Published in final edited form as:

J Struct Funct Genomics. 2009 April ; 10(2): 151–156. doi:10.1007/s10969-009-9063-1.

Crystal structure of YfeU protein from *Haemophilus influenzae*: a predicted etherase involved in peptidoglycan recycling

Y. Kim, P. Quartey, R. Ng, T. I. Zarembinski, and A. Joachimiak

Midwest Center for Structural Genomics and Structural Biology Center, Biosciences, Argonne National Laboratory, 9700 South Cass Ave., Bldg 202, Argonne, IL 60439, USA

Y. Kim: ; P. Quartey: ; R. Ng: ; T. I. Zarembinski: ; A. Joachimiak: andrzejj@anl.gov

Abstract

The crystal structure of the *H. influenzae* YfeU protein, was determined at 1.90 Å resolution using multi-wavelength anomalous diffraction. YfeU belongs to a very large conserved family of proteins found mainly in bacteria but also in archaea and eukaryota. The protein is a homolog of eukaryotic glucokinase regulator and is predicted to be a sugar phosphate isomerase or aminotransferase. Here we describe the structure of YfeU and discuss the possible function as an etherase possibly involved in peptidoglycan recycling.

Keywords

Alpha-beta-alpha sandwich; Phosphosugar isomerase; Etherase

Introduction

Haemophilus influenzae is a non-motile, Gram negative coccobaccilli in the family *Pasteurellaceae* that is widespread in the human population [1]. *H. influenzae* type b (Hib) is a major cause of bacterial meningitis among children. *H. influenzae* is also one of the first free-living organisms to have its entire chromosome sequenced. This is of high importance because genetic transformation in *H. influenzae* has included drug resistance and synthesis of capsular antigens [1].

The crystal structure of the *H. influenzae* YfeU, a 303 residue protein, was determined at 1.90 Å resolution using multi-wavelength anomalous diffraction (MAD). YfeU belongs to a very large (>3,000 members) conserved family of proteins (Pfam1380, COG2103) found mainly in bacteria but also in archaea and eukaryota. The protein is a homolog of eukaryotic glucokinase regulator and is predicted to be a sugar phosphate isomerase or aminotransferase. This protein is also found as a domain in a very large number of multi-domain proteins. Although YfeU is assigned to be a putative regulator or an enzyme, the exact biochemical or cellular function of YfeU from *H. influenzae* has not been established. Here we describe the structure of YfeU and discuss the possible function as an etherase probably involved in peptidoglycan recycling.

Correspondence to: A. Joachimiak, andrzejj@anl.gov.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

Materials and methods

Cloning, protein purification and crystallization

The ORF of the *H. influenzae* protein HI0754 was amplified from genomic DNA with *KOD* DNA polymerase using conditions and reagents provided by the vendor (Novagen, Madison, WI). The gene was cloned into a pMCSG5 vector [2] using a modified ligation independent cloning protocol [3]. This process generated an expression clone producing a fusion protein with an N-terminal His₆ tag and a Factor Xa recognition site. The fusion protein was over-produced in *E. coli* BL21-derivative harboring a plasmid encoding three rare tRNAs (Arg [AGG/AGA] and Ile [ATA]). A selenomethionine (SeMet) derivative of the expressed protein was prepared as described by Walsh et al. [4]. The protein was purified by re-suspension of IPTG-induced bacterial cells in binding buffer (500 mM NaCl, 5% glycerol, 50 mM HEPES, pH 8.0, 20 mM imidazole, and 10 mM β -mercaptoethanol). The cells were lysed by the addition of lysozyme to 1 mg/ml in the presence of a protease inhibitor mixture cocktail (Sigma P8849) (0.25 ml/5 g cells) and sonication for 3 min. After clarification by centrifugation (30 min at 30,000 g) and passage through 0.4 and 0.2 μ m filters, the lysate was applied to 5 ml of Ni-NTA from Qiagen. The protein was eluted from the column with 50 mM HEPES, 500 mM NaCl, 250 mM imidazole, 10 mM β -mercaptoethanol and 5% glycerol. Factor Xa was added to the protein to cleave the His₆ tag. The cleaved protein was further purified on a MONO Q column using FPLC (Pharmacia). The purified protein was concentrated with simultaneous buffer exchange (20 mM HEPES, pH 7.5, 200 mM NaCl, 2 mM DTT), using the Ultrafree centrifugal device (Millipore, Inc., Bedford, MA.) with 5 kDa cutoff. Before crystallization, any particulate matter was removed from the sample by passage through a 0.2 μ m Ultrafree-MC centrifugal filtration device (Millipore).

