WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Solution NMR structure of *Escherichia coli* ytfP expands the structural coverage of the UPF0131 protein domain family

James M. Aramini,[1,2]* Yuanpeng J. Huang,[1,2] G.V.T. Swapna,[1,2] John R. Cort,[2,3] P.K. Rajan,[1,2] Rong Xiao,[1,2] Ritu Shastry,[1,2] Thomas B. Acton,[1,2] Jinfeng Liu,[2,4] Burkhard Rost,[2,4] Michael A. Kennedy,[2,3] and Gaetano T. Montelione[1,2,5]*

[1] Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854

[2] Northeast Structural Genomics Consortium, Rutgers University, Piscataway, New Jersey 08854

[3] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

[4] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032

[5] Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

INTRODUCTION

The 113-residue protein from the open reading frame *ytfP* of *Escherichia coli* [SWISS-PROT ID: YTFP_ECOLI; NESG target ID: ER111; referred to hereafter as ytfP] is a member of the UPF0131 Pfam-A protein domain family (Pfam identifier: PF03674). *E. coli* genomics efforts have established that the *ytfP* gene shares an operon with two other open reading frames, *ytfM* and *ytfN*, and overexpression of each of these genes is highly toxic to the cell.[1] The largely uncharacterized UPF0131 protein domain family includes representatives from all three kingdoms, archaea, eubacteria, and eukaryotes, with over half originating from bacterial sources. At least one member of this family, BtrG from *Bacillus circulans* (SWISS-PROT ID: Q9F1Z7_BACCI) is thought to play a critical role in the biosynthesis of butirosin, a class of aminoglycoside antibiotics that features a distinctive (2*S*)-4-amino-2-hydroxybutyrate (AHBA) moiety.[2,3] It is thought that BtrG acts in the final stages of butirosin synthesis by cleaving the γ-L-Glu protecting group to yield the final product.[3] The function of the *E. coli* homologue ytfP has not yet been characterized, but it probably has similar biochemical function in antibiotic biosynthesis.

Although pair wise sequence alignments within the family vary extensively in degree of sequence identity, clusters of highly conserved residues are present in the sequence. A multiple sequence alignment of *E. coli* ytfP with a selected subset of UPF0131 protein domains is shown in Figure 1(A). X-ray crystal structures of UPF0131 proteins from the archeaon *Pyrococcus horikoshii* (SWISS-PROT ID: Y828_PYRHO; PDB ID: 1V30[6]; ~31% sequence identity with ytfP) and *Mus musculus* (SWISS-PROT ID: Q923B0_MOUSE; PDB ID: 1VKB[7]; ~20% sequence identity with ytfP) have been recently described. In this article, we present the solution NMR structure of *E. coli* ytfP and compare it with these two recent crystal structures of UPF0131 protein domain family homologues. This *E. coli* ytfP structure represents the first NMR structure from this family, as well as the first structure of a bacterial representative.

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: Protein Structure Initiative of the National Institutes of Health; Grant number: U54-GM074958.

*Correspondence to: James M. Aramini or Gaetano T. Montelione, CABM-Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854.

E-mail: jma@cabm.rutgers.edu or guy@cabm.rutgers.edu

Received 21 January 2007; Accepted 1 February 2007

Published online 21 May 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.21450

