Journal of Biomolecular NMR **27:** 389–392, 2003.
KLUWER/ESCOM

Letter to the Editor: Solution structure of the hypothetical protein YqgF from *Escherichia coli* reveals an RNAse H fold

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

Received 31 March 2003; Accepted 5 June 2003

Key words: heteronuclear NMR, Holliday junction resolvase, structural genomics

Biological context

YqgF is an *Escherichia coli* hypothetical protein, which was found essential for *E. coli* growth (Freiberg et al., 2001). YqgF is a representative member of a functionally uncharacterized protein family found in many bacteria (Aravind et al., 2000). Iterative database searches suggested that YqgF is likely a homolog of the *E. coli* RuvC protein, a Holliday junction resolvase. YqgF was predicted to be a nuclease (Aravind et al., 2000), since its predicted secondary structure profile was similar to that of the RuvC protein, which possesses an RNAse H fold. More recently, the YqgF-homologous domains were identified in bacterial Tex orthologues and eukaryotic Spt6p orthologues within their CXZ regions (Ponting, 2002). Spt6p is a transcription elongation factor of *Saccharomyces cerevisiae* whereas the molecular function of Tex is currently unknown. Here, we describe the solution structure of YqgF determined by nuclear magnetic resonance (NMR) spectroscopy. The structure of YqgF confirmed that YqgF is indeed structurally similar to RuvC; however, substantial differences between YqgF and RuvC were revealed. Unlike RuvC and other known Holliday junction resolvases, YqgF is likely monomeric in solution based on ^{15}N relaxation and size exclusion chromatography data. These findings provide valuable information to understand the biological function of YqgF and validate YqgF as a potential anti-microbial drug target.

Methods and results

Protein expression and purification have been described previously (Liu et al., 2002). Briefly, the *E. coli* yqgF gene was cloned into a pET29 expression vector to produce a final construct encoding the YqgF protein with an N-terminal S-tag and a C-terminal His-tag. The plasmid was transformed into *E. coli* strain BL21(DE3) for protein expression. The expressed protein was purified using a combination of Ni-NTA and size exclusion chromatography. The final protein product contains additional RGSMADIGS and LEHHHHHH polypeptide sequences at its N- and C-termini, respectively.

All NMR spectra were acquired at 25 °C on a Varian INOVA 600 MHz spectrometer, processed with FELIX980 (Accelrys Inc.), and analyzed with NMRView (Johnson and Blevins, 1994). The resonance assignments were obtained based on the following NMR experiments: HNCA, HNCOCA, CBCA(CO)NH, HNCACB, HNCO, ^{15}N-edited TOCSY-HSQC and HCCH-TOCSY (Kay, 1997). The backbone resonance assignments were obtained based on the HNCA, HNCOCA, CBCA(CO)NH, HNCACB experiments. The side chain aliphatic resonance assignments were obtained from analysis of 3D ^{15}N-edited TOCSY and ^{13}C-HCCH-TOCSY experiments. The aromatic ring proton resonances were assigned based on 2D TOCSY, NOESY and 3D ^{13}C- and ^{15}N-NOESY data. Stereospecific assignments of the prochiral methyl groups of valine and leucine residues were obtained by an analysis of the relative peak sign in a 2D ^{13}C,^{1}H-CT-HSQC spectrum recorded on a sample obtained using 10% ^{13}C labeling as described previously (Neri et al., 1989). The side chain ^{1}H and ^{13}C chemical shifts of YqgF have been deposited in the BioMagResBank (accession code BMRB-5758).

The structures of YqgF were calculated using XPLOR following the same protocol as described previously (Liu et al., 2001). The following 3D NOESY datasets were acquired: 3D ^{15}N-edited NOESY-

*To whom correspondence should be addressed, E-mail: dingjiang.liu@spcorp.com

Table 1. Structural statistics for the 25 conformers of YqgF[a]

Distance restraints		
Total experimental restraints		1399
All NOE distance restraints		1112
Intraresidue		427
Interresidue sequential ($\|i-j\|=1$)		283
Interresidue medium range ($1<\|i-j\|<5$)		158
Interresidue long range ($\|i-j\|\geqslant 5$)		244
Hydrogen bond restraints[b]		106
Diheral angle restraints		181
ϕ, φ		91,90
CHARMM Lennard-Jones energies (kcal•mol^{-1})[c]		-109 ± 19
Ramachandran analysis (%)		
Residues in favored regions		74
Residues in additional allowed regions		29
Residues in generously allowed regions		5
Residues in disallowed regions		0.9
Average pairwise r.m.s. deviations (Å)		
	backbone	all heavy atoms
Residues 3–95 and 124–136	0.85 ± 0.12	1.37 ± 0.12

[a]Ensemble of the 25 structures were included for the calculation from 100 calculated structures. All of them had no violations of distance restraints > 0.5 Å and dihedral angle restraints > 5°.
[b]Two restraints for each hydrogen bond included.
[c]The CHARMM Lennard-Jones van der Waals energy term, which was not included in the force field of the simulated annealing or restrained minimization, was used to assess the atomic packing in the protein structure.

