underline{C}^{\alpha\beta}$(CO)NHN. Side-chain spin system identification was accomplished by using aliphatic and aromatic (4,3)D $\underline{HC}$CH.[7] Assignments were obtained for 100% of the assignable backbone (excluding the N-terminal NH_3^+, the Pro ^{15}N and the $^{13}C'$ shifts) and $^{13}C^\beta$, and for 99% of the side-chain chemical shifts (excluding Lys NH_3^+, Arg NH_2, OH, side-chain $^{13}C'$, and

Grant sponsor: National Institutes of Health; Grant number: P50 GM62413-01; Grant sponsor: National Science Foundation; Grant number: MCB 0416899.

*Correspondence to: Thomas Szyperski, Department of Chemistry, University at Buffalo, the State University at New York, Buffalo, NY 14260. E-mail: szypersk@chem.buffalo.edu

Received 1 June 2005; Accepted 10 June 2005

Published online 9 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20666

TABLE I. Statistics of yqbG NMR Structure

Completeness of stereospecific assignments[a] [%]	
$^{\alpha}CH_2$ of Gly	25 (1/4)
$^{\beta}CH_2$	55 (41/75)
Val and Leu methyl groups	80 (20/25)
Conformationally restricting distance constraints (i and j are residue numbers)	
Intraresidue [i = j]	595
Sequential [(i − j) = 1]	945
Medium range [1 < (i − j) < 5]	906
Long range [(i − j) ≥ 5]	742
Total	3188
Dihedral angle constraints	
ϕ	53
ψ	53
Average number of constraints per residue	23.7
Average number of long-range distance constraints per residue	5.3
CYANA target function [Å^2]	0.78 ± 0.16
Average number of distance constraints violations per CYANA conformer	
0.2–0.5 Å	0
>0.5 Å	0
Average number of dihedral-angle constraint violations per CYANA conformer	
>5°	0
Average RMSD to the mean CYANA coordinates (Å)	
α-Helices I–V[b], backbone heavy atoms N, C$^{\alpha}$, C′	0.36 ± 0.07
α-Helices I–V, all heavy atoms	0.77 ± 0.08
Residues 3–73, 99–113, backbone heavy atoms N, C$^{\alpha}$, C′	0.49 ± 0.12
Residues 3–73, 99–113, all heavy atoms	0.87 ± 0.17
Heavy atoms of molecular core including best-defined side chains[c]	0.41 ± 0.08
PROCHECK G-factors[d] (ϕ and ψ/all dihedral angles)	−0.42/−0.60
MOLPROBITY clash score[e]	14.2 ± 4.3
AutoQF R/P/DP scores (%)[f]	97/97/84
Ramachandran plot summary for residues 3–73, 99–113 (%)	
Most favored regions	80
Additionally allowed regions	18
Generously allowed regions	2
Disallowed regions	0

[a]Relative to pairs with nondegenerate chemical shifts.
[b]α-Helices I–V comprise residues 6–12, 16–20, 23–41, 55–72, 105–111. Note that residues 114–131, including α-helix VI (residues 125–129), are "globally" disordered (Fig. 1).
[c]Backbone heavy atoms N, C$^{\alpha}$, and C′ of helices I–V and side-chain heavy atoms of residues 4, 6, 9, 11, 13–14, 18, 22, 25–26, 30–31, 33–35, 37–39, 41, 53–54, 57, 59–62, 64, 69–71, 102, 104, 106–108, 112.
[d]Scores defined in Ref.[18].
[e]Scores defined in Ref.[19].
[f]Scores defined in Ref.[20].

aromatic quaternary ^{13}C shifts; Table I). Stereospecific assignments were obtained for 55% of the β-methylene groups exhibiting nondegenerated proton chemical shifts and for 80% of the Val and Leu isopropyl moieties (Table I). Chemical shifts were deposited in the BioMagResBank (accession code: 6366). Upper distance limit constraints for structure calculations were obtained from NOESY (Table I). In addition, backbone dihedral angle constraints were derived from chemical shifts as described[10] for residues located in globally well-defined helices I–V (Table I). The programs CYANA[11,12] and AUTOSTRUCTURE[13] were used in parallel to automatically assign long-range NOEs. Coinciding assignments ("consensus assignments") were retained and established the starting point for manual completion of iterative NOE assignment, peak picking, and structure calculation. The final structure calculations were performed by using version 2.0 of CYANA,[12] (see: http://www.las.jp/prod/cyana/eg).