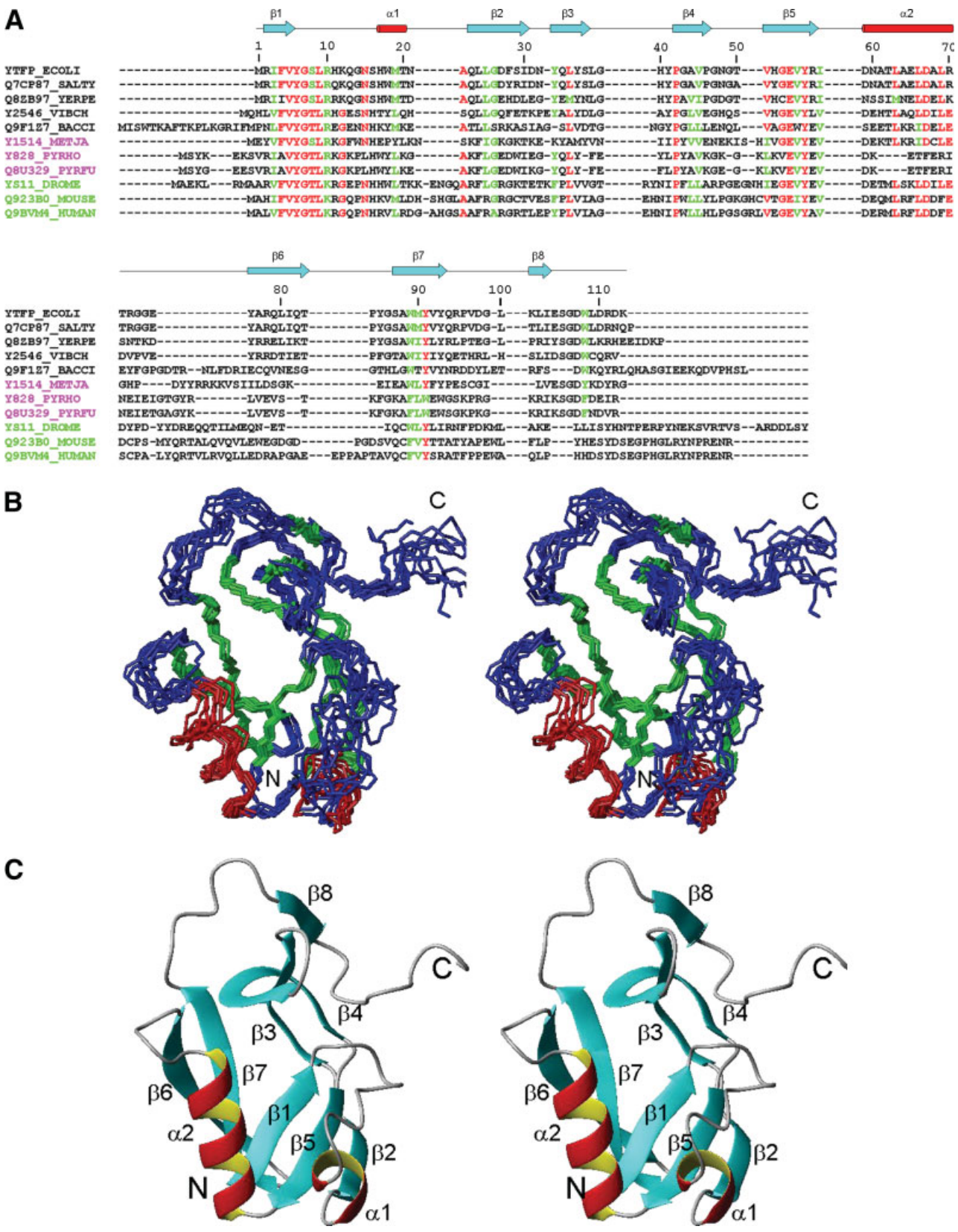

Figure 1

*(**A**) A subset of the multiple sequence alignment of the entire UPF0131 protein domain family (Pfam release 20.0) aligned using Clustal X.*[4] *Representatives from bacteria, archaea, and eukaryotes, indicated by their Swiss-Prot IDs, are listed in black, magenta, and green, respectively. Amino acid residues identical or similar in 67% of the entire family are shown in red and green, respectively; conserved residues were colored using the BOXSHADE server. The sequence numbering for yftP from* E. coli *and the secondary structural elements found in its NMR structure described in this paper (1XHS) are shown above the alignment. (**B**) Stereoview of the solution structure of ytfP showing the backbone atom superposition of the final ensemble of 10 conformers representing the solution structure of ytfP; β-strand elements are shown in green, α-helices are in red. (**C**) Stereoview of the ribbon representation of a representative conformer (lowest CNS energy) from the ensemble generated using MOLMOL.*[5] *The secondary structural elements are labeled.*