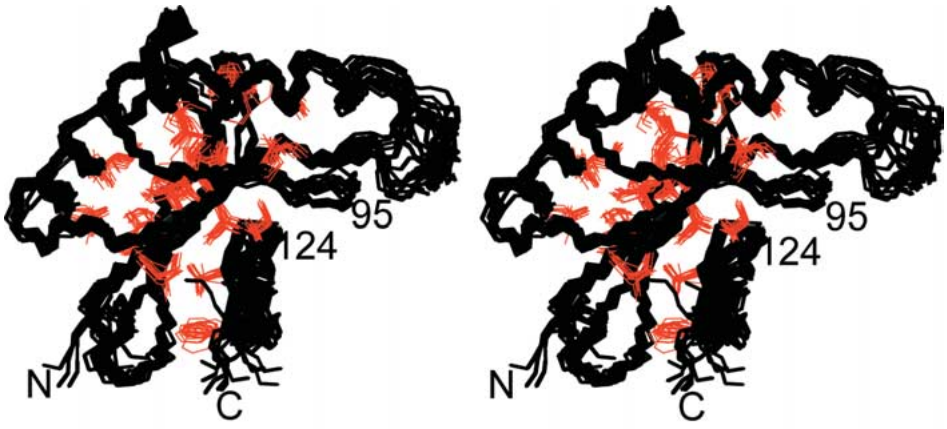

Figure 1. Stereoview of the NMR structure of YqgF. Backbone (N, C^{α} and C') of the 25 conformers of YqgF, with best fit for N, C^{α} and C' of residues 3–95 and 124–136. Residues 96–123, which are disordered in the structure, were omitted for clarity. The side chain heavy atoms of residues L5, L6, F8, I15, V17, V19, L44, I45, I57, I58, V59, A76, F79, I83, V91, L93, I128, L130 and F138 are highlighted in red.

HSQC with a U-[^{15}N,^{2}H]-labeled sample (τ_m = 100 ms), ^{15}N-edited NOESY-HSQC with a U-^{15}N-labeled sample (τ_m = 100 ms) and CN-NOESY-HSQC with a U-[^{13}C,^{15}N]-labeled sample (τ_m = 125) both in D_2O and H_2O. The hydrogen bond restraints were based on the identification of slow-exchanging HN protons in a ^{1}H-^{15}N HSQC spectrum acquired 3 hours after dissolving the YqgF protein into a D_2O buffer. A total of 1112 NOE distance restraints obtained from NOESY datasets, and 53 hydrogen bond restraints identified based on slow hydrogen-deuterium exchange rate were used in the structural calculation (Table 1). A total of 91 ϕ and 90 φ angle restrains, which were obtained by the program TALOS based on the backbone chemical shift data, were also included for the calculation. The overall RMSD of