Results and Discussion. The statistics of the yqbG structure determination (Table I) show that a high-quality NMR structure was obtained (Fig. 1). Protein yqbG (PDB ID: 1ZTS) contains five α-helices I–V [Fig. 1(b)] comprising residues 6–12, 16–20, 23–41, 55–72, and 105–111. These five helices are locally and globally well defined. In contrast, the polypeptide segment 74–99 and the C-terminal tail of residues 114–131, which comprise the locally well-defined short α-helix VI (residues 125–129), are disordered in solution (Fig. 1). We speculate that the presence of these long flexible segments prevents crystallization of yqbG. The CATH/GRATH[14] structure classification scheme assigns the fold of yqbG to the class "mainly α-helical" having the architecture of an "orthogonal" bundle.

No protein with meaningful structural similarity could be identified by using the programs DALI[15] or CE.[16] The highest z scores are at about the threshold of being significant (i.e., 3.0 and 4.1 for DALI and CE, respectively),

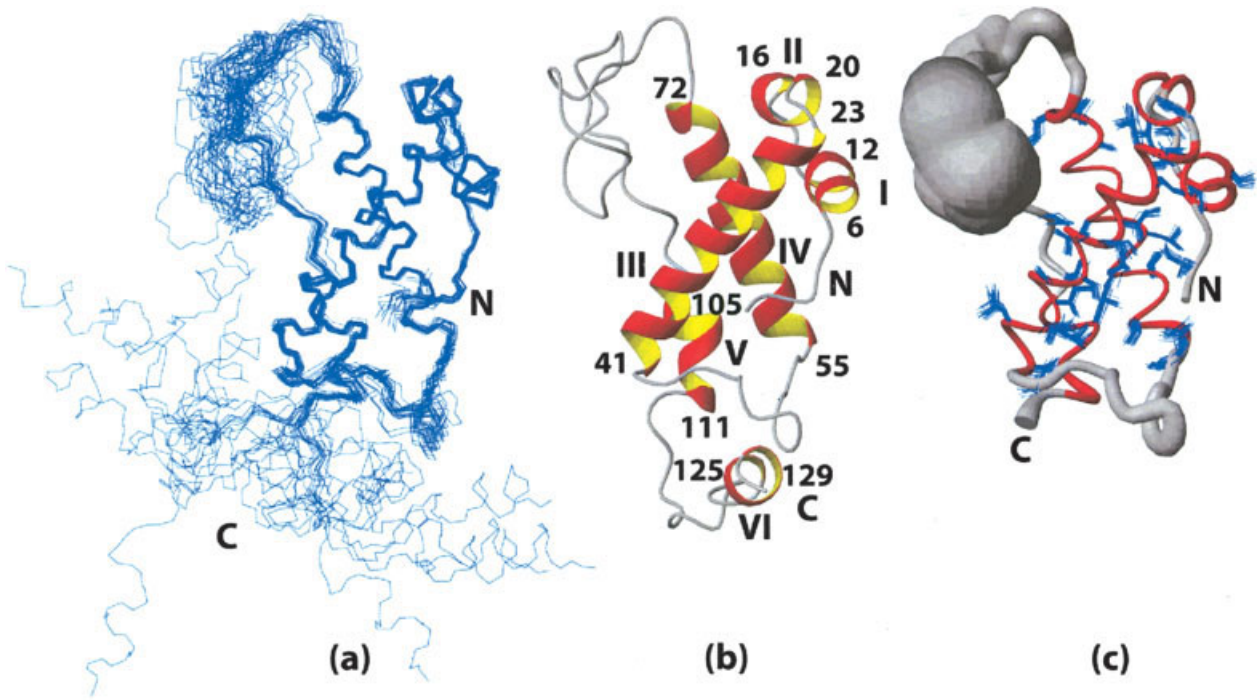

Fig. 1. NMR structure of protein yqbG revealing a novel α-helical fold. **a:** 20 CYANA conformers with the lowest residual target function chosen to represent the NMR solution structure of protein yqbG are shown after superposition for minimal root-mean-square deviation (RMSD) of the backbone heavy atoms N, C^{α}, and C′ of globally well-defined α-helices I–V (Table I). **b:** Ribbon drawing of the CYANA conformer with the lowest target function. The α-helices are shown in red and yellow, residue numbers (helix I: residues 6–12, II: 16–20, III: 23–41, IV: 55–72, V: 105–111, 125–129) are indicated, and the N- and C-termini of the polypeptide chain are labeled with N and C. **c:** Sausage presentation of backbone and best defined side-chains. A spline function was drawn through the C^{α} positions of residues 3–113, where the thickness of the cylindrical rod is proportional to the mean of the global displacements of the C^{α} atoms in the 20 CYANA conformers calculated after superposition as described in (a). The α-helices I–V are shown in red, and coil regions are displayed in gray. A superposition of the best defined side-chains with the lowest global displacement for the side-chain heavy atoms (for residue numbers, see Table I) is shown in blue to indicate precision of the determination of side-chain conformations.

but visual inspection reveals that this is due to the presence of the two long helices, helix III and IV [Fig. 1(b)], arranged in an antiparallel manner. Such arrangement of two helices is quite common and does, on its own, not point at significant structural similarity of two protein architectures. Furthermore, except for one protein, different candidates for structural similarity were identified by the two programs. Thus, we conclude that protein yqbG possesses a hitherto uncharacterized α-helical fold [Fig. 1(b)]. Notably, a PROSITE[17] motif search indicates that yqbG and its three sequence homologues share the Casein kinase II phosphorylation site Thr-Pro-Asp-Glu/Asp at the N-terminal end (residues 5–8 in yqbG).

