Published in final edited form as:

J Struct Funct Genomics. 2011 March ; 12(1): 27–32. doi:10.1007/s10969-010-9099-2.

Solution NMR and X-ray crystal structures of membrane-associated Lipoprotein-17 domain reveal a novel fold

Rajeswari Mani,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Sergey Vorobiev,

Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, USA

G. V. T. Swapna,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Helen Neely,

Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, USA

Haleema Janjua,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Colleen Ciccocanti,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Rong Xiao,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Thomas B. Acton,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

John K. Everett,

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

John Hunt, and

© Springer Science+Business Media B.V. 2010

Correspondence to: Gaetano T. Montelione, guy@cabm.rutgers.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s10969-010-9099-2) contains supplementary material, which is available to authorized users.

Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, USA

Gaetano T. Montelione

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium (NESG), Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854, USA

Gaetano T. Montelione: guy@cabm.rutgers.edu

Abstract

The conserved Lipoprotein-17 domain of membrane-associated protein Q9PRA0_UREPA from *Ureaplasma parvum* was selected for structure determination by the Northeast Structural Genomics Consortium, as part of the Protein Structure Initiative's program on structure–function analysis of protein domains from large domain sequence families lacking structural representatives. The 100-residue Lipoprotein-17 domain is a “domain of unknown function” (DUF) that is a member of Pfam protein family PF04200, a large domain family for which no members have characterized biochemical functions. The three-dimensional structure of the Lipoprotein-17 domain of protein Q9PRA0_UREPA was determined by both solution NMR and by X-ray crystallography at 2.5 Å. The two structures are in good agreement with each other. The domain structure features three α -helices, α 1 through α 3, and five β -strands. Strands β 1/ β 2, β 3/ β 4, β 4/ β 5 are anti-parallel to each other. Strands β 1 and β 2 are orthogonal to strands β 3, β 4, β 5, while helix α 3 is formed between the strands β 3 and β 4. One-turn helix α 2 is formed between the strands β 1 and β 2, while helix α 1 occurs in the N-terminal polypeptide segment. Searches of the Protein Data Bank do not identify any other protein with significant structural similarity to Lipoprotein-17 domain of Q9PRA0_UREPA, indicating that it is a novel protein fold.

Keywords

Structural genomics; Lipoprotein-17 domain of Q9PRA0_UREPA; Domains of unknown function (DUFs); Lipoprotein; *Ureaplasma parvum*; Northeast Structural Genomics Consortium (NESG); Solution NMR; X-ray crystal structure

Introduction

The 100-residue domain of the conserved membrane-associated lipoprotein from *Ureaplasma parvum* (NESG ID: UuR17A, Swiss Prot ID: Q9PRA0_UREPA) was selected for three-dimensional (3D) structure determination by the Northeast Structural Genomics Consortium (NESG; <http://www.nesg.org>) as part of the NIH Protein Structure Initiative program on structural coverage of large domain families. Q9PRA0_UREPA [10] is a conserved putative lipoprotein from *U. parvum*. It contains 834 amino acid residues and has a polypeptide chain molecular mass of 95.9 kDa. Amino acid sequence analysis suggests the presence of at least two conserved domains with unknown functions: Lipoprotein-17 (lipoprotein associated domain, pfam04200) and DUF31 (pfam01732) (Pfam protein family [4]). Lipoprotein-17 domain of Q9PRA0_UREPA is a member of the pfam04200 domain family (<http://pfam.sanger.ac.uk/family?acc=PF04200>). It is approximately 100 residues in length (residues 139–239) and has been found exclusively in six *Mycoplasma* and *Ureaplasma* species, both of which belong to *Mycoplasmataceae* family [4]. Some *Mycoplasma* proteins contain 3–7 repeats of the Lipoprotein-17 domain [4]. The pfam04200 domain family has more than 113 known members, none of which have significant sequence

similarity with any protein of known 3D structure [4]. Here we report the high-quality NMR solution structure (construct spanning residues 128–239) and 2.5-Å resolution X-ray structure (construct spanning residues 133–249) of the Lipoprotein-17 domain of Q9PRA0_UREPA, the first member of the pfam04200 domain family for which a 3D structure has been reported.