MATERIALS AND METHODS

Uniformly ^{13}C,^{15}N-enriched *E. coli* ytfP was cloned, expressed, and purified following standard protocols of the NESG consortium.[8] Briefly, the full-length *ytfP* gene from *E. coli* was cloned into a pET21d (Novagen) derivative, yielding the plasmid pER111-21. The resulting ytfP open reading frame contains an additional eight nonnative residues at the C-terminus (LEHHHHHH) of the protein. *E. coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pER111-21, and cultured in MJ minimal medium[9] containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as the sole nitrogen and carbon sources. Initial growth was carried out at 37°C until the OD_{600} of the culture reached ≈ 1.0 units, followed by induction of protein expression by the addition of IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 1 m*M*. The cells were harvested after 5 h by centrifugation and lysed by sonication. *U*-^{13}C,^{15}N ytfP was purified in a two step protocol consisting of Ni-NTA affinity column (Qiagen) and gel filtration column (HiLoad 26/60 Sephadex 75, Amersham Pharmacia Biotech) chromatography. The final yield of pure protein (>97% homogeneity by SDS-PAGE; 14.7 kDa by MALDI-TOF mass spectrometry) was ~21 mg/L. Samples of *U*-^{13}C,^{15}N ytfP for NMR spectroscopy were concentrated by ultracentrifugtion to a concentration of 1.3 m*M* in 95% H_2O/5% D_2O or 100% D_2O solution containing 20 m*M* ammonium acetate, 100 m*M* NaCl, 10 m*M* DTT, 5 m*M* $CaCl_2$ at pH 5.5. An NMR sample of ytfP with 5% ^{13}C and 100% ^{15}N labeling for selected experiments (see below) was prepared and purified in a similar fashion.

All NMR data were collected at 20°C on four-channel Varian INOVA 500, 600, 750, and 800 MHz NMR spectrometers, processed with NMRPipe 2.3,[10] and visualized using SPARKY 3.91.[11] Complete ^1H, ^{13}C, and ^{15}N resonance assignment of *E. coli* ytfP have been determined and deposited in the BioMagResDB (BMRB accession number 6448).[12] Stereospecific isopropyl methyl assignments for all Val and Leu residues were deduced from characteristic cross-peak fine structures in a high resolution 2D ^1H-^{13}C HSQC spectrum of 5%-^{13}C, 100%-^{15}N ytfP.[13] Resonance assignments were validated using the Assignment Validation Suite (AVS) software package.[14]

NOE distance constraints were derived from 3D ^{15}N-edited NOESY (τ_m = 80 ms), 3D ^{13}C-edited aliphatic and aromatic NOESYs (τ_m = 80 ms), and 4D ^{13}C/^{13}C-NOESY in 100% ^2H_2O (τ_m = 100 ms). Three-bond $^3J(H^N$-$H^\alpha)$ scalar couplings were determined from a 3D HNHA spectrum of *U*-^{13}C,^{15}N ytfP.[15] Slowly exchanging backbone amide protons were identified by dissolving lyophilized *U*-^{13}C,^{15}N ytfP in 100% ^2H_2O and monitoring the decay of the ^1H-^{15}N HSQC signal over time. ^1H-^{15}N heteronuclear NOEs were measured on 5% ^{13}C, 100% ^{15}N-enriched ytfP using a standard gradient sensitivity-enhanced 2D heteronuclear NOE experiment.[16]

Structure calculations were performed using the program AutoStructure 2.1.0,[17,18] interfaced with XPLOR-NIH 2.0.6.[19] The input for the AutoStructure program consisted of a resonance assignment list, manually edited peak lists with intensities for the 3D ^{15}N-edited, 3D ^{13}C-edited, and 4D ^{13}C/^{13}C NOESY spectra, $^3J(H^N$-$H^\alpha)$ values, broad ϕ,ψ angle constraints (±40° and ±50°, respectively) derived from chemical shift data using the program TALOS,[20] and slow amide hydrogen exchange data; TALOS dihedral constraints were used only for residues with confidence scores of 10. In each AutoStructure cycle, structure calculations were performed using an XPLOR-NIH 2.0.6 simulated annealing refinement protocol.[19] The best 10 of 56 structures (lowest energy) from the final cycle of AutoStructure were refined by restrained molecular dynamics in explicit water using CNS 1.1[21] based on published procedures.[22]

