PROTEINS: Structure, Function, and Bioinformatics 55:775–777 (2004)

STRUCTURE NOTE

Crystal Structure of YIGZ, a Conserved Hypothetical Protein From *Escherichia coli K12* With a Novel Fold

Frances Park,* Ketan Gajiwala, Galina Eroshkina, Eva Furlong, Dongmei He, Yelena Batiyenko, Rich Romero, Jon Christopher, John Badger, Jorg Hendle, Johann Lin, Tom Peat, and Sean Buchanan
Structural GenomiX, Inc. (SGX), San Diego, California

Introduction. The *yigZ* gene from *Escherichia coli K12* is widely conserved among thermophiles, archaea, and pathogens including *Yersinia pestis, Vibrio cholerae* and *Salmonella typhimurium*. While the function for this protein remains uncharacterized, domain conservation patterns suggest that the *yigZ* gene product plays an indispensable function, making it a possible antimicrobial drug target.

The closest mammalian homologue of *yigZ* is the gene *impact* (Accession ID: AAG35736). Mouse *impact* is an imprinted gene but little else is known of its function. Imprinted genes are expressed in a parent-of-origin-dependent manner,[1] and generally have roles in differentiation, development, and regulation of cell proliferation. Aberrations in imprinted genes or their regulation have been implicated in various human diseases including Prader-Willi syndrome, Angelman syndrome, diabetes mellitus, bipolar affective disorder, and some malignant tumors.[2] Knowledge of the YIGZ protein structure provides a potential template to model the structure of the IMPACT protein and provide clues regarding biochemical function. This report describes the x-ray structure determination of YIGZ to 2.8 Å resolution. Structural homology provides some insights into the function of this protein and reveals a novel polypeptide chain fold.

Materials and methods. *yigZ* PCR product from the *E. coli* genome (Accession ID: NP_418290[3]) was TOPO isomerase cloned into HexaHis bacterial expression plasmids with a forward primer (5′-ATGGAAAGCTGGTTAATTCCTGC-3′) and a reverse primer (5′-CTTCTTCAATCGCTAACAATTGC-3′). The final amino acid sequence of the protein is as follows: MSLMESWLIP AAPVTVVEEI KKSRFITMLA HTDGVEAAKA FVESVRAEHP DARHHCVAWV AGAPDDSQQL GFSDDGEPAG TAGKPMLAQL MGSGVGEITA VVVRYYGGIL LGTGGLVKAY GGGVNQALRQ LTTQRKTPLT EYTLQCEYHQ LTGIEALLGQ CDGKIINSDY QAFVLLRVAL PAAKVAEFSA KLADFSRGSL QLLAIEEEGG SHHHHHH

Selenomethionine protein was expressed in *E. coli* BL21[DE3] cells. After lysis, the protein was purified via nickel ion affinity chromatography [50 mM TrisHCl, pH7.8, 500 mM NaCl, 10 mM imidazole, 10 mM methionine, 10% glycerol, 1 mM DTT with a linear gradient of 10–500mM imidazole]. The protein was then subjected to gel filtration chromatography with Superdex75 (Pharmacia). S-met protein was expressed and purified using similar procedures.

Crystals were obtained within four days at 9°C via hanging drop vapor diffusion with equal volumes of protein [5 mg/ml, 1.0 mM β-mercaptoethanol (βME), 150 mM NaCl, 10 mM HEPES pH 7.5, 10 mM methionine, 10% glycerol] and reservoir [150 mM MES pH 6.5, 150 mM $(NH_4)_2SO_4$, 30% PEG MME 5000 (w/v), 28.4 mM βME] solutions. Crystals were transferred to a cryoprotective solution consisting of 10% glycerol, 10% ethylene glycol, and 80% reservoir.

Diffraction data were collected from frozen S-met and Se-Met crystals at the APS COM-CAT beam line, processed and reduced with MOSFLM, SCALA, and TRUNCATE.[4] Four selenium atom sites were located using SnB[5] and refined with SHARP.[6] The SAD phased electron density map was improved by solvent flattening with SOLOMON[4] and used for initial model building. Refinement was performed against the S-Met dataset with iterative cycles of manual model building with XTALVIEW/XFIT,[7] REFMAC,[4] and CNX.[8] Quality assessment of the model was performed with PROCHECK,[9] WHATCHECK,[10] and SFCHECK[11] (Table I).

Results and discussion. The YIGZ protein is a 23.2 kD monomer consisting of two domains separated by a six amino acid linker. The N′ terminal domain (residues 3–136) consists of a five-stranded anti-parallel β sheet with three α helices. The C′ terminal domain (residues 139–208) consists of a four-stranded anti-parallel β-sheet with two α-helices packed against one face of the β-sheet [Fig. 1(A)].

