

STRUCTURE NOTE

Crystal structure of YagE, a putative DHDPS-like protein from *Escherichia coli* K12

Sankar Manicka,¹ Yoav Peleg,² Tamar Unger,² Shira Albeck,² Orly Dym,² Harry M. Greenblatt,³ Gleb Bourenkov,⁴ Victor Lamzin,⁴ Sankaran Krishnaswamy,¹ and Joel L. Sussman^{2,3*}

¹ School of Biotechnology, Madurai Kamaraj University, Madurai 625021, Tamilnadu, India

² Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel

³ Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

⁴ European Molecular Biology Laboratory, c/o DESY, 22607 Hamburg, Germany

Key words: DHDPS; dihydrodipicolinate synthase; NAL subfamily; N-acetyl neuraminatase; 2-keto-3-deoxy gluconate aldolase.

INTRODUCTION

The incorporation of genetic material from other organisms into bacterial genomes potentiates novel forms of drug resistance or pathogenicity. One method of incorporation is through infection by a bacteriophage, which enters the lysogenic life cycle. Excision of the phage genome as a result of induction may include some of the host's genome, and subsequent infection and lysogeny in other bacteria would incorporate this non-phage DNA into these hosts. Loss of ability to excise itself from the host leads to the permanent addition in the host genome of the phage genes and other genetic material that may have accumulated during previous lysogenic phases. In this project, targets were chosen initially from proteins encoded by prophages present in *E. coli* K12 based on possible similarities to known virulence factors from pathogenic bacteria, whose three-dimensional structures are unknown. One such target was the *yagE* gene located in the *E. coli* K12 genome that is part of prophage CP4-6¹ encoding a 33-kDa putative dihydrodipicolinate synthase (DHDPS)-like protein (UniProtKB/Swiss-prot: P75682). The DHDPS-like domain belongs to the N-acetyl neuraminatase (NAL) subfamily² containing an eightfold α/β barrel (TIM barrel) with a small C-terminal α -helical domain (Inter pro: IPR005263). The DHDPS domain is present in a wide variety of enzymes. For example, NAL, which is involved in sialic acid metabolism, trans-*o*-hydroxybenzylidene pyruvate hydratase aldolase (HBPHA) in naphthalene degradation, D-4-deoxy-5-oxoglucuronate dehydratase (DOGDH) in glucuronate metabolism, 2-keto-3-deoxy gluconate aldolase (KDGA) in the Entner-Duodoroff pathway, and dihydrodipicolinate synthase (DHDPS) in lysine biosynthesis. All these enzymes contain the DHDPS-like domain and the small C-terminal α -helical region [Fig. 1(A)].

Of all the TIM barrel protein folds described so far, the most common is the eightfold α/β barrel and all family members are enzymes except narbonin.⁴ Proteins that contain DHDPS-like domains catalyze a wide variety of reactions, all of which involve formation of a Schiff's base intermediate.⁵ A lysine residue in β -strand 6 on the floor of the active site region forms a Schiff's base with one of the substrates,⁶ which in most of these enzymes is pyruvate. Substrate promiscuity is often reported for some of its members like KDG aldolase.⁷ The fold is

Grant sponsors: European Commission Sixth Framework Research and Technological Development Program "SPINE2-COMPLEXES" Project; Grant number: 031220; Ministry of Science, Culture and Sport, Israel (MOST) (Indo-Israel Project), Ministry of Science and Industry, India (Biotechnology Department), Bioinformatics Program of DBT, Nalvyco Foundation, Bruce Rosen Foundation, Jean and Julia Goldwurm Memorial Foundation, Kimmelman Center for Biomolecular Structure and Assembly, Benziyo Center for Neuroscience, Israel Structural Proteomics Center (ISPC), Divadol Foundation, Neuman Foundation.