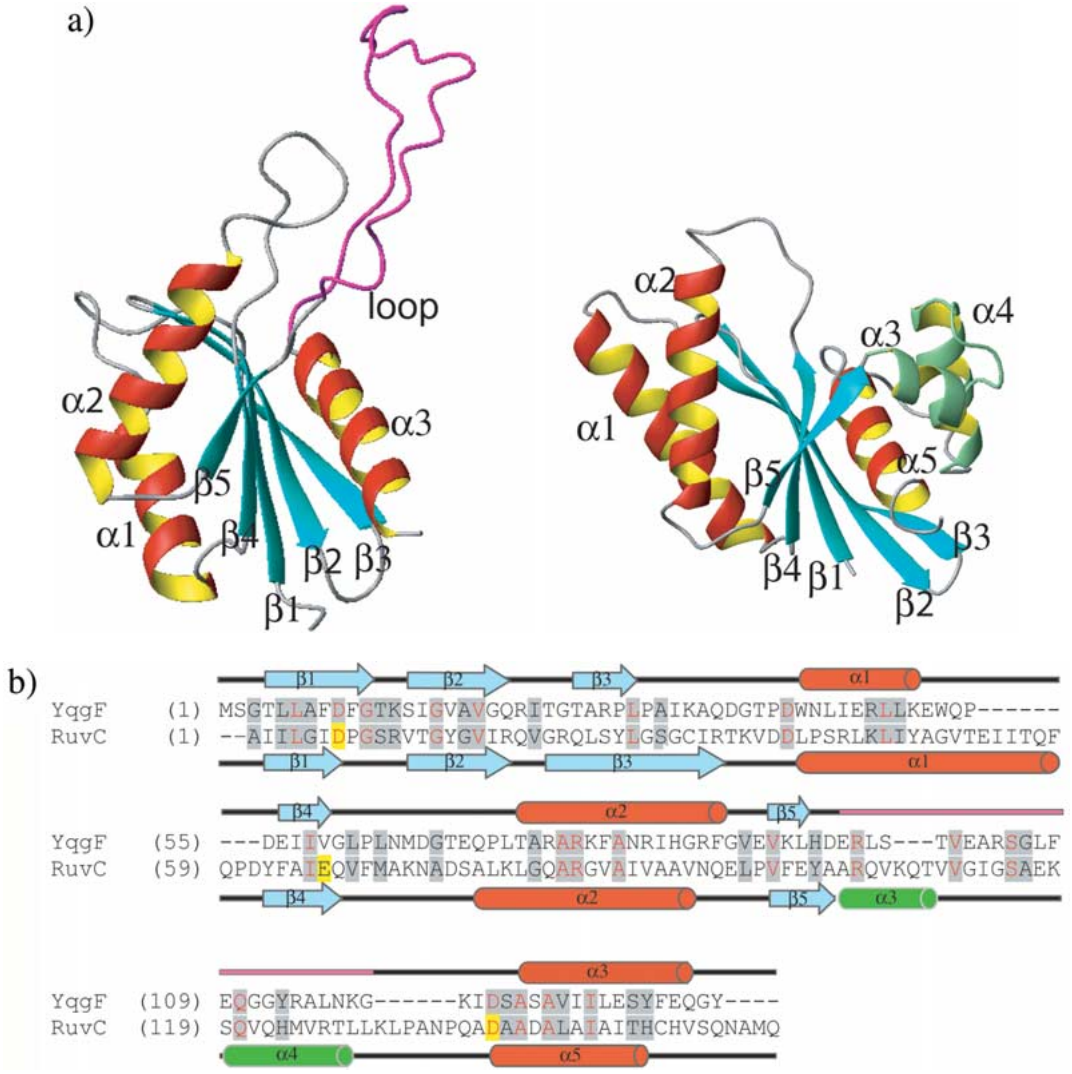

Figure 2. Structural comparison of YqgF and RuvC. (A) Ribbon representation of a representative NMR structure of YqgF (left) and the crystal structure of RuvC (right). The structures were superimposed with their common secondary structure elements as mentioned in the text and separated for clarity. Residues 96–123 of YqgF disordered in the structure is highlighted in magenta, the corresponding region in RuvC structure (helix α3 and α4) is highlighted in dark green. (B) Sequence alignment of YqgF and RuvC based on structural similarity. Conserved amino acid residues are shaded gray; red letters indicate identical residues. Asp 7, Asp 141 and Glu 66, which are important for catalytic activity of RuvC, are shaded yellow.

the 25 representative conformers (Figure 1) for the backbone atoms (C′, Cα and N) excluding the non-structured region (3–95, 124–136) is 0.85 ± 0.12 Å, and the RMSD for all the heavy atoms of the same region is 1.37 ± 0.12 Å. For the residues in the regular secondary structures (residues 4–21, 25–31, 42–50, 55–59, 73–86, 90–94, 124–137), the RMSDs of backbone and heavy atoms are 0.50 ± 0.09 Å and 1.08 ± 0.11 Å, respectively. The atom coordinates of the structure ensemble for YqgF have been deposited in the Protein Data Bank (PDB ID code: 1ovq).

The structure of YqgF consists of a central β-sheet sandwiched by two helices on one side and one helix on the other side (Figure 2a), which is typical for an 'α-β-α' fold. The central β-sheet is formed by 5 mixed β-strands in the order of 3-2-1-4-5. The strands 1, 2, and 3 are antiparallel to each other, whereas the strands 1, 4, and 5 are arranged in a parallel fashion. The secondary struc-

ture elements are arranged sequentially along the primary amino acid sequence in the following manner, β1(4-11)-β2(14-21)-β3(26-30)-α1(42-50)-β4(56-59)-α2(73-87)-β5(91-93)-loop-α3(124-136). A large segment of the sequence (96–123) is not structured since NOE data for this region are largely missing.

Relaxation rates were determined using standard methods as described previously (Liu et al., 2001). The backbone ^{15}N R_1 and R_2 data were obtained to estimate the overall correlation time of YqgF. The average ^{15}N R_1 and R_2 rates were determined as 1.61 ± 0.32 S^{-1} and 17.5 ± 1.1 S^{-1}. Base on the ratio of R_2/R_1, the overall correlation time τ_c was estimated to be 10.5 ns.