The PredAct™[12] program is used to identify and rank polar residues that are conserved in sequence alignments and cluster within 5Å of one another. Ser23, His54, Glu77, and Arg104 almost certainly participate in catalysis [Fig. 1(B)]. These four residues are absolutely conserved among mammalian IMPACT proteins, which share less than 30% sequence identity overall. Lys22, Arg24, Phe25, Asp75,

*Correspondence to: Frances Park, Structural GenomiX, Inc. (SGX), 10505 Roselle St., San Diego, CA. 92121. E-mail: frances_park@stromix.com.

Received 5 December 2003; Accepted 16 December 2003

Published online 14 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20087

TABLE I. Data Collection and Refinement Statistics

Data Collection		
	Space group	R3
	Cell dimensions	a = b = 148.75 Å c = 35.60Å α = β = 90° γ = 120°
	Molecules per asymmetric unit	1
	Wavelength λ	0.9795 Å
	Overall Resolution limits	37.19–2.80 Å
	Number of reflections collected	81,478
	Number of unique reflections	7,231
	Overall Redundancy of data (last shell)	11.3 (11.3)
	Overall Completeness of data (last shell)	100.0% (100.0%)
	Overall R_{SYM} (last shell)	0.104 (0.485)
	I/σ(I) (last shell)	18.1 (8.4)
Model Refinement		
Model	Total number of atoms	1620
	Number of water molecules	91
	Crystal solvent content	62.91 %
Refinement	Resolution limits	37.19–2.80 Å
	Number of reflections used	7231
	Completeness	100.0%
	R-factor for all reflections	0.2591
	Correlation coefficient	0.8672
	Number of reflections above 2 σ(F) and resolution from 5.0 Å—high resolution limit	5963
	used to calculate $R_{working}$	5689
	used to calculate R_{free}	274
	R-factor without free reflections	0.247
	R-factor for free reflections	0.316
	Error in coordinates estimated by Luzzati plot	0.3850 Å
Validation	Phi-Psi in most favored region	85.1%
	Phi-Psi in additional allowed region	14.3%
	Phi-Psi in generously allowed region	0%
	Phi-Psi in disallowed region	0%
	RMSD from ideal bond length	0.013
	RMSD from ideal bond angle	2.4°

[a] $R_{sym} = \Sigma|I_i - \langle I\rangle| / \Sigma|I_i|$, for equivalent reflections.
[b] $R_{working} = (\Sigma||Fo| - |Fc||) / \Sigma|Fo|)$, calculated without free reflections.
[c] $R_{free} = (\Sigma||Fo| - |Fc||) / \Sigma|Fo|)$, calculated with free reflections.

Gly76, Pro78, Ala82, Tyr105, Tyr106, Gly107, Leu111, Leu116, Tyr120, Asp74, and Thr81 probably contribute to substrate binding. These residues cluster on the distal side of the N′ domain within an electropositive surface charge (data not shown).

A search for structural homologues of YIGZ was performed using the DALI server (http://www.ebi.ac.uk/dali/).[13] Queries with the entire protein or with the N′ terminal domain did not yield significant hits. A query with the C′ terminal domain revealed some possible functions. The four highest scoring, albeit weak, structural homologies of the C′ terminal domain include three ribosomal proteins [1DAR[14] (Z-score = 9.6, 1.9 Å RMSD for 69 equivalent α-carbons), 1JQS[15] (Z-score = 7.0, 2.4 Å RMSD for 66 equivalent α-carbons), and 1RIS[16] (Z-score = 8.2, 2.1 Å RMSD for 67 equivalent α-carbons)] and the propeptide portion of procarboxypeptidase a2 [1AYE[17] (Z-score = 6.5, 2.1 Å RMSD for 64 equivalent α-carbons)]. The ribosomal proteins include bacterial peptide elongation factor G (EF-G) and S6. EF-G and its eukaryotic counterpart, EF-2, are members of the GTPase superfamily of proteins.[18] The ribosomal protein S6 from *Thermus thermophilus* interacts with RNA. Sequence alignments of EF-G, S6, and the C′ terminal domain suggest that this portion of the YIGZ protein may interact with nucleic acids. The C′ terminal domain of the YIGZ protein also resembles the pro-portion of procarboxypeptidase A2. Pancreatic carboxypeptidases catalyze C′ terminal exoproteolysis of alimentary proteins and esters during digestion.[19] They are secreted as zymogens with a 94–95 residue N′ terminal pro-segment, which occludes the active site. The A2 isoform has a preference for aliphatic and aromatic C-terminal residues and has a clear specificity for bulkier aromatic C-terminal residues.