Materials and methods

Protein sample preparation

The gene fragment coding for the Lipoprotein-17 domain of the Q9PRA0_UREPA protein from *U. parvum* (UniProtKB/Swiss-Prot ID Q9PRA0_UREPA) was cloned into an *E. coli* expression plasmid. The corresponding protein was expressed and purified following standard, largely-automated protocols of the NESG consortium, to produce protein samples for structure determination by both NMR and X-ray crystallography [1]. For NMR studies, a truncated UU045 gene construct (residues 128–239) was cloned into a pET21 (Novagen) expression vector derivative (pET21_NESG, Ref. [1]), yielding the plasmid pUuR17A-128-239-21.76. For crystallographic studies, a construct coding for residues 133–249 was cloned into pET21_NESG, yielding plasmid pUuR17A-133-249-21.21. Both constructs include an 8-residue affinity tag at the C-terminus (LEHHHHHH) to facilitate protein purification. Details of [U - ^{13}C , ^{15}N]- and [5% ^{13}C , U - ^{15}N]-enriched, and selenomethionine (SeMet)-labeled Lipoprotein-17 domain of the Q9PRA0_UREPA protein sample preparation are provided in the Supplementary Materials. Static light scattering data [1] (Supplementary Fig. S1) and ^{15}N T_1 and T_2 relaxation data (Supplementary Fig. S2) indicate that these protein constructs are monomeric in solution under the conditions used for NMR studies.

Protein NMR studies

Protein samples for NMR studies were prepared at concentrations of ~ 1.0 mM in 95% H_2O /5% $^2\text{H}_2\text{O}$ buffer containing 20 mM MES, 200 mM NaCl, 10 mM DTT, 5 mM CaCl_2 , and 0.02% NaN_3 , at pH 6.5. All NMR spectra were recorded at 25°C using cryogenic NMR probes. Details of NMR data collection and analysis are provided in the Supplementary Materials. Assignments were obtained for ca. 90% of backbone and side-chain resonances (excluding N-terminal $^{15}\text{NH}_3^+$, Lys $^{15}\text{NH}_3^+$, Arg $^{15}\text{N}^\zeta\text{H}_2$, OH of Ser, Thr and Tyr, $^{13}\text{C}^\gamma$ of Asp and Asn, $^{13}\text{C}^\delta$ of Glu and Gln, and aromatic $^{13}\text{C}^\gamma$ shifts). The 2D [^1H , ^{15}N]-HSQC spectrum of the Lipoprotein-17 domain of Q9PRA0_UREPA, labeled with sequence-specific resonance assignments, is shown in Supplementary Fig. S3. NMR data supporting the resonance assignments are summarized in Supplementary Fig. S4. Resonance assignments were deposited in the BioMagResBank on 12/22/2009 with accession code 16648. The locations of regular secondary structure elements were next identified based on backbone chemical shift data [24]. A NOESY peak list containing expected intra-residue, sequential, and α -helical medium-range NOE peaks was initially generated and then manually edited by visual inspection of NOESY spectra [19]. The programs CYANA [11, 12] and AutoStructure [3, 15] were used in parallel to automatically assign long-range NOEs. Statistical summaries of ^1H - ^1H upper-bound distance constraints used for structure calculations are summarized in Table 1. Backbone dihedral angle constraints were derived from chemical shift data using the program TALOS+ [22] (Table 1). The final structure calculation was performed with AutoStructure 2.0, and the 20 conformers with the lowest target function value were refined in an ‘explicit water bath’ [18] using the program CNS [6]. Coordinates were deposited in the Protein Data Bank on 12/22/2009 (PDB accession code 2KRT).

Protein crystallization and crystallography

Crystallization screening was performed using a micro-batch-under-oil crystallization method at 18°C [7], followed by crystallization optimization. Crystals of selenomethionine-derivatized Lipoprotein-17 domain of Q9PRA0_UREPA protein used for structure determination were grown in drops composed of 0.5 uL of protein and 0.5 μ l of precipitant solution (100 mM HEPES, pH 7.5, 100 mM potassium bromide, 40% (w/v) PEG8000) under mineral oil (Hampton Research). The crystals were transferred to paratone oil to remove excess mother liquor before being frozen in liquid propane for data collection at 100 K. Single-wavelength anomalous dispersion (SAD) data were collected at the peak wavelength of the selenium K edge using beamline X4A at the National Synchrotron Light Source. These diffraction data were processed with the HKL2000 package [20]. The crystal belongs to space group $P3_221$, with cell parameters of $a = b = 60.56$ Å, $c = 64.72$ Å. The program SHELXE/D [21] was used to locate a selenium site and to calculate phases to 2.8 Å resolution. Solvent-flattening calculations and partial model building were performed using RESOLVE [23]. The model was completed manually using Coot [9], and was refined against 2.49 Å data with the program PHENIX [2]. The quality of the model was inspected by the program PROCHECK [17]. There is one molecule of the Lipoprotein-17 domain of Q9PRA0_UREPA in the crystallographic asymmetric unit. Data collection and refinement statistics are shown in Table 2. The atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank on 10/08/2009 (PDB accession code 3K63).