Discussion and conclusions

The structure of YqgF was compared with the crystal structure of Holliday junction resolvase, RuvC (PDB ID code 1hjr; Ariyoshi et al., 1994) (Figure 2). RuvC contains five strands and 5 helices in the following order: β1(2-7)-β2(12-21)-β3(24-34)-α1(40-58)-β4(62-67)-α2(77-92)-β5(98-102)-α3(103-110)-α4(118-128)-α5(140-151). Superposition of the alpha carbon atoms for 72 residues within all the regular secondary structure elements of YqgF with the corresponding regions of RuvC resulted a RMSD of 3.0 Å. While the regions of the β1-β2-β3-α1-β4-α2-β5 fold from both proteins are structurally equivalent, the third helix of YqgF (α3) superimposed with the fifth helix of RuvC (α5). Sequence identity and similarity between these two proteins are 13% and 27%, respectively (Figure 2b). The region of YqgF that corresponds to helices α3 and α4 of RuvC does not align well and is unstructured (Figures 1 and 2). Asp 7, Asp 141 and Glu 66 are three acidic amino acids believed to be active site residues in RuvC. As predicted by Aravind et al., 2000, Asp9 and Asp122 of YqgF can structurally be superimposed with the corresponding active site residues, Asp 7 and Asp141 of RuvC; the distances between the Cα carbon atoms for each pair of residues are less then 2 Å. However, Glu 96 of YqgF does not superimpose well with the third RuvC active site residue, Glu 66; the distance between their C^{α} atoms is larger than 7 Å.

Almost all known Holliday junction resolvases form dimers, such as RuvC (Ariyoshi et al., 1994), T4 endonuclease VII (Raaijmakers et al., 1999), archaeal Holliday junction resolvase Hjc (Nishino et al., 2001) and yeast mitochondrial Holliday junction resolvase Ydc2 (Ceschini et al., 2001). The dimeric form of the protein is required for proper binding to the Holliday junction DNA structure and for the introduction of paired nicks on opposing strands of the four-way DNA junction substrate (Bond et al., 2001). However, YqgF is likely a monomer in solution. An overall isotropic rotational correlation time of 10.5-ns was obtained for YqgF is consistent with YqgF being a 15 kDa monomeric protein in solution (Maciejewski et al., 2000). This is further supported by the apparent molecular mass of 17 kDa as determined by size exclusion chromatography experiments (data not shown). Therefore, the biochemical and structural data presented here suggest that, although YqgF structurally resembles RuvC, it is less likely to function as a Holliday junction resolvase. Functional studies using standard Holliday junction cleavage and DNA binding assays will be required to confirm this hypothesis.

Acknowledgements

The authors thank Presanti Repaka for her work on protein expression and purification, Todd Black, Robert Palermo and Shane Taremi for valuable discussion and technical assistance.

References

Aravind, L., Makarova, K.S. and Koonin, E.V. (2000) *Nucl. Acids Res.*, **28**, 1427–1464.

Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H. and Morikawa, K. (1994) *Cell*, **78**, 1063–1072.

Bond, C.S., Kvaratskhelia, M., Richard, D., White, M.F. and Hunter, W.N. (2001) *Proc. Natl. Acad. Sci.*, **98**, 5509–5514.

Ceschini, S., Keeley, A., McAlister, M.S.B., Oram, M., Phelan, J., Pearl, L.H., Tsaneva, I.R. and Barrett, T.E. (2001) *EMBO J.*, **20**, 6601–6611.

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

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

Kay, L.E. (1997) *Biochem. Cell. Biol.*, **75**, 1–15.

Liu, D., Repaka, P., Taremi, S.S. and Wyss, D.F. (2002) *J. Biomol. NMR*, **23**, 159–160.

Liu, D., Wang, Y.-S., Gesell, J.J. and Wyss, D.F. (2001) *J. Mol. Biol.*, **314**, 543–561.

Maciejewski, M.W., Liu, D., Prasad, R., Wilson, S.H. and Mullen, G.P. (2000) *J. Mol. Biol.*, **296**, 229–253.

Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) *Biochemistry*, **28**, 7510–7516.

Nishino, T., Komori, K., Ishino, Y. and Morikawa, K. (2001) *J. Biol. Chem.*, **276**, 35735–35740.

Ponting, C.P. (2002) *Nucl. Acids Res.*, **30**, 3643–3652.

Raaijmakers, H., Vix, O., Toro, I., Golz, S., Kemper, B. and Suck, D. (1999) *EMBO J.*, **18**, 1447–1458.