

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/7823170>

Conserved protein YecM from *Escherichia coli* shows structural homology to metal-binding isomerases and oxygenases

ARTICLE in PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2003

Impact Factor: 2.63 · DOI: 10.1002/prot.10307 · Source: PubMed

CITATIONS

5

READS

13

8 AUTHORS, INCLUDING:



Rongguang Zhang

131 PUBLICATIONS 8,943 CITATIONS

SEE PROFILE



Roman Aleksander Laskowski

European Molecular Biology Laboratory

99 PUBLICATIONS 30,908 CITATIONS

SEE PROFILE



Andrzej Joachimiak

Argonne National Laboratory

400 PUBLICATIONS 16,462 CITATIONS

SEE PROFILE



Alex Savtchenko

Sanford Burnham Prebys Medical Discovery I...

157 PUBLICATIONS 4,147 CITATIONS

SEE PROFILE

Published in final edited form as:

Proteins. 2003 May 1; 51(2): 311–314. doi:10.1002/prot.10307.

Conserved Protein YecM From *Escherichia coli* Shows Structural Homology to Metal-Binding Isomerases and Oxygenases

Rong-Guang Zhang¹, Norma Duke¹, Roman Laskowski², Elena Evdokimova³, Tatiana Skarina³, Aled Edwards^{3,4}, Andrzej Joachimiak^{1,*}, and Alexei Savchenko³

¹Biosciences Division and Structural Biology Center, Argonne National Laboratory, Argonne, Illinois

²Department of Crystallography, Birkbeck College, London, United Kingdom

³Clinical Genomics Centre/Proteomics, University Health Network, Toronto, Ontario, Canada

⁴Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada

Introduction

The crystal structure of protein YecM¹ has been determined at 1.6 Å resolution as a part of the ongoing structural genomics initiative (<http://www.mcsg.anl.gov>). The YecM is a conserved, hypothetical *Escherichia coli* protein with sequence homologs found exclusively in bacteria, including *Salmonella typhimurium*, *Yersinia pestis*, *Vibrio cholerae*, *Haemophilus influenza*, and *Pasteurella multocida* (Fig. 1). YecM (188 residues) shows also sequence similarity to proteins in COG database (<http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palox-?COG3102>). YecM (Pfam-B domain 24546) was selected as a structural genomics target because it shows no sequence similarity with proteins of known three-dimensional structure and therefore, may contain a previously unobserved fold.

Materials and Methods

Protein Cloning Expression and Purification. The ORF of YecM was amplified, cloned, and protein was purified and concentrated following procedures described previously.⁴ The ORF of YecM was amplified by PCR from *E. coli* genomic DNA (ATCC). The gene was cloned into the *Nde*I and *Bam*HI sites of a modified pET15b cloning vector (Novagen) in which the TEV protease cleavage site replaced the thrombin cleavage site and a double-stop codon was introduced downstream from the *Bam*HI site. This construct provides for an N-terminal hexahistidine tag separated from the gene by a TEV protease recognition site (ENLYFQ ↓ G). The fusion protein was overexpressed in *E. coli* BL21-Gold (DE3) (Stratagene) harboring plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile).

Large-scale expression of the recombinant protein was performed as described previously.⁴ The sample was induced at an OD₆₀₀ of 0.6–0.8 with 0.4 mM IPTG after growth at 37°C. The cells were harvested by centrifugation, and the cell pellet was resuspended in 40 mL with binding buffer, supplemented with 1 mM each of the protease inhibitors PMSF and

*Correspondence to: Dr. Andrzej Joachimiak, Biosciences Division, Structural Biology Center, Argonne National Laboratory, 9700 South Cass Avenue, Argonne IL 60439. andrzejj@anl.gov.