Protein crystallization

The protein was crystallized by vapor diffusion in hanging drops by screening with Hampton Research Crystal Screens 1 and 2. Needle-like crystals were obtained under several conditions. These crystals were further optimized using Additive Screens from Hampton Research. The large (0.2 mm), data quality prisms were obtained in an optimized solution containing 0.2 M ammonium acetate, 0.1 M sodium citrate pH 6.5, 15% polyethylene glycol 4000 and 5 mM EDTA at 22°C. Crystals, which appeared after 3 days, were flash-frozen in liquid nitrogen with crystallization buffer plus 10% glycerol as a cryoprotectant prior to data collection. Crystals belong to C2 space group with cell dimensions $a = 104.66$ Å, $b = 41.27$ Å, $c = 77.25$ Å, $\alpha = \gamma = 90.00^\circ$, $\beta = 123.76^\circ$ and monomer in the asymmetric unit.

Data collection

Diffraction data were collected at 100 K at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The MAD data sets near selenium absorption peak (0.9794 Å), inflection point (0.9795 Å) and high energy remote (0.9464 Å) were collected from a SeMet labeled protein crystal using inverse-beam strategy up to 2.2, 2.5, 2.7 Å, respectively. One crystal (0.2 $\times$ 0.2 $\times$ 0.2 mm) was used to collect all three MAD data sets at 100 K. Data were collected with 8 s exposure/1.0°/frame (rotation of ω) using a 200 mm crystal to detector distance on the SBC-2 CCD detector. The total rotation range for each data set covered 360 degrees in ω , 180 degrees were with $\phi = 0$ and the other 180 degrees with $\phi = 180$, which was 30 degrees more than suggested range by the strategy module of HKL2000 [5]. In addition, the higher resolution native data were also collected to 1.90 Å from a different crystal (0.3 $\times$ 0.25 $\times$ 0.03 mm) of native protein with a 10 s exposure per 1 degree frame rotating 200 degrees in ω with $\phi = 0$. All data were processed and scaled with HKL2000 [5] (Table 1).

Structure determination and refinement

The structure was determined by MAD phasing using CNS [6] with the data sets collected at peak and inflection energies, and refined to 1.90 Å using CNS against the native data. The high-remote data was not included in the final phasing because the phasing became worse with the high-remote data. The initial partial model was traced using ARP/wARP [7]. Manual model building and adjustment of residues using O were required to complete the model for refinement. After several rounds of manual adjustment using O and CNS refinements, the model converged to a final R of 21.6% and a free R of 24.6% (Table 2). Electron density contoured at 1.0 σ was well connected for most of the main chain except for the first 11 N-terminal residues including the 3 residues (Ser, Asn, and Ala) left-over from the tag-cleavage and the 47 C-terminal residues which are disordered in the crystal structure. The stereochemistry of the structure was checked with PRO-CHECK [8] and the Ramachandran plot. The main chain torsion angles for all residues are in the allowed regions and additionally allowed regions.

Data deposition

Atomic coordinates of the protein YfeU from *H. influenzae* have been deposited into the Protein Data Bank (PDB) as 1NRI.

Discussion

The YfeU monomer is a three layer $\alpha/\beta/\alpha$ sandwich resembling an FAD/NAD(P)-binding Rossmann fold (Fig. 1a). In the middle, a left-hand-twisted sheet composed of five extended parallel β -strands $\beta 2$, $\beta 1$, $\beta 3$, $\beta 4$, and $\beta 5$ is flanked by four α -helices ($\alpha 1$, $\alpha 2$, $\alpha 3$, and $\alpha 8$) in the concave side and by three α -helices ($\alpha 5$, $\alpha 6$, and $\alpha 7$) in the convex side. This domain fold is highly homologous to those of well-conserved SIS (sugar isomerase) domains, which are found in many phosphosugar isomerases [9], phosphosugar binding proteins [10], and proteins that regulate the expression of genes involved in the synthesis of phosphosugars [11]. YfeU forms elaborate dimers with four α -helices ($\alpha 1$, $\alpha 2$, $\alpha 3$, and $\alpha 8$) contributing to the interface (Fig. 1b). The isomerase domain of glucosamine 6-phosphate synthase (GlmS), one of the SIS proteins, is bilobal in shape and consists of two topologically identical subdomains of approximately equal size in a single polypeptide chain. Each half-domain of an isomerase is structurally homologous to the YfeU monomer and the whole domain of an isomerase is similar to the YfeU dimer.