The global goodness-of-fit of the final ensemble of structures with the NOESY peak list data was assessed using a set of RPF-scores, a formalism based on information retrieval statistics.[17] Briefly, recall measures the percentage of NOESY peaks that are consistent with the interproton distances in the 3D structure, Precision measures the percentage of close distance proton pairs (<5 Å) in the 3D structure whose back-calculated NOESY cross peaks are observed in the NOESY peak lists, F-measure is the overall performance score calculated from the Recall and Precision, and Discriminating Power (DP)-score is a normalized F-measure that reflects how the query structure is distinguished from the freely-rotating chain model. In practice, DP-scores and F-measures greater than 0.7 and 0.9, respectively, indicate good structure/spectral data quality and accuracy. RPF-scores for the final ensemble of CNS refined ytfP structures were calculated against the raw 3D NOESY peak list data only. Global structure quality factors for the final ensemble of ytfP structures were determined using the PSVS 1.0 software package,[23] which outputs Verify3D,[24] Prosa II,[25] PROCHECK,[26] and MolProbity[27] raw and statistical *Z*-scores. The final ensemble of structures (excluding the C-terminal tag whose conformation is not well-defined by these data) have been deposited in the Protein Data Bank (PDB ID 1XHS).

RESULTS AND DISCUSSION

E. coli ytfP is comprised of eight β strands (β1, 2–5; β2, 23–30; β3, 33–37; β4, 42–46; β5, 51–57; β6, 76–83; β7, 86–93; β8, 101–103) and two α helices (α1, 17–20; α2, 60–70) arranged in a β-α-β-β-β-β-α-β-β-β topology. Stereoimages of the superimposed final ensemble and ribbon diagram of a representative structure are shown in Figure 1(B,C), and structural statistics are listed in Table I. In the core of the protein, five β-strands (β1, β2, β5, β6, and β7) come together to form an antiparal-

Table I
Summary of NMR Data and Structural Statistics for E. coli *ytfP*[a]

Completeness of resonance assignments[b]		
Backbone	97%	
Side chain	91%	
Aromatic	90%	
Stereospecific methyl	100%	
Conformationally-restricting constraints[c]		
Distance constraints		
Total	1687	
Intra-residue ($i = j$)	197	
Sequential ($\|i-j\| = 1$)	478	
Medium range ($1 < \|i-j\| \leq 5$)	257	
Long range ($\|i-j\| > 5$)	755	
Distance constraints per residue	14.9	
Dihedral angle constraints	211	
Hydrogen bond constraints		
Total	68	
Long range ($\|i-j\| > 5$)	50	
Number of constraints per residue	17.4	
Number of long range constraints per residue	7.1	
Residual constraint violations[c]		
Average number of distance violations per structure		
0.1–0.2 Å	7.5	
0.2–0.5 Å	3.7	
>0.5 Å	0.6	
Average RMS distance violation/constraint (Å)	0.02	
Maximum distance violation (Å)	1.06	
Average number of dihedral angle violations per structure		
1–10°	9.3	
>10°	0	
Average RMS dihedral angle violation/constraint (degree)	0.58	
Maximum dihedral angle violation (degree)	9.10	
RMSD from average coordinates (Å)[c,d]		
Backbone atoms	0.7	
Heavy atoms	1.2	
Ramachandran plot statistics[c,d]		
Most favored regions (%)	89.9	
Additional allowed regions (%)	10.1	
Generously allowed (%)	0.0	
Disallowed regions (%)	0.0	
Global quality scores[c]	Raw	*Z*-score
Verify3D	0.42	−0.64
ProsaII	0.45	−0.83
Procheck(phi-psi)[d]	−0.53	−1.77
Procheck(all)[d]	−0.49	−2.90
Molprobity clash	29.67	−3.57
RPF Scores[e]		
Recall	0.96	
Precision	0.87	
F-measure	0.91	
DP-score	0.79	

[a]Structural statistics were computed for the ensemble of 10 structures refined in explicit water using CNS 1.1.[21]
[b]Computed using AVS software[14] from the expected number of peaks, excluding: highly exchangeable protons (N-terminal, Lys, and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, nonprotonated aromatic carbons, and the C-terminal tag.
[c]Calculated for final ensemble excluding the C-terminal tag using PSVS 1.0 program.[23] Average distance violations were calculated using the sum over r^{-6}.
[d]Ordered residue ranges [S(phi) + S(psi) > 1.8]: 1–5, 9–10, 17–18, 23–38, 41–70, 76–85, 88–93, 101–104, 106–108.
[e]RPF scores[17] reflecting the goodness-of-fit of the final ensemble of structures (including C-terminal tag) to the NMR data.