The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory ("Argonne") under Contract No. W-31-109-ENG-38 with the U.S. Department of Energy. 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.

benzamidine, flash-frozen in liquid nitrogen and stored at -70°C . The purification procedure used buffers containing 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, and 5, 30, and 250 mM imidazole for the binding, wash, and elution buffers, respectively. The harvested cells were lysed by adding 0.5% NP-40 to the thawed sample before sonication (5×30 s; D.C. 50%; O.L. 6). Fresh protease inhibitors were added before the sample was clarified by centrifugation (30 min @ 17,000 rpm; Beckman Coulter Avanti J-25 centrifuge). The clarified lysate was passed by gravity through a DE52 column in series with a Ni^{2+} -column. The bound protein was removed with elution buffer, and its concentration was determined by the Bradford assay. The sample was then brought to a final concentration of 0.5 mM EDTA, followed by the addition of a final concentration of 0.5 mM DTT. The His₆-tag was removed by cleavage with recombinant His-tagged TEV protease (60 μg TEV per mg recombinant protein). The His-tag and His-tagged TEV protease are purified from the recombinant protein by passage through a second Ni^{2+} -column. The sample was prepared for crystallization by dialysis in 10 mM HEPES, pH 7.5, 500 mM NaCl, followed by concentration to 10 mg/mL using a BioMax concentrator (Millipore). Se-Met-labeled protein was prepared by using this same procedure.

Protein Crystallization

The protein was crystallized by vapor diffusion in hanging drops by mixing 2 μL of the protein at the concentration of 10 mg/mL with 2 μL of 2% PEG 400 and 2.2 M ammonium sulfate in 0.1 M HEPES buffer at pH 7.5. Crystals were flash-frozen in liquid nitrogen with crystallization buffer plus 10 or 20% glycerol or ethylene glycol as cryoprotectant before data collection.

The crystal structure of Se-Met-derivatized protein was determined by using multi wavelength anomalous diffraction (MAD). The diffraction data were collected at the Advanced Photon Source (APS) Structural Biology Center (SBC) sector 19ID and BM beamline. Data collection statistics are listed in Table I.

Discussion

In the crystal, the YecM is a monomer. The eight, mostly antiparallel β -strands form an extensively curved sheet that wraps around C-terminal α -helix and a presumed active site, forming a deep groove. This surface is decorated with highly conserved residues. The β -sheet floor is buttressed by four α -helices, two on either side of the curved sheet, yielding a pseudo-twofold axis running down the center of the structure, as shown in Figure 2. The longest α -helix runs across the convex surface of the β -sheet, shielding it from solvent. Despite low-sequence similarity, the program DALI⁵ revealed several structural homologues of YecM. The closest homologue was the isomerase, methylmalonyl-coenzymeA epimerase⁶ (Z score of 7.8, RMSD = 3.3 Å, 110 equivalenced residues, 15% sequence identity), containing an ancient metal-binding scaffold. In addition, strong structural similarities were found to the oxidoreductases catechol 2,3-dioxygenase from *Pseudomonas putida*⁷ (Z score 7.5, RMSD = 3.3 Å, 106 equivalenced residues, 10% sequence identity), 4-hydroxyphenylpyruvate dioxygenase from *Pseudomonas fluorescens*⁸ (Z score 7.3 RMSD = 2.9 Å, 114 equivalenced residues, 9% sequence identity), and biphenyl-cleaving extradiol dioxygenase from *Burkholderia cepacia*⁹ (Z score = 6.1, RMSD = 2.9 Å, 103 equivalenced residues, 11% sequence identity). Except for extradiol dioxygenase, all YecM structural homologs are oligomeric. YecM is a monomer because its oligomerization surface is blocked by an α -helix (residues 12–34). Somewhat looser correlation was found to both human lyase glyoxalase I¹⁰ (Z score = 3.7, RMSD = 4.8 Å, 72 equivalenced residues, 8% sequence identity), yeast poly (A) polymerase¹¹ (Z score = 3.1, RMSD = 4.4 Å, 105 equivalenced residues, 10% sequence identity), and bleomycin resistance protein¹² (Z score = 2.9, RMSD = 3.8 Å, 68 equivalenced residues, 12% sequence identity). These data provide strong evidence that proteins with very

low sequence identity (8–15%) can assume virtually this same fold and suggest that the degeneracy of the “secondary code” is very high.