The HMMS library search using the ProFunc server [12] shows that residues 23–253 of YfeU match PF01488—a protein family that includes both shikimate and quinate dehydrogenases. Shikimate 5-dehydrogenase catalyses the conversion of shikimate to 5-dehydroshikimate and quinate 5-dehydrogenase catalyses the conversion of quinate to 5-dehydroquininate. YfeU bears decent resemblance to the catalytic domain of the shikimate dehydrogenase (1P74 and 1P77) [13,14] (RMSD = 3.40 Å), however, the NADP binding domain of the shikimate dehydrogenase is dissimilar. Structural alignment suggests that access to a potential NADP binding site in YfeU is blocked by the N-terminal region (residues 9–17) of the protein and the conserved sequence GAGGXX is absent.

The InterProScan [15], which attempts to find sequence motifs in a database of protein families, domains, and functional sites, produced three matching sequence motifs to the HI0754 YfeU protein. The three matching motifs are as follows: PS01272 (glucokinase regulatory protein), PF01380 (sugar isomerase), and TIGR00274 (glucokinase regulatory protein). The Secondary Structure Matching (SSM) program retrieved 12 structural matches with Z-scores higher than 6.0, nine of which were with root mean square deviations (rmsd) of less than 2.25 Å [16]. Most of higher matches are of phosphoheptose isomerases; 3BJZ (*Pseudomonas aeruginosa*), 1TK9

(*C. jejuni*), 2I2 W (*E. coli*), 2I22 (*E. coli*), 1X92 (*P. aeruginosa*), 1X94 (*V. cholerae*), 3CVJ (*B. halodurans*). Particularly, the crystal structure of putative phosphoheptose isomerase 1 (PDB ID = 1TK9) [17] had ten secondary structure elements (SSE) of the seventeen total elements that superimposed with an rmsd of 1.58 Å between the C-alpha atoms of the matched residues and a Z-score of 8.2 [16]. The crystal structure of *P. aeruginosa* phosphoheptose isomerase in complex with reaction product d-glycero-d-mannopyranose-7-phosphate (PDB ID = 1X92) had nine matching SSEs and an rmsd of 1.69 Å. The crystal structure of YckF from *Bacillus subtilis* (PDB ID = 1M3S) had nine matching SSEs and an rmsd of 2.24 Å.

A BLAST [18] search retrieved 50 sequence homologues for this protein in the PDB. This protein exhibited a high similarity to the C-terminal two-thirds of several glucokinase regulatory proteins and proteins of unknown function. Numerous significant hits were found for the N-terminal 2/3 of the protein (residues 1–201), which are 63% similar to the sequence of a glucokinase regulatory protein-related protein in *Vibrio cholerae*, and 55% similar to YfeU from *E. coli*, a 31.2 kD protein of unknown function in the *cysP-amiA* intergenic region. The Superfamily HMM program [19], which searches against a library of Hidden Markov Models (HMM) [20] derived from Structural Classification of Proteins Database [21] (SCOP) superfamilies, found a single motif between residues 9 and 256 that matches the characteristic SCOP motif 53697 of the dlnria_domain superfamily under the SIS domain SCOP fold. Overall, there is no significant sequence similarity (sequence identities less than 20%) between YfeU and any of these SIS proteins. However, in the active site, where a phosphosugar binds to an isomerase, located near the dimer interface in YfeU, several residues are conserved (Fig. 2). Particularly, in the structure of the isomerase domain of glucosamine 6-phosphate synthase (1MOS), most of the residues surrounding the substrate analogue 2-deoxy-2-amino glucitol-6-phosphate (GlcNol-6P), in the binding patch, specifically, Gly301, Thr302, Ser349, Gly350, Thr352, Ser401, and Thr402 are well conserved with the residues Gly77, Thr78, Ser147, Gly148, Thr150, Ser198, and Ser199 in similar positions, respectively, in YfeU (Fig. 3). These residues in the isomerases are involved in either binding the substrate and/or maintaining the binding pocket in the structure. This suggests YfeU may assume similar enzymatic activity.