lel beta sheet. Interestingly, the region spanning strands β2 through β5 features a peculiar cross-over point; this unusual motif is also present in the crystal structures of the two homologues of *E. coli* ytfP (see below). One face of the protein features a large cavity (volume ∼2440 Å^3 and average depth ∼9.6 Å)[28] formed by the central beta sheet flanked by the two α helices, and capped by strands β3 and β4. This cavity is rich with basic and hydrophobic residues. The electrostatic surface potential diagram[29] of ytfP [Fig. 2(A)] shows a largely positively charged surface within and lining the cavity. Moreover, a ConSurf analysis[30] reveals that several residues in the cavity are highly conserved across the UPF0131 protein domain family, in contrast to the opposite face of the molecule [Fig. 2(B)]. Taken together, these results strongly suggest that the face of the protein comprising the cavity plays a key role in the biochemical function of ytfP.

The solution NMR structure of ytfP is similar to the recent X-ray crystal structures of two other UPF0131 protein domain family members, namely the 116-residue Y828_PYRHO protein from *Pyrococcus horikoshii* (1V30)[6] and the 149-residue Q923B0_MOUSE protein from *Mus musculus* (1VKB).[7] A Dali[32] search of the *E. coli* ytfP structure against the Protein Data Bank (PDB) reveals very strong hits for only these two protein structures (Dali *Z*-scores: 1V30, 12.2; 1VKB, 9.6; C^{α} RMSDs: 1V30, 2.6 Å; 1VKB, 3.0 Å), in spite of their moderate to low sequence identity to ytfP (1V30, 31%; 1VKB, 20%); no other protein structure in the PDB exhibits a Dali Z score above 2.0. Ribbon diagrams of the structures of the three homologues are shown in Figure 2(C). Like *E. coli* ytfP (1XHS), structures of both Y828_PYRHO and Q923B0_MOUSE feature a cavity formed by a five stranded antiparallel β-sheet flanked by α-helices, two β strands capping the cleft, and the crossover between strands β2 through β5. However, there are minor differences in topology between the three structures. In Y828_PYRHO, the second helix is split into two and the C-terminus (including the tag) adopts a helical conformation. In Q923B0_MOUSE, a short insert between α1 and β2 folds into a helix, and the final short β-strand in ytfP is replaced by a short helix, followed by a much longer unstructured tail. More importantly, the cavities in both Y828_PYRHO and Q923B0_MOUSE do not possess the positively charged electrostatic surface observed in the ytfP structure (not shown). Specifically, the two basic residues in *E. coli* ytfP, Lys12 and Arg70, are found only in some gamma proteobacteria, and are nonbasic highly conserved residues (primarily, Gly at position 12 and Glu at position 70) in the vast majority of the UPF0131 protein domain family. Thus, we conclude that while these three proteins are distant in terms of sequence, they share a highly unique common fold; we further postulate that specific amino acid changes in the vicinity of the active site of these proteins alter their biochemical specificity. While it is likely that, like the *B. circulans* homologue, *E. coli* ytfP