*Correspondence to: Joel L. Sussman, Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel.
E-mail: joel.sussman@weizmann.ac.il
Received 21 August 2007; Revised 18 January 2008; Accepted 30 January 2008
Published online 24 March 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22023

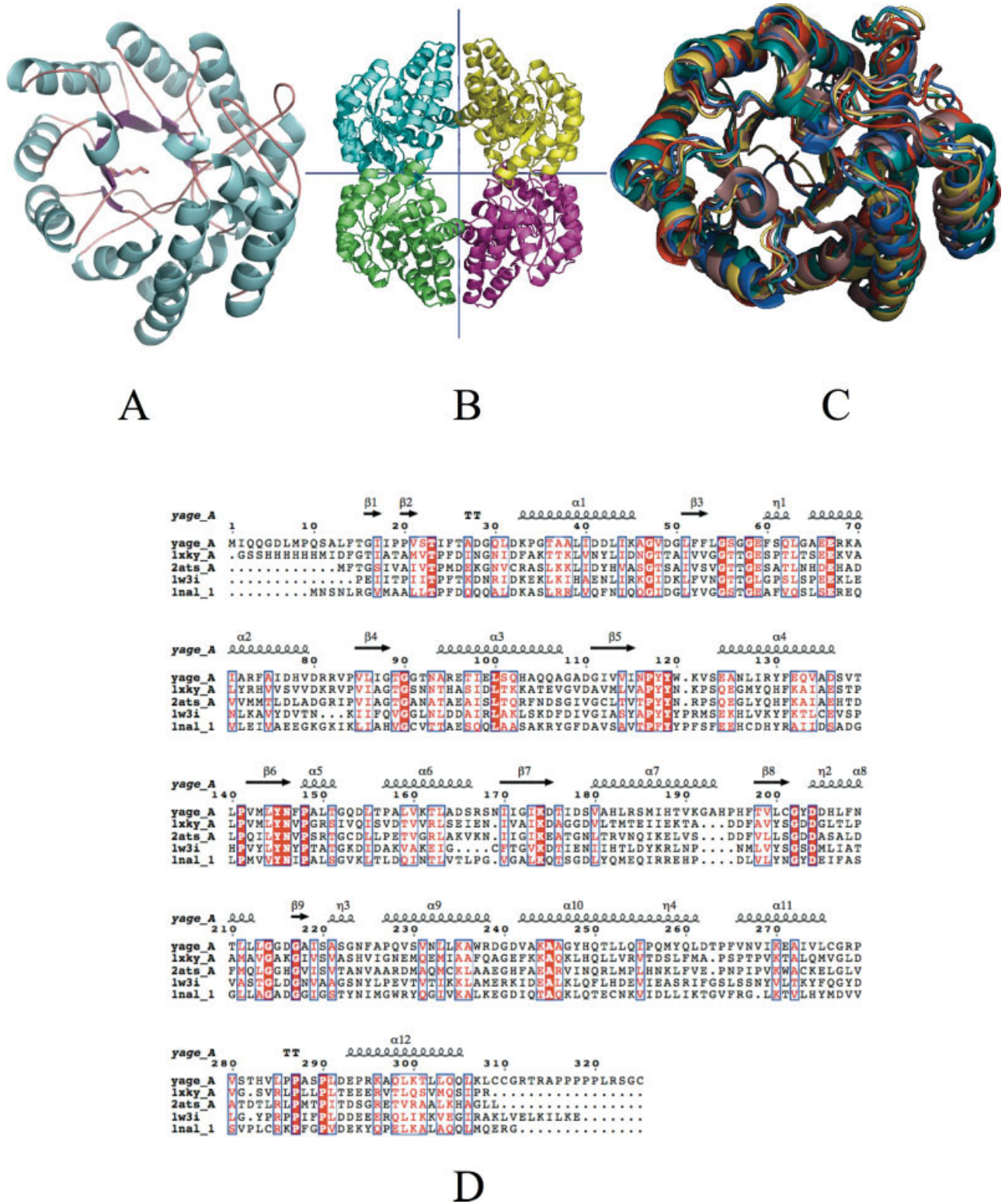


Figure 1

Comparison of YagE structure and sequence with related proteins (A) cartoon representation of YagE monomer. (B) Cartoon representation of YagE tetramer showing the twofold symmetry axes between monomers. Chain A is colored green, Chain B, cyan, Chain C, magenta, and Chain D, yellow. The third axis is perpendicular to the page. (C) Superposition of YagE monomer with EcDHDSP, BaDHDPS, EcNAL, and SsKDGA monomers. (D) Sequence alignment of YagE with EcDHDPS (1XKY), BaDHDPS (2ATS), EcNAL (1NAL), and SsKDGA (1W31) figure generated using ESPript.³

very amenable to directed evolution studies as seen in the case of NAL, where DHDPS activity can be gained by a single point mutation (L142R) while NAL activity is retained.²

Crystal structures of proteins with a DHDPS-like domain have been determined from several organisms. Some of the available PDB models include DHDPS from *E. coli*⁸ (EcDHDPS, PDB: 2ATS), *Thermotoga maritima*⁹ (PDB ID code: 1O5K), *Bacillus anthracis*¹⁰ (BaDHDPS, PDB ID code: 1XKY), *Nicotiana glauca*¹¹ (PDB ID code: 1NAL), *Haemophilus influenzae*¹³ (PDB ID code: 1F6K), and KDG aldolase from *Sulfolobus solfataricus* (SsKDGA, PDB ID code: 1W3I). The sequence similarity of YagE to any of these structures is very low ($\leq 30\%$). Thus the availability of a crystal structure of YagE would help address the question of the functionality and structural relatedness of YagE in relation to the DHDPS-like domain proteins.

METHODS

Cloning