Further investigation of the YecM structural homologs reveals that all the proteins bind a divalent metal cation; methylmalonyl-coenzyme A epimerase⁶ was shown to bind Co^{+2} , catechol 2,3-dioxygenase(4), 4-hydroxyphenylpyruvate dioxygenase⁸, biphenyl-cleaving extradiol dioxygenase,⁹ and bleomycin¹¹ all bind Fe^{+2} , and glyoxalase I¹⁰ binds Zn^{+2} . Table II shows comparison of residues involved in metal binding of methylmalonyl-coenzyme A epimerase and glyoxalase I and equivalent residues in YecM. This comparison suggests that YecM may be a metal-binding protein and may function as an enzyme. Thus far, no divalent metal ion specificity has been reported in the literature for YecM protein. Our structural analysis would argue strongly in favor of such a circumstance. However, because the structurally homologous proteins bind a variety of divalent metal cations, the exact identity of which cation YecM might preferentially bind remains undetermined, with Co^{+2} or Zn^{+2} being good candidates.

Acknowledgments

We thank all members of the Structural Biology Center at Argonne National Laboratory for their help in conducting experiments and Lindy Keller for help in preparation of this manuscript.

Grant sponsor: National Institutes of Health; Grant number: GM62414-01; Grant sponsor: U.S. Department of Energy, Office of Biological and Environmental Research; Grant number: W-31-109-Eng-38.

References

1. The YecM coordinates have been submitted to the Protein Data Bank, entry 1K4N. Sequence information may also be found under Swiss Protein Databank acquisition number: P52007. A four-residue cloning tag (Gly-His-Met-Ile) has been added to the beginning of sequence P52007; in addition, the final two residues (Ser-Ala) are absent.
2. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988;85:2444–2448. [PubMed: 3162770]
3. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
4. Zhang, Rg; Skarina, T.; Katz, JE.; Beasley, S.; Khachatryan, A.; Vyas, S.; Arrowsmith, CH.; Clarke, S.; Edwards, A.; Joachimiak, A.; Savchenko, A. Structure of *Thermotoga maritima* SurE—stationary phase survival protein—a novel acid phosphatase. *Structure* 2001;9:1095–1106. [PubMed: 11709173]
5. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123–138. [PubMed: 8377180]
6. McCarthy AA, Baker HM, Sherwy SC, Patchett ML, Baker EN. Crystal structure of methylmalonyl-coenzyme A epimerase from *P. shermanii*; a novel enzymatic function on an ancient metal binding scaffold. *Structure* 2001;9:637–646. [PubMed: 11470438]PDB acquisition number: 1JC4
7. Kita A, Kita S, Fujisawa I, Inaka K, Ishida T, Horiike K, Nozaki M, Miki K. An archetypical extradiol-cleaving catecholic dioxygenase; the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from *Pseudomonas putida* mt-2. *Structure* 1994;7:25–34. [PubMed: 10368270] PDB acquisition number: 1MPY
8. Serre L, Sailland A, Sy D, Boudec P, Rolland A, Pebay-Peyroula E, Cohen-Addad C. Crystal structure of *Pseudomonas fluorescens* 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. *Structure* 1999;7:977–988. [PubMed: 10467142]PDB acquisition number: 1CJX
9. Han S, Eltis LD, Timmis KN, Muchmore SW, Bolin JT. Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading pseudomonad. *Science* 1995;270:976–980. [PubMed: 7481800]PDB acquisition number: 1HAN

10. Cameron AD, Ridderstrom M, Olin B, Kavarana MJ, Creighton DJ, Mannervik B. Reaction mechanism of glyoxalase I explored by an X-ray crystallographic analysis of the human enzyme in complex with a transition state analogue. *Biochemistry* 1999;38:13480–13490. [PubMed: 10521255] PDB acquisition number: 1QIP
11. Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moor CL, Bohm A. Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science* 2000;289:1346–1349. [PubMed: 10958780] PDB acquisition number: 1FA0
12. Maruyama M, Kumagai T, Matoba Y, Hayashida M, Fujii T, Hata Y, Sugiyama M. Crystal structures of the transposon Tn5-carried bleomycin resistance determinant uncomplexed and complexed with bleomycin. *J Biol Chem* 2001;276:9992–9999. [PubMed: 11134052] PDB acquisition number: 1ECS
13. WebLab ViewerPro. Accelrys; 2001.