DOI 10.1002/prot

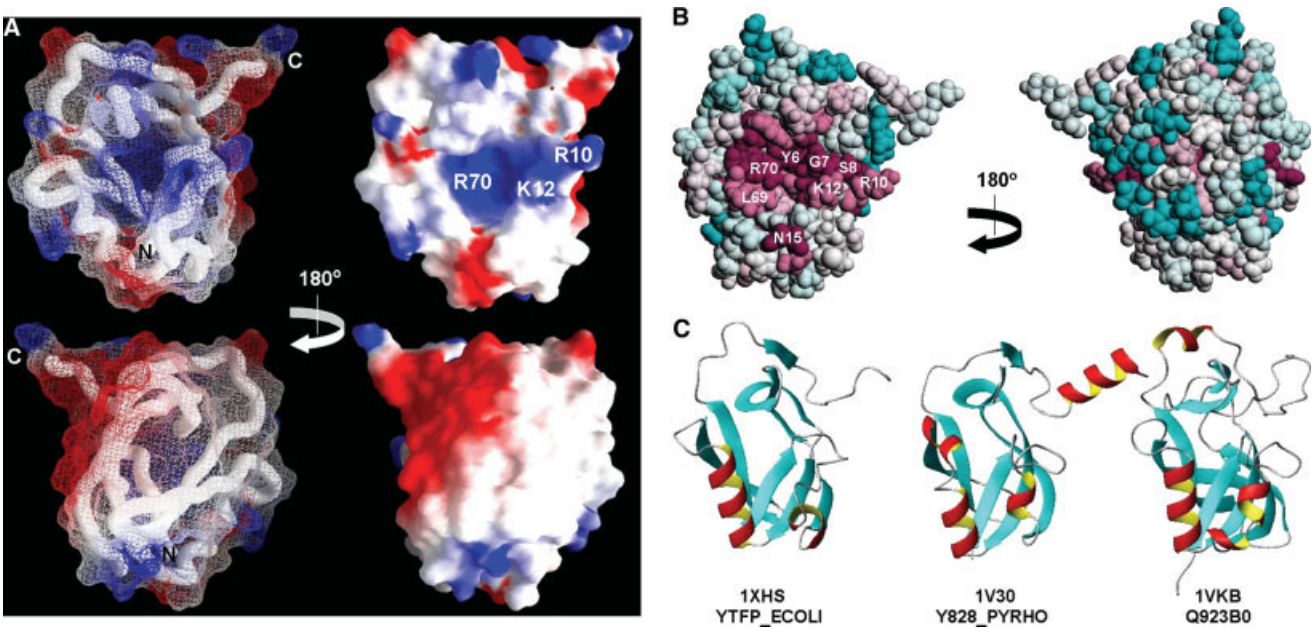

Figure 2

*(**A**) GRASP[29] electrostatic potential surfaces showing the face of ytfP containing the positively charged (blue) cavity (top) and the opposite face of the protein (bottom). (**B**) Two ConSurf[30] images of ytfP based on the multiple sequence alignment of the entire UPF0131 protein domain family. Residue coloring, reflecting the degree of residue conservation over the entire family, ranges from magenta (highly conserved) to cyan (variable). (**C**) Ribbon diagrams of the pairwise structural alignments of the solution structure of* E. coli *ytfP (1XHS; residues 1–113), crystal structure of* P. horikoshii *Y828_PYRHO (1V30[6]; 7–116 plus C-terminal tag), and crystal structure of* M. musculus *Q923B0 (1VKB[7];1–149) using the CE server.[31]*

functions in antibiotic biosynthesis, confirmation of its biological role will require further investigations.

The structure of *E. coli* ytfP presented here increases the number of protein sequences that can be homology modeled with reasonable accuracy, generally considered to be $> \sim$40% sequence identity and PSI-Blast *E* value less than 10^{-10} between target and template.[33–35] Modeling leverage defined as the total number of protein sequences in the UniProt database (release 7.6) for which a portion greater than 50 residues can potentially be modeled by the ytfP structure, and novel modeling leverage, defined as the modeling leverage excluding protein sequences that could be modeled with structures available in the PDB on the date that the ytfP structure was deposited (9/20/04), were determined as defined elsewhere.[35] The total and novel leverage values for ytfP are 121 and 11 protein structures, respectively. Moreover, ytfP provides an additional modeling template for many of these sequences, which can be used to improve modeling accuracy.[34] Hence, additional unique structural information is provided by ytfP that was not available from structures in the PDB on the date ytfP was deposited. While the structures of Q923B0_MOUSE and Y828_PYRHO can be used to homology model subsets of the eukaryotic and archaeal UPF0131 proteins, respectively, the homology modeling coverage space of the *E. coli* ytfP structure presented here encompasses an entirely different region of the phylogenetic tree (see Fig. 3), namely all of the UPF0131 family members from the gamma proteobacteria. This structure should be valuable in studies of structure–function relationships and mechanisms of antibiotic biosynthesis.