More specific functional information can be derived from recent mechanistic studies of *E. coli* MurQ. These studies revealed that MurQ is an etherase—*N*-acetylmuramic acid 6-phosphate hydrolase that converts MurNAc 6-phosphate into GlcNAc 6-phosphate and lactate [22] and plays important role in peptidoglycan recycling. *H. influenza* YfeU shares 55% sequence identity (E value 2e-95) and 72% sequence similarity with MurQ. The Glu83 and Glu114 in MurQ when mutated to Ala or Gln, respectively, significantly reduce MurQ enzymatic activity. These residues (equivalent Glu89 and Glu120 in *H. influenza* YfeU) are fully conserved in YfeU family (Fig. 2). Interestingly the residues Gly77, Thr78, Glu120, Ser147, Gly148, Thr150, Ser198, and Ser199 from one subunit form a well defined pocket that is complemented by Glu89 from the opposite subunit in the dimer (Fig. 3b). All these residues are also conserved in YfeU family (Fig. 2). Based on the sequence conservation of the presumed active site we propose to transfer function from MurQ and suggest that YfeU is an etherase and is involved in recycling cell wall peptidoglycan in bacteria. The specific substrate specificity of YfeU will require further biochemical analyses. The access to this site appears to be regulated by the loop region 224–240 of the opposite subunit (Fig. 3b). As a result the enzyme has two independent active sites but is regulated by the opposite subunit.

Acknowledgments

We wish to thank all members of the Structural Biology Center at Argonne National Laboratory for their help in conducting experiments. This work was supported by National Institutes of Health Grant GM62414, GM074942 and by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

References

1. Todar, K. Todar's online textbook of bacteriology. University of Wisconsin-Madison Department of Bacteriology; Madison, WI: 2004.
2. Stols L, Gu M, Dieckman L, Raffin R, Collart FR, Donnelley MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif* 2002;25:8–15. [PubMed: 12071693]
3. Dieckman L, Gu M, Stols L, Donnelley MI, Collart FR. High throughput methods for gene cloning and expression. *Protein Expr Purif* 2002;25:1–7. [PubMed: 12071692]
4. Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A. Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr D Biol Crystallogr* 1999;55(6):1168–1173. [PubMed: 10329779]
5. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, vol 276: Macromolecular Crystallography, part A 1997:307–326.
6. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905–921. [PubMed: 9757107]
7. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* 2003;374:229–244. [PubMed: 14696376]
8. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283–291.
9. Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B. The mechanism of sugar phosphate isomerization by glucosamine 6-phosphate synthase. *Protein Sci* 1999;8(3):596–602. [PubMed: 10091662]
10. Peschke U, Schmidt H, Zhang HZ, Piepersberg W. Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78-11. *Mol Microbiol* 1995;16(6):1137–1156. [PubMed: 8577249]
11. Sorensen KI, Hove-Jensen B. Ribose catabolism of *Escherichia coli*: characterization of the rpiB gene encoding ribose phosphate isomerase B and of the rpiR gene, which is involved in regulation of rpiB expression. *J Bacteriol* 1996;178(4):1003–1011. [PubMed: 8576032]
12. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 2005;33:W89–W93. [PubMed: 15980588]
13. Ye S, von Delft F, Brooun A, Knuth MW, Swanson RV, McRee DE. The crystal structure of shikimate dehydrogenase (AroE) reveals a unique NADPH binding mode. *J Bacteriol* 2003;185(14):4144–4151. [PubMed: 12837789]
14. Benach J, Lee I, Edstrom W, Kuzin AP, Chiang Y, Acton TB, Montelione GT, Hunt JF. The 2.3 Å crystal structure of the shikimate 5-dehydrogenase orthologue YdiB from *Escherichia coli* suggests a novel catalytic environment for an NAD-dependent dehydrogenase. *J Biol Chem* 2003;278(21):19176–19182. [PubMed: 12624088]
15. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, Bucher P, Copley RR, Courcelle E, Das U, Durbin R, Falquet L, Fleischmann W, Griffiths-Jones S, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lopez R, Letunic I, Lonsdale D, Silventoinen V, Orchard SE, Pagni M, Peyruc D, Ponting CP, Selengut JD, Servant F, Sigrist CJA, Vaughan R, Zdobnov EM. The InterPro database, 2003 brings increased coverage and new features. *Nucleic Acids Res* 2003;31:315–318. [PubMed: 12520011]
16. Krissinel, E.; Henrick, K. Protein structure comparison in 3D based on secondary structure matching (SSM) followed by Ca alignment, scored by a new structural similarity function. In: Kungl, AJ.; Kungl, PJ., editors. Proceedings of the 5th international conference on molecular structural biology; Vienna. 3–7 September; 2003. p. 88
17. Seetharaman J, Rajashankar KR, Solorzano V, Kniewel R, Lima CD, Bonanno JB, Burley SK, Swaminathan S. Crystal structures of two putative phosphoheptose isomerases. *Proteins* 2006;63(4):1092–1096. [PubMed: 16477602]