The initial clone of the *yagE* gene was obtained from Genobase, Tokyo, in a pCA24N vector. Subsequently, using 5'-TCACGGTACCCAAGGAGATCTCATGCCG as forward and 5'-CTTAATTCGCGCCGCTTAAGCAAA-GCTTGAGCTGTTGCAG as reverse primers, YagE was recloned into the *KpnI* and *NotI* sites in pET28a-Tev-His, a locally modified pET28a of NovagenTM. The expressed protein had an N-terminal 6X-His-tag with a Tev cleavage site (MGSSHHHHHSAGENLYFQGQQGDLMPQ-SALFT). The His-tag was not removed for crystallization. The C-terminal had 16 additional amino acids (CGR-TRAPPPPLRSGC) before the termination codon.

Expression and purification

YagE expressed as soluble protein when log phase cultures grown at 37°C were induced with 100 μ M isopropyl β -D-1 thiogalactopyranoside (IPTG) and allowed to grow further for expression at 30°C. The cells were then harvested using a Sorvall RC 250 with $\sim 6700g$ for 20 min at 4°C and sonicated in 50 mM tris-HCl pH 7.5, 500 mM sodium chloride, 20 mM imidazole (lysis buffer) containing 1 mM phenylmethylsulphonyl fluoride, 1 U/mL DNase I, 1 mg/mL lysozyme. The lysate was then spun down using an ultra centrifuge at 40,000g for 40 min at 4°C. The supernatant was loaded onto a 5-mL Hi-Trap chelating column (Amersham Pharmacia) and protein was eluted using lysis buffer containing 500 mM imidazole. The eluted fraction was loaded on to a Superdex 200 column (Amersham Pharmacia) and washed with lysis buffer containing 1 mM DTT, and 1 mM EDTA, without imidazole. The peak fractions were then pooled and dia-

lyzed overnight against 50 mM tris-HCl pH 7.5, 50 mM sodium chloride, 10% w/v glycerol, 1 mM DTT, 1 mM EDTA using SnakeskinTM 7 kDa cut off dialysis membrane. The protein was then concentrated to 10 mg/mL and setup for crystallization in microbatch using the IMPAXTM crystallization robot.