ACKNOWLEDGMENTS

We thank J. B. Spencer, M. Inouye, and M. Suzuki for valuable discussions. High field NMR experiments used in the structure determination were acquired in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

NOTE ADDED IN PROOF

Lytle et al. (Acta Crystallogr F 2006;62:490–493) have recently described the solution NMR structure of *Arabidopsis thaliana* protein At5g39720.1 (PDB ID: 2G0Q) which, although it is from a different protein domain family (PF06094) and shares very little sequence similarity with *E. coli* ytfP, exhibits a similar overall 3D structure.

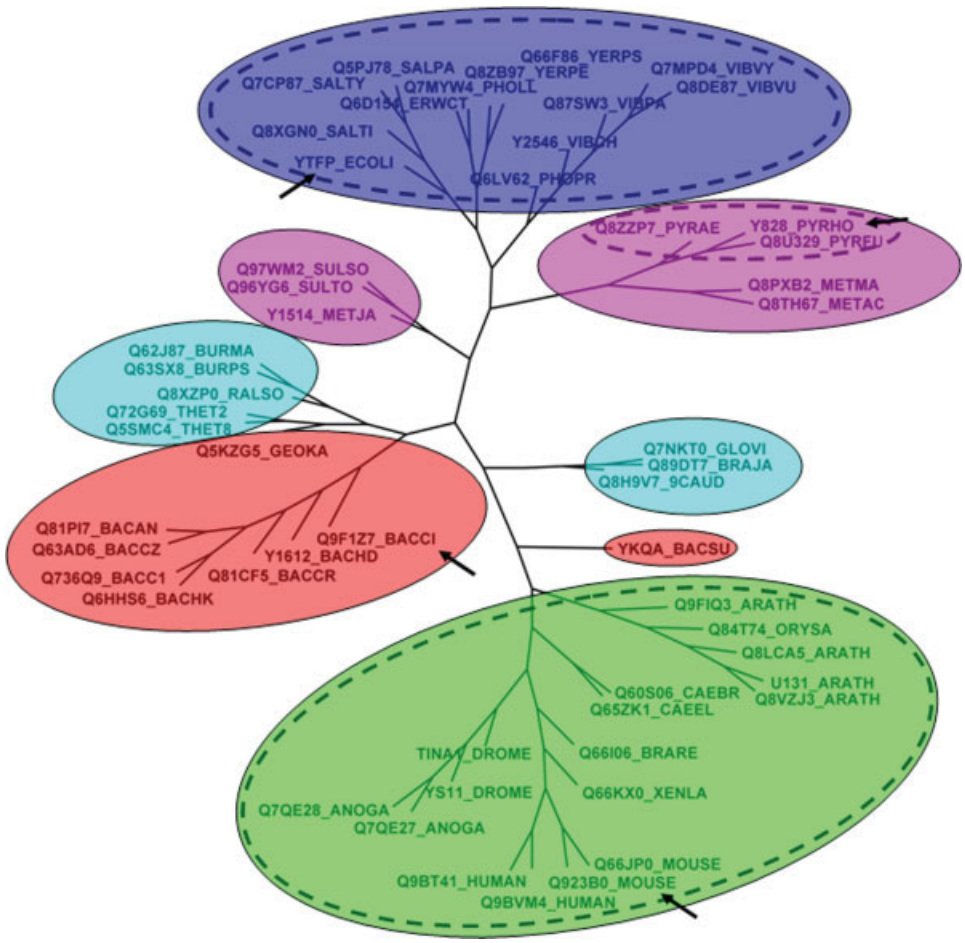

Figure 3

Radial N-J phylogenetic tree for the UPF0131 protein domain family (release 20.0) based upon the ClustalX multiple sequence alignment of the entire family. The tree was generated using the program PhyloDraw.[36] *Regions of the tree are colored on the basis of kingdom or phylum: blue, gamma proteobacteria; red, firmicutes; cyan, other bacteria; magenta, archaea; green, eukaryotes. Arrows denote the positions of genes from* E. coli, P. horikoshii, *and* M. musculus *whose protein structures have been solved, as well as that from* B. circulans, *the lone member of the family whose biological role has been deciphered. Dashed circles encompass members of the family with* $\geq$*35% sequence identity and Blast* $E < 10^{-10}$ *(from NCBI BlastP using UniProt release 6.1) to one of the three family members whose structures have been solved (blue,* E. coli*; magenta,* P. horikoshii*; green,* M. musculus*).*

