Journal of Biomolecular NMR, **23:** 159–160, 2002.
KLUWER/ESCOM

Letter to the Editor: Backbone ^{1}H, ^{15}N and ^{13}C resonance assignments of YqgF, an *Escherichia coli* protein of unknown structure and function

Dingjiang Liu, Prasanti Repaka, S. Shane Taremi & Daniel F. Wyss*
Schering-Plough Research Institute, Department of Structural Chemistry, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033, U.S.A.

Received 6 February 2002; Accepted 25 March 2002

Key words: antibiotics resistance, anti-microbial, NMR assignments, structural genomics, YqgF

Biological context

The spread of resistant bacterial strains has made conventional antibiotics less effective (Hand, 2000). Therefore, the discovery of novel classes of anti-microbial agents has become important. High throughput genomic DNA sequencing has delivered a large amount of genomic sequence data. Many complete microbial genome sequences have recently become available (http://www.ncbi.nlm.nih.gov/Entrez/, National Center for Biotechnology Information, Bethesda, MD). Such data represent a valuable basis for the identification of novel broad-spectrum antibiotics. Structural and functional characterization efforts have been focused on so-called 'hypothetical' proteins, since their biological functions have not been characterized experimentally and cannot readily be deduced from sequence analysis using current bioinformatic databases. These proteins are believed to be important for structural biology in the discovery of new protein folds (http://www.nature.com/nsb/structural_genomics/; see for example Burley, 2000) and in drug discovery as potential new drug targets for developing novel anti-infective agents (Freiberg et al., 2001). YqgF is an *Escherichia coli* gene annotated as a hypothetical protein and was found essential for *E. coli* growth (Freiberg et al., 2001). It contains 138 amino acids with a molecular weight of 15.2 kDa. Sequence analysis with BLAST (Altschul et al., 1997) revealed YqgF is conserved among many bacteria (Aravind et al., 2000; Freiberg et al., 2001); no homologous protein was identified from higher organisms. In addition, BLAST searches against current protein structures in the PDB (http://www.rcsb.org/pdb/) did not reveal any significant hits. However, it was recently suggested from sequence analysis that YqgF is likely to function as a Holliday junction resolvase (Aravind et al., 2000). To assist in revealing the biological function of YqgF and validate YqgF as a potential drug target, solution NMR techniques have been used to determine its solution structure. Here, we report the sequence-specific backbone resonance assignments of YqgF.

*To whom correspondence should be addressed. E-mail: daniel.wyss@spcorp.com

Methods and results

The gene sequence of *E. coli* YqgF was cloned into a pET29 expression vector (Novagen) with a cleavable N-terminal S-tag and a C-terminal His_6-tag. Recombinant YqgF was overexpressed in *E. coli* strain BL21(DE3). The cell cultures were grown at 37 °C in M9 minimal medium with $^{15}NH_4Cl$ as the sole nitrogen source and/or ^{13}C-glucose as the sole carbon source. Recombinant protein expression was induced with IPTG (1 mM final concentration) at an optical density (600 nm) of 1.2. After 3 h of additional growth, the cells were harvested by centrifugation and the cell pellets were frozen at −20 °C. The cell pellets were thawed at room temperature and resuspended into a lysis buffer containing BPER (PIERCE), 10% glycerol, 0.3 M NaCl, 0.2 mM dithiothreitol (DTT), and 100 units ml^{-1} of DNAase I (Sigma). After centrifugation, the supernatant was applied to a 4 ml l^{-1} Ni-NTA agarose column (Qiagen) and a washing buffer containing 25 mM imidazole. YgqF was eluted off from the column with elution buffer containing 250 mM imidizole. The N-terminal S-tag was removed from purified protein by a 12-h digestion with 5 units mg^{-1} of thrombin (Enzyme research