A selenomethionine-containing version of YagE was produced from the same *E. coli* BL21 (DE3) clone by growing the bacteria in M9 medium supplemented with 0.5% glucose till 0.6 OD₆₀₀ at 37°C. Then the cells were induced with 100 μ M IPTG and allowed to grow further for expression at 30°C supplemented with 100 mg/L of the amino acids lysine, threonine, and phenylalanine along with 50 mg/L of leucine, iso-leucine, valine, and seleno-methionine.¹⁴ The protein was purified as described for the native protein, and crystallized using the microbatch method.

Crystallization and data collection

Hampton screens Index and PIP (Hampton Research) were used and several conditions produced some crystals within a few minutes of setting up. Large, single crystals grew in conditions which contained 25% PEG 3350 in buffers with acidic pH within 24–48 h. Some of the crystals which were tested gave smeared diffraction spots in the c^* direction. Optimization of crystallization conditions gave well-diffracting orthorhombic crystals after 48 h in hanging-drop vapor diffusion method growing from 100 mM HEPES pH 6.5, 200 mM magnesium chloride, 25% PEG 3350. A full X-ray data set to 2.0 Å resolution was collected using an RU-H3R rotating anode with an RAXIS IV⁺⁺ detector (see Table I).

Initial attempts to solve the structure using molecular replacement (MOLREP) were unsuccessful, due to YagE's low sequence similarity to known PDB models. Therefore, a seleno-methionine containing protein was produced as described earlier. Well-diffracting monoclinic crystals were obtained from 100 mM bis-tris pH 5.5, 200 mM magnesium chloride, and 12% PEG 3350. Crystal diffracted to 2.0 Å resolution and a MAD data set was collected at ID14-4 beamline at ESRF, Grenoble, France (see Table I).

MAD phasing and structure solution

PHENIX (Solve/Resolve)^{15,16} was used to find the positions of the selenium atoms, calculate initial phases, and carry out density modification for phase refinement. The chain of one subunit was traced manually using COOT,¹⁷ starting from residue numbers 30 to 327 numbered as per the ORF in the clone. The other monomers were found using this as a search model (labeled C) in MOLREP¹⁸ from the CCP4 program suite.¹⁹ Two monomers (labeled A and B) were correctly oriented with respect to the electron density map, whereas the third (labeled D) was badly oriented. To locate subunit

Table I

Data Collection and Refinement Statistics

	Native	Seleno-methionine		
Data collection				
X ray source	RU-H3R		ID14-4 ESRF	
Space group	P2 ₁ 2 ₁ 2		P2 ₁	
Cell dimensions				
<i>a, b, c</i> (Å)	141.6, 156.4, 55.8		56.3, 149.8, 78.9	
α, β, γ (°)	90, 90, 90		90, 108.4, 90	
Molecules/AU	4		4	
Solvent content (%)	41.3		42.4	
		Peak	Inflection	Remote
Wavelength (Å)	1.54178	0.9794	0.9796	0.9762
Distance (mm)	160	320		
Exposure time	15 min	1 s		
Resolution (Å)	50–2.09 (2.16–2.09)	50–2.15 (2.23–2.15)	50–2.15 (2.23–2.15)	50–2.14 (2.22–2.14)
<i>R</i> _{sym} (%)	13.5 (56.3)	6.4 (14.0)	6.0 (16.6)	6.3 (19.0)
Reflections	73,983	67,391	67,619	68,585
Completeness (%)	99.3 (95.1)	99.9 (99.7)	99.9 (99.8)	99.8 (99.4)
Redundancy	6.7 (5.8)	7.6 (7.4)	3.7 (3.6)	3.8 (3.7)
Refinement				
Resolution (Å)	19.92–2.20 (2.26–2.20)	36.4–2.15 (2.20–2.15)		
Protein atoms	9,052	9,052		
Solvent atoms	464	649		
No. of reflections	60,553	63,872		
<i>R</i> _{work} / <i>R</i> _{free} (%)	18.0/23.7 (20.8/30.1)	16.6/21.8 (17.1/24.9)		

Values in parentheses are for highest resolution shells. R_{sym} or $R_{\text{merge}} = \sum_{\text{hkl}} \sum_j |I_{\text{hkl}}(j) - \langle I_{\text{hkl}} \rangle| / \sum_{\text{hkl}} \sum_j I_{\text{hkl}}$. $R_{\text{work}} = \sum |F_o| - |F_c| / \sum |F_o|$.