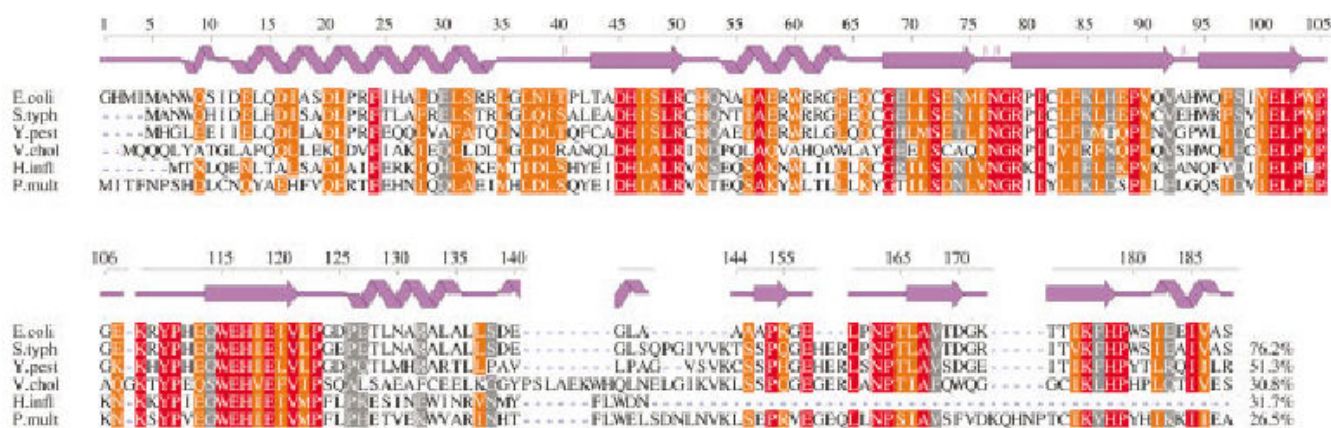


Fig. 1.
Multiple alignment of YecM sequence homologs using FASTA² and CLUSTAL W.³
Secondary structure elements of YecM are indicated above the sequence.