Acknowledgements. The authors thank M. Ciaro and K. Cunningham for technical assistance. This work was supported by the National Institutes of Health and the National Science Foundation (T.S.).

REFERENCES

1. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2002;30:276–280.
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
3. Wu CH, Yeh LL, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Ledley RS, Kourtesis P, Ledley, RS, Suzek BE, Vinayaka CR, Zhang J, Barker WC. The Protein Information Resource. Nucleic Acids Res 2003;31:345–347.
4. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B. High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. J Biomol NMR 1996;7:131–141.
5. Kim S, Szyperski T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. J Am Chem Soc 2003;125:1385–1393.
6. Atreya HS, Szyperski T. G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment. Proc Natl Acad Sci USA 2004;101:9642–9647.
7. Shen Y, Atreya HS, Liu G, Szyperski T. G-matrix Fourier transform NOESY based protocol for high-quality protein structure determination. J Am Chem Soc 2005;127:9085–9099.
8. Güntert P, Dötsch V, Wider G, Wüthrich K. Processing of multi-dimensional NMR data with the new software PROSA. J Biomol NMR 1992;2:619–629.
9. Bartels C, Xia T, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 1995;6:1–10.
10. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shifts and sequence homology. J Biomol NMR 1999;13:289–302.
11. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new

software CANDID and the torsion angle dynamics algorithm CYANA. J Mol Biol 2002;319:209–227.
12. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program CYANA. J Mol Biol 1997;273:283–298.
13. Huang YJ, Swapna GVT, Rajan PK, Ke H, Xia B, Shukla K, Inouye M, Montelione GT. Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from *Escherichia coli*. J Mol Biol 2003;327:521–536.
14. Orengo CA, Michie AD, Jones S, Jones DT, Swindles, MB, Thornton, JM. CATH—a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.
15. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
16. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
17. Bairoch A, Bucher P, Hofmann K. The PROSITE database, its status in 1997. Nucleic Acids Res 1997;25:217–221.
18. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton, JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
19. Word JM, Bateman RC, Presley BK, Lovell SC, Richardson DC. Exploring steric constraints on protein mutations using MAGE/PROBE. Protein Sci 2000;9:2251–2259.
20. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 2005;127:1665–1674.