D, subunit A of the AB dimer was overlaid on C using the graphics program PyMOL,²⁰ which placed subunit B accurately in the density for D.

The peak data with resolution range 50–2.1 Å were taken for the structure refinement. Five percent of the data chosen randomly were set aside for *R*_{free} calculations. A simple rigid body refinement of the subunits, followed by refinement with REFMAC5²¹ using tight main-chain, and loose side-chain noncrystallographic symmetry (NCS) restraints. The likelihood weighted, and likelihood gradient maps were used for visual inspection, manual rebuilding between refinement cycles, and for the assignment of water molecules.

The refined structure of YagE tetramer from the Se-Met crystal was used to solve the orthorhombic native crystal data by molecular replacement using MOLREP. Refinement of the orthorhombic structure was carried out using REFMAC5 applying tight main-chain and loose-side chain NCS restraints throughout the refinement, since relaxation of these restraints did not improve *R*_{free}. Adjustments to the structure, including addition of water molecules continued until the *R*_{work} and *R*_{free} had converged to 0.20 and 0.27, respectively. Two final cycles of refinement, where each chain was defined as a separate TLS group, significantly reduced the *R*_{work} and *R*_{free} to their final values as given in Table I. Atomic coordinates and structure factors for the orthorhombic and monoclinic crystal forms of YagE were deposited in the Protein Data Bank under PDB id code 2V8Z and 2V9D, respectively.

Automated structure solution

In parallel to the seleno-methionine derivative, we used automated molecular replacement (BALBES)²² and model building software (ARP/wARP)²³ to solve the structure using the native orthorhombic data set alone.

Seven candidate search models for molecular replacement were automatically selected with BALBES: the monomer, dimer, and tetramer of BaDHDPS (PDB ID code 1XKY, 30% sequence identity), the monomer of EcDHDPS (PDB ID code 2ATS, 28% identity), and the monomer, dimer, and tetramer of *Mycobacterium tuberculosis* DapA (PDB ID code 1XXX, 28% identity). All searches with monomers gave a solution with 2ATS having provided the best result. The solution has been refined with REFMAC5, using tight NCS restraints spanning the whole tetramer and the *R*_{work} and *R*_{free} factors reduced to 0.39 and 0.47. The resulting phases were input to ARP/wARP which then automatically refined and built the model comprising 1145 traced residues (split into 26 fragments), 1043 of which were unambiguously assigned to the amino acid sequence. The automated structure determination took only a few hours on a desktop computer.

Structure analysis

The structures were analyzed using COOT and Biosym (InsightII). The intermolecular and intersubunit interactions in the P2₁ and P2₁2₁2 structures were compared using the PISA server.²⁴