While the structure of the YIGZ protein from E. coli K12 has been determined, its biochemical and biological functions remain unclear. Structural evidence and sequence

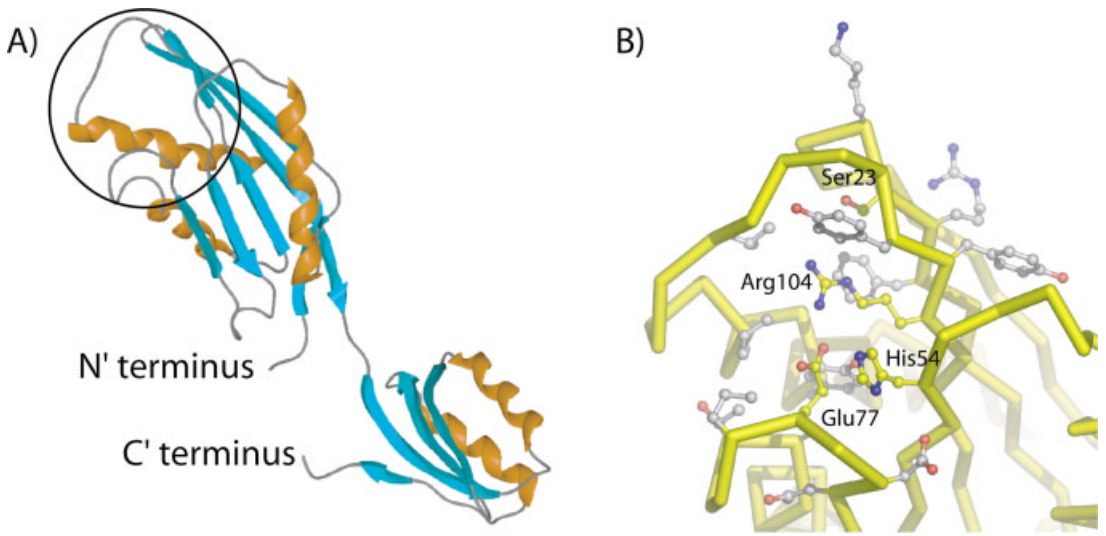

Fig. 1. A: SPOCK[20] ribbon diagram of YIGZ from *Escherichia coli K12*. β-strands and α-helices are shown in cyan and orange, respectively. Circle indicates region illustrated in part B. B: Putative catalytic residues and substrate binding site residues are rendered as ball-and-stick figures in yellow and grey, respectively. Figure produced with PyMOL.[21]

homology of the YIGZ protein suggest that the C′ terminal domain binds nucleic acids, but essentially shed no light on the function(s) supported by the balance of the polypeptide chain. Functional studies of the YIGZ protein or a member of the IMPACT families may be accelerated by our X-ray crystallographic results.

Acknowledgments. We thank the staff at the Advanced Photon Source and COM-CAT, for their help during data collection. The atomic coordinates (code: 1VI7) have been deposited in the Protein Data Bank (http://www.rcsb.org/).

REFERENCES

1. Morison IM, Reeve AE. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. Hum Mol Genet 1998;7:1599–1609.
2. Nakao M, Sasaki H. Genomic imprinting: significance in development and diseases and the molecular mechanisms. J Biochem (Tokyo) 1996;120:467–473.
3. Blattner FR, Plunketti GI, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF and others. The complete genome sequence of Escherichia coli K-12. Science 1997;277:1453–1462.
4. CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
5. Weeks CM, Miller R. The design and implementation of SnB version 2.0. J Appl Crystallogr 1999;32(Part 1):120–124.
6. La Fortelle Ed, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In: Sweet RM, Carter CW, editors. Methods in enzymology, vol. 276. New York: Academic Press; 1997. p 472–494.
7. McRee DE. XtalView Xfit—A versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125: 156–165.
8. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS and others. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54(Pt 5):905–921.
9. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck—a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr 1993;26(APR):283–291.
10. Vriend G. What If—a molecular modeling and drug design program. J Mol Graph 1990;8(1):52–56.
11. Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr 1999;55(Pt 1):191–205.
12. Gajiwala K; patent pending. Pub. No.: US-2003-0158671-A1.
13. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
14. al-Karadaghi S, Aevarsson A, Garber M, Zheltonosova J, Liljas A. The structure of elongation factor G in complex with GDP: conformational flexibility and nucleotide exchange. Structure 1996;4:555–565.
15. Agrawal RK, Linde J, Sengupta J, Nierhaus KH, Frank J. Localization of L11 protein on the ribosome and elucidation of its involvement in EF-G-dependent translocation. J Mol Biol 2001;311: 777–787.
16. Lindahl M, Svensson LA, Liljas A, Sedelnikova SE, Eliseikina IA, Fomenkova NP, Nevskaya N, Nikonov SV, Garber MB, Muranova TA and others. Crystal structure of the ribosomal protein S6 from Thermus thermophilus. Embo J 1994;13:1249–1254.
17. Garcia-Saez I, Reverter D, Vendrell J, Aviles FX, Coll M. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. Embo J 1997;16:6906–6913.
18. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: conserved structure and molecular mechanism. Nature 1991;349: 117–127.
19. Aviles FX, Vendrell J, Guasch A, Coll M, Huber R. Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. Eur J Biochem 1993;211: 381–389.
20. Christopher JA. SPOCK: The structural properties observation and calculation kit. College Station, TX: Texas A&M University; 1997.
21. DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002.