REFERENCES

1. Mori H, Isono K, Horiuchi T, Miki T. Functional genomics of *Escherichia coli* in Japan. Res Microbiol 2000;151:121–128.
2. Ota Y, Tamegai H, Kudo F, Kuriki H, Koike-Takeshita A, Eguchi T, Kakinuma K. Butirosin-biosynthetic gene cluster from *Bacillus circulans.* J Antibiot (Tokyo) 2000;53:1158–1167.
3. Li Y, Llewellyn NM, Giri R, Huang F, Spencer JB. Biosynthesis of the unique amino acid side chain of butirosin: possible protective-group chemistry in an acyl carrier protein-mediated pathway. Chem Biol 2005;12:665–675.
4. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23:403–405.
5. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.
6. Tajika Y, Sakai N, Tamura T, Yao M, Watanabe N, Tanaka I. Crystal structure of hypothetical protein PH0828 from *Pyrococcus horikoshii.* Proteins 2004;57:862–865.
7. Klock HE, Schwarzenbacher R, Xu Q, McMullan D, Abdubek P, Ambing E, Axelrod H, Biorac T, Canaves JM, Chiu H-J, Deacon AM, DiDonato M, Elsliger M-A, Godzik A, Grittini C, Grzechnik SK, Hale J, Hampton E, Han GW, Haugen J, Hornsby M, Jaroszewski L, Koesema E, Kreusch A, Kuhn P, Miller MD, Moy K, Nigoghossian E, Paulsen J, Quijano K, Reyes R, Rife C, Sims E, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, White A, Wolf G, Hodgson KO, Wooley J, Lesley SA, Wilson IA. Crystal structure of a conserved hypothetical protein (gi: 13879369) from Mouse at 1.90 Å resolution reveals a new fold. Proteins 2005;61:1132–1136.
8. Acton TB, Gunsalus KC, Xiao R, Ma L-C, Aramini JM, Baran MC, Chiang Y-W, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Paranji RK, Shastry R, Shih L-Y,

DOI 10.1002/prot

Swapna GVT, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and protein production platform of the Northeast Structural Genomics Consortium. Methods Enzymol 2005;394:210–243.
9. Jansson M, Li Y-C, Jendeberg L, Anderson S, Montelione GT, Nilsson B. High level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. J Biomol NMR 1996;7:131–141.
10. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
11. Goddard TD, Kneller DG. SPARKY 3. San Fransisco: University of California.
12. Aramini JM, Swapna GVT, Huang YJ, Rajan PK, Xiao R, Shastry R, Acton TB, Cort JR, Kennedy MA, Montelione GT. ^{1}H, ^{13}C, ^{15}N resonance assignments for *Escherichia coli* ytfP, a member of the broadly conserved UPF0131 protein domain family. J Biomol NMR 2005;33:197.
13. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry 1989;28:7510–7516.
14. Moseley HNB, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 2004;28:341–355.
15. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J($H^{N}H^{\alpha}$) coupling constants in ^{15}N-enriched proteins. J Am Chem Soc 1993;115:7772–7777.
16. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay L.E. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 1994;33:5984–6003.
17. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and *F*-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 2005;127:1665–1674.
18. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins 2006;62:587–603.
19. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. J Magn Reson 2003;160:65–73.
20. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–302.
21. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, 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 1998;54:905–921.
22. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50:496–506.
23. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics criteria. Proteins 2007;66:778–795.
24. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164–170.
25. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
26. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–86.
27. Lovell SC, Davis IW, Arendall WB, III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ, ψ, and Cβ deviation. Proteins 2003;50:437–450.
28. Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 2005;33:D266–D268.
29. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
30. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 2003;19:163–164.
31. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Prot Eng 1998;11:739–747.
32. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
33. Rost B. Twilight zone of protein sequence alignments. Prot Eng 1999;12:85–94.
34. Sánchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, Madhusudhan MS, Mirkovic N, Šali A. Protein structure modeling for structural genomics. Nat Struct Biol 2000;7:986–990.
35. Liu J, Montelione GT, Rost B. Novel leverage of structural genomics. Nature Biotechnology, in press.
36. Choi JH, Jung HY, Kim HS, Cho HG. PhyloDraw: a phylogenetic tree drawing system. Bioinformatics 2000;16:1056–1058.