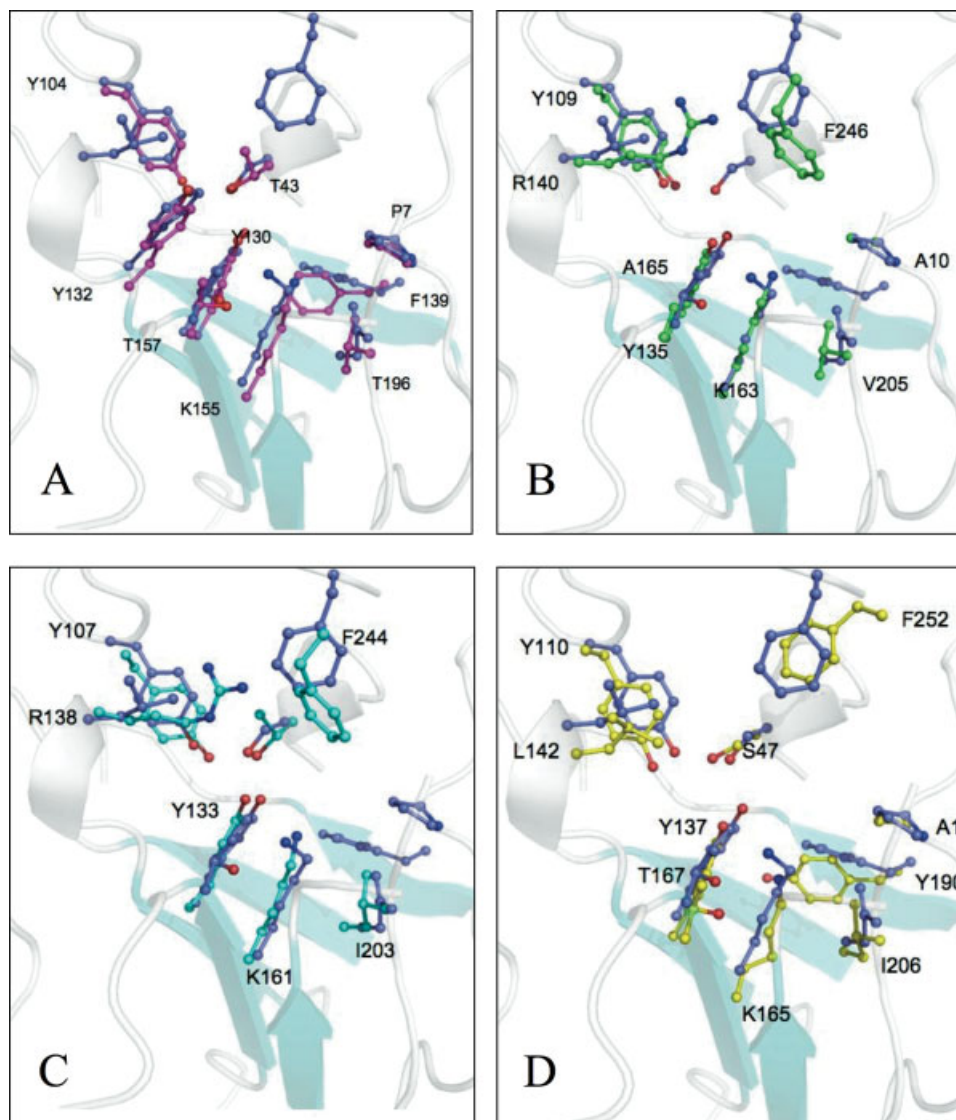


Figure 2

Superposition of active site region residues of YagE shown in blue with (A) 1W3I shown in magenta, (B) 1XKY shown in green, (C) 2ATS shown in cyan, and (D) 1NAL shown in yellow, with amino-acids labels are from the respective PDB structures.

The PDB co-ordinates of EcDHDPS (PDB ID code: 2ATS), EcNAL (PDB ID code: 1NAL), SsKDGA (PDB ID code: 1W3I), and BaDHDPS (PDB ID code: 1XKY) were used. Pairwise structural alignments with YagE were done using the align function in PyMOL which performs a sequence alignment followed by a structural alignment, with additional cycles of refinement, if needed, to reject structural outliers [Fig. 1(C)]. This helped to identify the similarities in the active site and the conserved catalytic lysine. The progressive multiple sequence alignment using clustalW was in agreement with the structure-based sequence alignment [Fig. 1(D)] generated using the align function.

RESULTS AND DISCUSSION

Overall structure and packing

The atomic co-ordinates for the purification tag Tev cleavage site and first 11 amino acids of the YagE protein were not observed in the final structure. In both crystal forms, the asymmetric unit consists of the same tetramer of YagE with 222 symmetry [Fig. 1(B)].⁸ The interface between subunits A and B (or C and D) is more extensive (1660 Å² and eight salt bridges) than the AC or BD interface (990 Å², and four salt bridges). The inter-subunit molecular interfaces are very similar to those of EcDHDPS, BaDHDPS, EcNAL, and SsKDGA, even

though they all crystallized in different space groups and have very low sequence identity, suggesting that the tetrameric arrangement of YagE is biologically relevant.