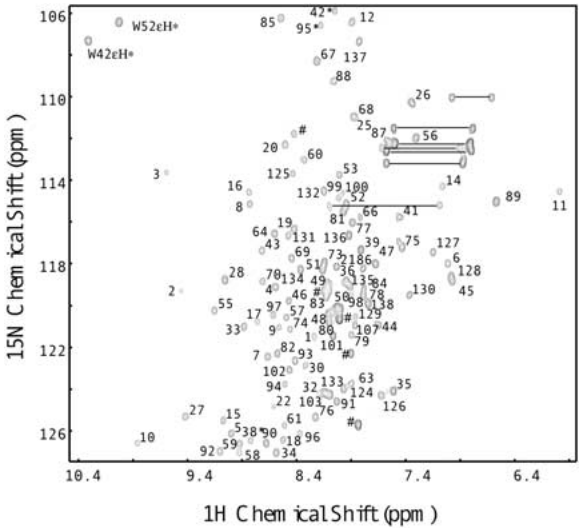

Figure 1. 2D ^{1}H-^{15}N HSQC of 0.8 mM YqgF acquired at 600 MHz. The peaks in the spectrum are labeled with residue numbers. Pairs of peaks connected by a horizontal line are from NH_2 resonances of Asn and Gln side chains. Peaks labeled with an asterisk are aliased due to the small sweep width used; peaks labeled with # are from the C-terminal tag. Note that in addition to 21 missing backbone amide peaks, those of residues 23, 24, 72, and 106 are too weak to be seen in this spectrum.

labs) at 4 °C. The cleaved products were separated using a Superdex-75 column. The final protein product contains additional RGSMADIGS and LEHHHHHH polypeptide sequences at its N- and C-termini, respectively. The correct molecular weight was confirmed by MALDI mass spectrometry. The yield of purified protein was 20 mg l^{-1}. NMR samples containing 0.3–1.0 mM protein were prepared in 75 mM K_iPO_4, pH 7.5, 5 mM dithiothreitol (DTT), 0.015% NaN_3, 0.4 mM 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF), 5% D_2O.

NMR experiments were performed at 25 °C on a Varian *INOVA* 600 MHz spectrometer. Spectra were processed with FELIX980 (Accelrys Inc.) and analyzed with NMRView (Johnson and Blevins, 1994) on a Silicon Graphics workstation. The HNCA, HNCOCA, CBCA(CO)NH, HNCACB, HNCO, ^{15}N-edited NOESY-HSQC, and ^{15}N-edited TOCSY-HSQC experiments (Proteinpack, Varian Inc.) were used to obtain ^{1}HN, ^{1}Hα, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}C′ resonance assignments.

Extent of assignments and data deposition

The current construct includes 9 and 8 extra residues at the N- and C-termini, respectively, which are excluded from the assignment statistics. The ^{1}H-^{15}N HSQC spectrum of YqgF shown in Figure 1 displays excellent proton and nitrogen chemical shift dispersion which is indicative of a well-folded protein. However, the number of cross peaks that were observed in the spectrum was less than expected from its amino acid sequence. Of the possible 132 non-proline backbone amide resonances, 21 were not observed in any of the acquired NMR spectra. As a consequence, no backbone amide assignments are available for the corresponding residues 13, 37, 65, 104, 105 and 108 to 123. Missing peaks are presumed to be from residues in slow conformational exchange and consequently broadened beyond detection. Lowering the pH to pH 7.0 resulted in one or two more peaks in the ^{1}H-^{15}N HSQC spectrum, but decreased the solubility and stability of the protein (pI = 6.7). 122 ^{13}Cα and 109 ^{13}Cβ chemical shifts were assigned by analyzing HNCA, HNCACB and HNCOCA, CBCA(CO)NH triple resonance data. In addition, 109 carbonyl and 103 ^{1}Hα chemical shifts were obtained from an HNCO and an ^{15}N-edited TOCSY-HSQC experiment, respectively. The ^{1}H, ^{15}N, and ^{13}C chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 5270.

Acknowledgements

We thank Dr Charles McNemar for providing protein N-terminal sequencing data, Drs Urooj Mirza and Yan-Hui Liu for carrying out Mass spectrometry analysis.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) *Nucl. Acids Res.*, **25**, 3389–3402.

Aravind, L., Makarova, K.S., and Koonin, E.V. (2000) *Nucl. Acids Res.*, **28**, 3417–3432.

Burley, S.K. (2000) *Nat. Struct. Biol.*, **7**, 932–934.

Freiberg, C., Wieland, B., Spaltmann, F., Ehlert, K., Brötz, H. and Labischinski, H. (2001) *J. Mol. Microbiol. Biotechnol.*, **3**, 483–489.

Hand, W.L. (2000) *Adolesc. Med.*, **11**, 427–438.

Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603–614.