Title

Crystal structure of a putative quorum sensing-regulated protein (PA3611) from the Pseudomonas-specific DUF4146 family

Journal

Proteins Structure Function and Bioinformatics, 82(6)

ISSN

0887-3585

Authors

Das, Debanu
Chiu, Hsiu-Ju
Farr, Carol L
et al.

Publication Date

2014-06-01

DOI

10.1002/prot.24455

Peer reviewed

Published in final edited form as:
Proteins. 2014 June ; 82(6): 1086–1092. doi:10.1002/prot.24455.

Crystal structure of a putative quorum sensing-regulated protein (PA3611) from the Pseudomonas-specific DUF4146 family

Debanu Das[1,2], **Hsiu-Ju Chiu**[1,2], **Carol L. Farr**[1,3], **Joanna C. Grant**[1,4], **Lukasz Jaroszewski**[1,5,6], **Mark W. Knuth**[1,4], **Mitchell D. Miller**[1,2], **Henry J. Tien**[1,3], **Marc-André Elsliger**[1,3], **Ashley M. Deacon**[1,2], **Adam Godzik**[1,5,6], **Scott A. Lesley**[1,3,4], and **Ian A. Wilson**[1,3,*]

[1]Joint Center for Structural Genomics, http://www.jcsg.org

[2]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California

[3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California

[4]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California

[5]Center for Research in Biological Systems, University of California, San Diego, La Jolla, California

[6]Program on Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, California

Abstract

Pseudomonas aeruginosa is an opportunistic pathogen commonly found in humans and other organisms and is an important cause of infection, especially in patients with compromised immune defense mechanisms. The *PA3611* gene of *P. aeruginosa* PAO1 encodes a secreted protein of unknown function, which has been recently classified into a small Pseudomonas-specific protein family called DUF4146. As part of our effort to extend structural coverage of novel protein space and provide a structure-based functional insight into new protein families, we report the crystal structure of PA3611, the first structural representative of the DUF4146 protein family.

Keywords

Pseudomonas-specific protein family; DUF4146; Pfam PF13652; virulence factor; quorum-sensing; JCSG; structural genomics

[*]Correspondence to: Dr. Ian Wilson, The Scripps Research Institute, Department of Integrative Structural and Computational Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037. wilson@scripps.edu.

INTRODUCTION

Pseudomonas aeruginosa is a ubiquitous environmental bacterium, which is found in soil, marshes and coastal marine habitats, as well as on plant and animal tissues. It is an opportunistic pathogen that is one of the top three causes of infection in humans[1,2]. People afflicted with cystic fibrosis and compromised host defense mechanisms are at increased risk of infections from *P. aeruginosa*. *P. aeruginosa* PAO1 has a large 6.3 Mbp genome with 5,570 predicted open reading frames (ORFs)[1]. As with other organisms, a substantial number of its genes lack functional characterization, although many of these have been assigned putative functional roles based on transcriptome profiling[2–5] and structural genomics approaches[6]. The *PA3611* gene of *P. aeruginosa* PAO1 encodes a secreted protein of unknown function with a molecular weight of ~14 kDa (residues 1– 136) and a calculated isoelectric point of 8.86. PSI-BLAST[7] searches identify ~60 homologues of PA3611 (UniProt ID: Q9HY15), which are all domains of unknown function (DUF) found solely in different strains of Pseudomonas. These proteins have been recently classified into a small Pseudomonas-specific family in Pfam[8], PF13652 (DUF4146), and are all secreted proteins of similar size comprising a single DUF4146 domain. An earlier proteomics analysis using 2D-PAGE and MALDI-TOF mass spectrometry revealed that PA3611 may be a Quorum Sensing (QS)-regulated protein and a potential virulence factor[9]. *P. aeruginosa* PAO1 has 195 known virulence factors according to the Virulence Factor Database[10] and the Pseudomonas Genome Database[11] (http://www.pseudomonas.com). Here we report the crystal structure of PA3611 at 1.6 Å resolution, which was determined using the semi-automated, high-throughput pipeline of the Joint Center for Structural Genomics (JCSG), as part of the NIGMS Protein Structure Initiative (PSI). The structure provides the first structural representative of the PF13652 (DUF4146) protein family.