18. Altschul SF, Madden TL, Schäffer 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. [PubMed: 9254694]
19. Madera M, Vogel C, Kummerfeld SK, Chothia C, Gough J. The SUPERFAMILY database in 2004: additions and improvements. *Nucleic Acids Res* 2004;32:D235–D239. [PubMed: 14681402]
20. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 2001;313:903–919. [PubMed: 11697912]
21. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–540. [PubMed: 7723011]
22. Hadi T, Dahl U, Mayer C, Tanner ME. Mechanistic studies on *N*-acetylmuramic acid 6-phosphate hydrolase (MurQ): an etherase involved in peptidoglycan recycling. *Biochemistry* 2008;47(44): 11547–11558. [PubMed: 18837509]
23. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the clustal series of programs. *Nucleic Acids Res* 2003;31:3497–3500. [PubMed: 12824352]

Fig. 1. Structure of YfeU protein. **a** A monomer with N- and C-termini labeled. **b** Structure of dimer

Fig. 2.

Primary sequence alignment using ClustalX [23] showing conservation of residues (red) in the family and mapping with the secondary structure elements in YfeU protein (gi|68196068 YfeU, *Enterococcus faecium*, gi|34810976 *Haemophilus influenzae*, gi|75818318 *Vibrio cholerae*, gi|22125205 *Yersinia pestis*, gi|49612710 *Erwinia carotovora*, gi|28379837 *Lactobacillus plantarum*, gi|16130353 *Escherichia coli*, gi|24052849 *Shigella flexneri*, gi|16077238 *Bacillus subtilis*, gi|27468807 *Staphylococcus epidermidis*, gi|65318242 *Bacillus anthracis*, gi|46907889 *Listeria monocytogenes*, gi|49482433 *Staphylococcus aureus*, gi|56708533 *Francisella tularensis*, gi|56543764 *Zymomonas mobilis*)

Fig. 3. Structural comparison of the isomerase domain of glucosamine 6-phosphate synthase (1MOS) with YfeU. **a** Stereo plot of SSM superposition [12] of the isomerase domain of 1MOS and the YfeU dimer. The isomerase domain of 1MOS is shown in dark grey and the YfeU in cyan and orange. **b** Close-up of the binding patch with the same scheme as in (a), Conserved residues inside of the patch are indicated in red, the corresponding null mutations, Glu89 and Glu120, identified from *E. coli* MurQ [22] are shown by a stick model, and the substrate GlcNol-6P for 1MOS is also shown as a stick model

Table 1

Summary of crystal and data collection

Crystal				
Unit cell	$a = 104.66 \text{ \AA}, b = 41.27 \text{ \AA}, c = 77.25 \text{ \AA}, \alpha = \gamma = 90.00^\circ, \beta = 123.76^\circ$			
Space group	C2 (C 1 2 1)			
MW Da (residues)	32,822			
Mol (AU)	1			
SeMet (AU)	8			
	MAD Edge	Peak	Remote	Native High resolution
Wavelength (Å)	0.9794	0.9795	0.9464	0.9795
Resolution range (Å)	2.50	2.2	2.7	1.88
No. of unique reflections	9,841	14,164	7,857	22,432
Completeness (%)	99.9	99.6	99.9	98.8
<i>R</i> merge (%)	10.4	12.6	11.8	7.8

Table 2

Crystallographic statistics

Phasing	Centric (Inf) FOM	Phasing power	Acentric (Inf/peak) FOM	Phasing power	No. of reflections
Resolution range (Å)	0.343	1.13	0.573	2.32/2.56	14,163
37.5–2.19					
FOM after modification	Density	0.86			
<i>Refinement</i>					
Resolution range (Å)			43.50–1.90		
No. of reflections			21,252		
σ cutoff on F			0		
<i>R</i> -value (%)			21.6		
Free <i>R</i> -value (%)			24.6		
Rms deviations from ideal geometry					
Bond length (1–2) (Å)			0.005		
Angle (°)			1.1		
Dihedral (°)			21.8		
Improper (°)			0.81		
No. of atoms					
Protein			1,879		
Water			196		
Mean B-factor (Å ²)			23.0		
Ramachandran plot statistics (%)					
Residues in most favored regions			90.4%		
Residues in additional allowed regions			8.3%		
Residues in disallowed region			0.0%		