Results and discussion

The solution NMR structure of the Lipoprotein-17 domain of Q9PRA0_UREPA (residues 128-239, PDB ID 2KRT) is shown in Fig. 1. Structure quality statistics for this NMR structure are presented in Table 1 (a more complete set of structure quality statistics are presented in Supplementary Table S1). This solution NMR structure exhibits excellent quality assessment scores. The domain structure comprises three α -helices, $\alpha 1$, $\alpha 2$, and $\alpha 3$, and five β -strands (Fig. 1). Strands $\beta 1$ – $\beta 2$, $\beta 3$ – $\beta 4$, and $\beta 4$ – $\beta 5$ are anti-parallel to each other (Fig 1). Strands $\beta 1$ – $\beta 2$ lie orthogonal to the β -sheet formed by strands $\beta 3$, $\beta 4$, and $\beta 5$, while helix $\alpha 3$ is formed between strands $\beta 3$ and $\beta 4$. A single-turn helix $\alpha 2$ is formed between strands $\beta 1$ and $\beta 2$, while helix $\alpha 1$ is formed by the N-terminal polypeptide segment. The protein core, consisting of residues 128–239, is well ordered, while the C-terminal segment, including the hexaHis purification tag, is flexible in solution.

The 2.5-Å X-ray crystal structure of the Lipoprotein-17 domain of Q9PRA0_UREPA (residues 133–249, PDB ID 3K63) is shown in Fig. 2a, and the corresponding refinement statistics are presented in Table 2. The X-ray and NMR structures are superimposed in Fig. 2b. The two structures are generally in good agreement with each other. All five β -strand structures are similar in the NMR and X-ray structures. Differences between the NMR and X-ray structures are observed in flexible loop regions, the disordered C-terminal segment (which is longer in the construct used for crystallographic studies), and in the orientation of N-terminal helix $\alpha 1$. The X-ray structure shows two single-turn helices (residues 160–162 and 172–174) which are not observed in the NMR structure. This difference may be due to the weak intensity of these resonances in the NMR spectra, presumably due to exchange broadening resulting from internal motions. While such dynamics are sometimes functionally important, we cannot speculate on their functional relevance in Lipoprotein-17, as the protein's function is currently unknown. The difference in relative orientation of helix $\alpha 1$ in the NMR and X-ray structures (Fig. 2b) may be due to flexibility that is commonly seen in such N-terminal polypeptide segments.

A Dali [13] search for structurally-similar proteins did not produce any significant hits with other protein structures in the Protein Data Bank. The highest Dali Z-scores are at about the

threshold of being significant (about 4.0). A comparison of the structure of Lipoprotein-17 domain of Q9PRA0_UREPA with a putative tail lysozyme from *Geobacter sulfurreducens* (PDB ID 2IA7, Fig 2d) which has the highest Dali Z-score = 4 and sequence identity of only 7% (using a structure-based sequence alignment), does not reveal much structural similarity. The overall structure of 2IA7 comprises a central core of three pleated antiparallel β -strands and two α -helices. Visual inspection of the two structures (Fig. 2c, d) shows that Lipoprotein-17 domain of Q9PRA0_UREPA has 5 β -strands while the putative lysozyme has only 3 β -strands. The superimposition of the main-chain structures of the Lipoprotein-17 domain and the 2IA7 showed that both structures have α 1 helices in their N-terminal regions, and β 3, β 4, and β 5 strands of the Lipoprotein-17 domain overlap well with β -strands of 2IA7, but the putative tail lysozyme lacks the strands β 1– β 2 and has long α 2 helix instead. These do not appear to be similar folds. Hence we conclude that the Lipoprotein 17 domain of QPRA0_UREPA has a novel fold.