Each subunit displays the same fold as the other members of the NAL subfamily: an eightfold α/β TIM barrel along with three additional helices in the C-terminal end. Even though YagE has only 23, 24, and 27% sequence identity with EcNAL, SsKDGA, EcDHDPS, respectively, their monomeric structures are all very similar, with the β -strands being highly superposable [Fig. 1(C)]. The YagE dimer is also similar to other members of the NAL subfamily and can be superposed with a root-mean-square deviation ranging from 2 to 2.6 \AA^2 for all atoms.

Comparison with automated molecular replacement structure

The overall structure of the BALBES/ARP/wARP generated protein is almost identical to the structure of the orthorhombic form derived from the monoclinic MAD data. The automated chain tracing almost completed the tetrameric structure from just the native orthorhombic dataset by iteratively searching the database for probable structural homolog candidates at a very low sequence identity of 28%. The final structure traced by ARP/wARP had an R factor of 0.20 and R_{free} of 0.25. Each chain was of variable length, with Chain A starting at residue 31 and Chain C starting at residue 41. The most complete chain of all was Chain B, which was built without any breaks from residue 12 to 314. There were three large stretches missing in the whole tetramer in the middle of chains: in Chain A, residues 73–81, and in Chain C, residues 50–64 and 83–91. Analysis of the final refined model showed that these areas had higher than average atomic displacement parameters (ADPs), also known as temperature factors. The average main-chain ADP for chain A was 15.3 \AA^2 (maximum 26.6 \AA^2) and for chain C 17.4 \AA^2 (maximum 31.9 \AA^2). The mean main-chain ADP for the stretch A73–A81 was 23.5 \AA^2 (maximum 26.6 \AA^2), that for C50–C64 was 23.3 \AA^2 (maximum 26.7 \AA^2), and that for C83–C91 was 24.9 \AA^2 (maximum 31.9 \AA^2). Thus, aside from the elevated ADPs, two of the three sections contain at least one main-chain atom with the highest temperature factor for the whole chain.

The ARP/wARP traced structure was perfectly superposable with the final refined model of native and Sernet derivative solved structures with an rmsd of 0.12 \AA (all atoms).

Comparison of YagE with other members of NAL subfamily

Members of the NAL subfamily of proteins are known to have very similar active sites and a single amino acid change can drastically alter the function. This is seen, for example, in the case of NAL, which can gain DHDPS

activity by altering a leucine to arginine at position 142. The possible active site region of YagE shows closest sequence resemblance to the active site of KDG aldolase from SsKDGA [Fig. 2(A)] and NAL from EcNAL [Fig. 2(D)], suggesting that, possibly, this protein can perform either of these functions. Although the active site of EcDHDPS [Fig. 2(C)] and BaDHDPS [Fig. 2(B)] shows similarities, the important residue that differentiates between NAL and DHDPS, namely Leu142 (in 1NAL), is also present in YagE at that particular position suggesting that the protein might perform more of a NAL-related than DHDPS-related function.

In summary, the high-resolution structure of YagE provides a first view of this protein, which is identified here as a putative member of the well-characterized NAL subfamily of proteins. This analysis has suggested two possible functions based on sequence and structure alignment with known NAL subfamily members. Even though the exact molecular function for the protein is not immediately evident, the structure provides a framework to deduce and assay molecular function based on clustered conserved residues and the putative active site. The automated model building almost completed the structure from just the native dataset, with an RMS deviation after refinement of 0.12 \AA . This is a remarkable result considering the low sequence identity (28%) of the homolog candidate to YagE and implies that, it may be possible, in many more cases to solve structures via molecular replacement, even when the apparent sequence identity is quite low.