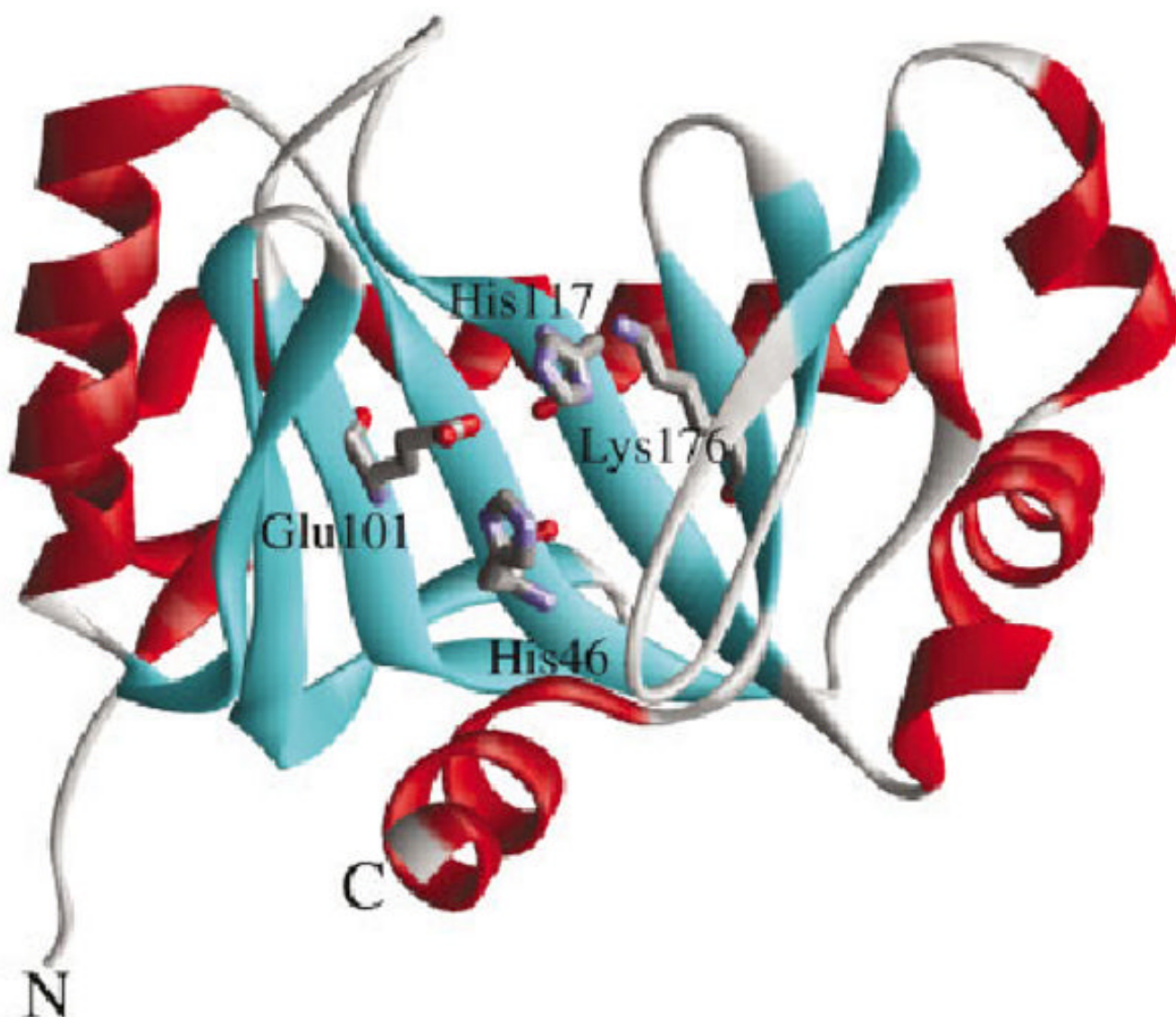


Fig. 2.
Structure of YecM. Pseudo-twofold axis approximately along line-of-sight. Presumed metal-binding residues are labeled as well as C- and N-termini. Diagram was created by using program WebLab.¹³

TABLE I
Summary of YecM Crystal Data, MAD Data Collection and Refinement

| Unit cell parameters (angstroms, degrees) | a 38.661 Å, b = 41.783 Å, c = 105.748 Å; $\alpha = \beta = \gamma = 90.00$ degrees | | |
|---|--|-------------|-------------|
| Space group | P2 ₁ 2 ₁ 2 ₁ (#19) | | |
| Molecular weight | 21,485.4 Da | | |
| Molecules per asymmetric unit (a.u.) | 1 | | |
| Se-Met residues per a.u. | 3 | | |
| MAD data | Edge energy | Peak energy | High energy |
| Wavelength (Å) | 0.9798 | 0.9796 | 0.94656 |
| Resolution limit (Å) | 1.6 | 1.6 | 1.6 |
| Number unique reflections: | 24315 | 24345 | 24275 |
| Overall data completeness (%) | 99.4 | 99.6 | 99.3 |
| Overall data redundancy | 7.0 | 7.0 | 6.9 |
| Overall Rmerge (%) | 4.8 | 4.7 | 4.9 |
| Figure of merit (FOM) | | 0.5370 | |
| Phasing power | 1.5177 | 1.4771 | 0.9914 |
| Resolution range (Å) | 50.0–1.6 | | |
| No. of reflections (all) | 24450 | | |
| No. of reflections (observed) | 23643 | | |
| Reflections observed (%) | 96.7 | | |
| σ cutoff | 2.0 | | |
| Overall R-value (%) | 19.9 | | |
| Free R-value (%) | 21.7 | | |
| RMS deviations from ideal geometry | | | |
| Bond length | 0.005 | | |
| Angle | 1.3 | | |
| Dihedral | 24.4 | | |
| Improper | 0.87 | | |
| Number of: | | | |
| Protein nonhydrogen atoms | 1462 | | |
| Water molecules | 253 | | |
| Mean B-factor (Å ²) | 18.5 | | |
| Ramachandran plot statistics (%) | | | |
| Residues in most favored regions | 95.3 | | |
| Residues in additional allowed regions | 4.1 | | |
| Residues in generously allowed regions | 0.7 | | |
| Residues in disallowed region | 0.0 | | |

TABLE II
Comparison of Metal-Binding Site Residues in *E. coli* YecM, *Propionibacterium shermanii* Methylmalonyl-Coenzyme A Epimerase, and Human Glyoxalase I

| YecM | Methylmalonyl-coenzyme A epimerase | Glyoxalase I |
|---------|------------------------------------|--------------|
| His 046 | His 012 | Gln 033 |
| Glu 101 | Gln 065 | Glu 099 |
| His 117 | His 091 | His 126 |
| Lys 176 | Glu 141 | Glu 172 |