The electrostatic surface potential for the Lipoprotein-17 domain of Q9PRA0_UREPA is shown in Fig. 1c. A hydrophobic patch is located in the center of the molecule where the orthogonal β -strands are located. However, the 3D structure does not yet provide any useful clues into the biochemical function of Q9PRA0_UREPA.

The NMR and X-ray structures presented here are the first 3D structures from the large Pfam domain family Lipoprotein-17 (PF04200). Further understanding of the biological activity of Lipoprotein-17 domains will require further studies of the domain, as well as the full-length Q9PRA0-UREPA protein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Prof. M. Inouye for advice regarding the biology and biochemistry of bacterial lipoproteins. This work was supported by a grant from the Protein Structure Initiative of National Institute of General Medical Sciences, U54-GM074958. NMR (2KRT) and X-ray structures [3JVC (2.7 Å resolution); 3K63 (2.5 Å resolution)] of the Lipoprotein-17 domain have been submitted to the Protein Data Bank. Chemical shifts, NMR constraints, NOESY FID data, and NOESY peak list data have been deposited in the BioMagResDatabase (16648). Expression plasmids for Lipoprotein-17 domain of Q9PRA0_UREPA are available from the PSI Materials Repository (<http://psimr.asu.edu/>).

References

1. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, 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. [PubMed: 15808222]
2. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1948–1954. [PubMed: 12393927]
3. Baran MC, Huang YJ, Moseley HN, Montelione GT. Automated analysis of protein NMR assignments and structures. *Chem Rev.* 2004; 104:3541–3556. [PubMed: 15303826]
4. Bateman A, Birney E, Cerruti L, Durbin R, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. *Nucleic Acids Res.* 2002; 30:276–280. [PubMed: 11752314]

5. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins*. 2007; 66:778–795. [PubMed: 17186527]
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*. 1998; 54:905–921. [PubMed: 9757107]
7. Chayen NE, Stewart PDS, Maeder DL, Blow DM. An automated system for micro-batch protein crystallization and screening. *J Appl Cryst*. 1990; 23:297–302.
8. DeLano, WL. The pymol manual. DeLano Scientific; San Carlos: 2002.
9. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
10. Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature*. 2000; 407:757–762. [PubMed: 11048724]
11. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol*. 1997; 273:283–298. [PubMed: 9367762]
12. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*. 2002; 319:209–227. [PubMed: 12051947]
13. Holm L, Sander C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci*. 1995; 20:478–480. [PubMed: 8578593]
14. 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. [PubMed: 15701001]
15. 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. [PubMed: 16374783]
16. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*. 1996; 14:29–32.
17. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck—a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993; 26:283–291.
18. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. *Proteins*. 2003; 50:496–506. [PubMed: 12557191]
19. Liu GH, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrow-smith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc Natl Acad Sci USA*. 2005; 102:10487–10492. [PubMed: 16027363]
20. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997; 276:307–326.
21. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:1772–1779. [PubMed: 12351820]
22. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS + : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. 2009; 44:213–223. [PubMed: 19548092]
23. Terwilliger TC. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol*. 2003; 374:22–37. [PubMed: 14696367]
24. Wishart DS, Sykes BD. The C-13 Chemical-Shift Index—a simple method for the identification of protein secondary structure using C-13 chemical-shift data. *J Biomol NMR*. 1994; 4:171–180. [PubMed: 8019132]
25. Word JM, Bateman RC, Presley BK, Lovell SC, Richardson DC. Exploring steric constraints on protein mutations using MAGE/PROBE. *Protein Sci*. 2000; 9:2251–2259. [PubMed: 11152136]

Fig. 1. Solution NMR structure of the Lipoprotein-17 domain of Q9PRA0_UREPA from *U. parvum* (residues 128-239, PDB ID 2KRT). **a** Stereoview of the backbone trace of 20 representative conformers, after superposition of backbone N, C α and C' atoms of the regular secondary structure elements for minimal root-mean-square deviation (RMSD). The N- and C-termini are labeled as "N" and "C". **b** Ribbon drawing of the Lipoprotein-17 domain of Q9PRA0_UREPA conformer with the lowest CNS energy. α -helices $\alpha 1$ – $\alpha 3$, β strands $\beta 1$ – $\beta 5$, and the N- and C-termini are labeled. **c** Surface representation of the Lipoprotein-17 domain of Q9PRA0_UREPA conformer with the lowest CNS energy. Surface colors represent the electrostatic potential: blue—positive, red—negative. All figures were prepared with the programs MOLMOL [16] and PYMOL [8].