ACKNOWLEDGMENTS

JLS is the incumbent of the Morton and Gladys Pickman Chair of Structural Biology.

REFERENCES

- Blattner FR, Plunkett G, III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. The complete genome sequence of *Escherichia coli* K-12. *Science* 1997;277:1453–1474.
- Joerger AC, Mayer S, Fersht AR. Mimicking natural evolution in vitro: an *N*-acetylneuraminase lyase mutant with an increased dihydrodipicolinate synthase activity. *Proc Natl Acad Sci USA* 2003;100:5694–5699.
- Gouet P, Courcelle E, Stuart DI, Metz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 1999;15:305–308.
- Hennig M, Schlesier B, Dauter Z, Pfeffer S, Betzel C, Hohn WE, Wilson KS. A TIM barrel protein without enzymatic activity? Crystal-structure of narbonin at 1.8 \AA resolution. *FEBS Lett* 1992;306:80–84.
- Lawrence MC, Barbosa JA, Smith BJ, Hall NE, Pilling PA, Ooi HC, Marcuccio SM. Structure and mechanism of a sub-family of enzymes related to *N*-acetylneuraminase lyase. *J Mol Biol* 1997;266:381–399.

6. Laber B, Gomis-Ruth FX, Romao MJ, Huber R. *Escherichia coli* dihydrodipicolinate synthase. Identification of the active site and crystallization. *Biochem J* 1992;288 (Part 2):691–695.
7. Theodossis A, Walden H, Westwick EJ, Connaris H, Lamble HJ, Hough DW, Danson MJ, Taylor GL. The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in *Sulfolobus solfataricus*. *J Biol Chem* 2004;279:43886–43892.
8. Mirwaldt C, Korndorfer I, Huber R. The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 Å resolution. *J Mol Biol* 1995;246:227–239.
9. Joint Centre for Structural Genomics. Crystal structure of dihydrodipicolinate synthase from *Thermotoga maritima* at 1.8 Å resolution, unpublished PDB code 1o5k.
10. Blagova E, Levdikov V, Milioti N, Fogg MJ, Kalliomaa AK, Brannigan JA, Wilson KS, Wilkinson AJ. Crystal structure of dihydrodipicolinate synthase (BA3935) from *Bacillus anthracis* at 1.94 Å resolution. *Proteins* 2006;62:297–301.
11. Blickling S, Beisel HG, Bozic D, Knablein J, Laber B, Huber R. Structure of dihydrodipicolinate synthase of *Nicotiana sylvestris* reveals novel quaternary structure. *J Mol Biol* 1997;274:608–621.
12. IZard T, Lawrence MC, Malby RL, Lilley GG, Colman PM. The three-dimensional structure of *N*-acetylneuraminidase from *Escherichia coli*. *Structure* 1994;2:361–369.
13. Barbosa JA, Smith BJ, DeGori R, Ooi HC, Marcuccio SM, Campi EM, Jackson WR, Brossmer R, Sommer M, Lawrence MC. Active site modulation in the *N*-acetylneuraminidase sub-family as revealed by the structure of the inhibitor-complexed *Haemophilus influenzae* enzyme. *J Mol Biol* 2000;303:405–421.
14. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 1993;229:105–124.
15. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 1999;55 (Part 4):849–861.
16. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 2000;56 (Part 8):965–972.
17. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60 (Part 1, Part 12):2126–2132.
18. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* 2000;56 (Part 12):1622–1624.
19. Potterton E, McNicholas S, Krissinel E, Cowtan K, Noble M. The CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* 2002;58 (Part 11):1955–1957.
20. DeLano LW. The PyMOL molecular graphics system 0.1. San Carlos, CA: DeLano Scientific LLC; 2006.
21. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53(Part 3):240–255.
22. Long F, Vagin AA, Young P, Murshudov GN. BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* 2008;64:125–132.
23. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 1999;6:458–463.
24. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007;372:774–797.