Fig. 2.
a. 2.5-Å resolution X-ray structure of the Lipoprotein-17 domain of Q9PRA0_UREPA (residues: 133–249, PDB ID: 3K63). **b** Overlaid NMR (PDB ID: 2KRT, *green*) and X-ray (PDB ID: 3K63, *blue*) structures of Lipoprotein-17 domain. **c** Solution NMR structure of Lipoprotein-17 domain of Q9PRA0_UREPA (residues: 128–239, PDB ID: 2KRT) and **d** X-ray structure of putative tail lysozyme (PDB ID: 2IAN, Dali z-score 4.0, sequence id: *ca.* 7%)

Table 1

NMR structure quality statistics for Lipoprotein-17 domain of Q9PRA0_UREPA

Conformationally restricting distance constraints ^a	
Intra-residue [$i = j$]	384
Sequential [$ i - j = 1$]	592
Medium range [$1 < i - j < 5$]	440
Long range [$ i - j \geq 5$]	625
Total	2169
Dihedral angle constraints	
φ	64
ψ	64
Number of constraints per residue ^b	18.7
Number of long range constraints per residue ^b	5.4
Average r.m.s.d. to the mean CNS coordinates [$\text{\AA}$]	
Regular secondary structure elements ^c backbone heavy atoms N, C $^{\alpha}$, C $^{\prime}$	0.7 ± 0.05
Regular secondary structure, ^c all heavy atoms	1.1 ± 0.08
RPF R/P/F/DP [14]	92/93/92/76
Ramachandran plot summary for all ordered (by PSVS) ^d	
Most favorable regions	85.1
Additionally allowed regions	14.5
Generously allowed regions	0.4
Disallowed regions	0.0
Global quality scores (raw/Z-score) ^d	
VERIFY3D	0.26/− 3.21
PROSA2	0.43/− 0.91
PROCHECK G-factor φ and ψ [17]	0.42/− 1.34
PROCHECK G-factor all dihedral angles	−0.40/− 2.37
MOLPROBITY [25] Clash Score	22.04/2.26

^aAnalysed for all residues including HisTag^bThere are 116 residues with conformationally restricting constraints^cRegular secondary elements: alpha helices 6–14, 39–43, 75–81; beta strands : 22–26, 50–54, 61–62, 83–92, 65–69, 98–106^dOrdered residues 6–55, 60–70, 74–108, 110–114, defined based on dihedral angle order parameters $S(\varphi)$ and $S(\psi) > 1.8$. Z-scores were computed relative to corresponding structure quality measures for high resolution X-ray crystal structures [5]

Table 2

X-ray data and refinement statistics for Lipoprotein-17 domain of Q9PRA0_UREPA

<i>Data collection</i>	
Space group	P 3 ₂ 2 1
X-ray wavelength (Å)	0.97908
Unit-cell parameters (Å)	$a = b = 60.56, c = 64.72$
Resolution (Å)	50–2.49
No. of unique reflections	9252
Completeness (%)	99.5 (97.9)
Redundancy	11.2 (10.6)
$I/\sigma(I)$	36.7 (3.6)
$R_{\text{merge}}^{\#}$	0.089 (0.812)
<i>Refinement</i>	
Resolution range (Å)	27.538–2.494
$R_{\text{work}}/R_{\text{free}}^{+}$	0.2131/0.248
No. of subunits in ASU	1
No. of protein non-H atoms	985
No. of water molecules	14
No. of heterogen atoms	0
R.m.s. deviation in bond lengths ^{##} (Å)	0.008
R.m.s. deviation in bond angles ^{##} (°)	1.639
Ramachandran plot	
Most favored regions (%)	84.8
Additional allowed regions (%)	15.2
Generously allowed regions (%)	0

Values in parentheses are for the highest resolution shell (2.49–2.59 Å)

$^{\#}R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum over all reflections and $\sum_i$ is the sum over i measurements of reflection hkl

$^{+}R = \frac{\sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{hkl} F_{\text{obs}}}$; R_{free} is calculated for a randomly chosen set (5%) of reflections which were not used for structure refinement and R_{work} is calculated for the remaining reflections

^{##}Root-mean-square deviation of bond lengths and bond angles from ideal geometry