MATERIALS AND METHODS

Protein production and crystallization

Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method[12]. The gene encoding PA3611 (gi|15598807) was amplified by polymerase chain reaction (PCR) from *Pseudomonas aeruginosa* PAO1 genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5’-ctgtacttccagggcGCCTCGCTCAAGGATTTCGAACTGAGC-3’; reverse primer, 5’-aattaagtcgcgttaCTTCTTGCCCTGGATGCGGCAGCTGCCG-3’, target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector, pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHHHHHHENLYFQ/G), was PCR amplified with V-PIPE (Vector) primers (forward primer: 5’-taacgcgacttaattaactcgtttaaacggtctccagc-3’, reverse primer: 5’-gccctggaagtacaggttttcgtgatgatgatgatgatg-3’). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the I-PIPE / V-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Using the PIPE method, the gene segment encoding residues Met1-Ala19 were deleted from the construct used for structure determination for expression of soluble protein

because it is predicted to contain a signal peptide based on SignalP[13]. Expression was performed in a selenomethionine-containing medium at 25°C. Selenomethionine was incorporated via inhibition of methionine biosynthesis[14], which does not require a methionine auxotrophic strain. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 μg/ml, and the cells were harvested and frozen. After one freeze/thaw cycle, the cells were homogenized and sonicated in lysis buffer [50 mM HEPES pH 8.0, 50 mM NaCl, 10 mM imidazole, 1 mM Tris(2-carboxyethyl)phosphine-HCl (TCEP)] and the lysate was clarified by centrifugation at 32,500 × g for 30 minutes. The soluble fraction was passed over nickel-chelating column (GE Healthcare) pre-equilibrated with lysis buffer, the column washed with wash buffer [50 mM HEPES pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP], and the protein was eluted with elution buffer [20 mM HEPES pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, 1 mM TCEP]. The eluate was buffer exchanged with TEV buffer [20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP] using a PD-10 column (GE Healthcare), and incubated with 1mg of TEV protease per 15 mg of eluted protein for 2 hr at ambient temperature followed by overnight at 4°C. The protease-treated eluate was passed over nickel-chelating column (GE Healthcare) pre-equilibrated with HEPES crystallization buffer [20 mM HEPES pH 8.0, 200 mM NaCl, 40 mM imidazole, 1 mM TCEP] and the column was washed with the same buffer. The flow-through and wash fractions were combined and concentrated to 10.5 mg/ml by centrifugal ultrafiltration (Millipore) for crystallization trials. PA3611 was crystallized using the nanodroplet vapor diffusion method[15] with standard JCSG crystallization protocols[16]. Sitting drops composed of 200 nl protein solution mixed with 200 nl crystallization solution in a sitting drop format were equilibrated against a 50 μl reservoir at 277 K for 22 days prior to harvest. The crystallization reagent consisted of 2.0 M $(NH_4)_2SO_4$, 0.2 M Li_2SO_4, and 0.1 M 3-(Cyclohexylamino)-1-propanesulfonic acid (CAPS) pH 10.5. Ethylene glycol was added to a final concentration of 15% (v/v) as a cryoprotectant. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM)[17] at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). The diffraction data were indexed in orthorhombic space group $P2_12_12_1$. The oligomeric state of PA3611 in solution was determined to be monomeric using a 1 × 30 cm^2 Superdex 200 size exclusion column (GE Healthcare)[12] coupled with miniDAWN (Wyatt Technology) static light scattering (SEC/SLS) and Optilab differential refractive index detectors (Wyatt Technology). The mobile phase consisted of 20 mM Tris pH 8.0, 150 mM NaCl, and 0.02% (w/v) sodium azide. The molecular weight was calculated using ASTRA 5.1.5 software (Wyatt Technology).

Data collection, structure solution and refinement

MAD data were collected at SSRL on beamline 9-2 at wavelengths corresponding to the high-energy remote (λ_1), inflection point (λ_2) and peak (λ_3) of a selenium MAD experiment using the BLU-ICE[18] data collection environment. The data sets were collected at 100 K using a MarMosaic 325 CCD detector (Rayonix, USA). The MAD data were integrated and reduced using XDS[19] and scaled with the program XSCALE. The heavy atom sub-structure and phasing calculations were performed using SOLVE[20]. RESOLVE[21] was used for density modification and ARP/wARP[22] was used for automatic model building to 1.60 Å resolution. Model completion and crystallographic refinement were performed

with the λ_1 data set using COOT[23] and REFMAC5[24]. The refinement protocol included the experimental phase restraints in the form of Hendrickson–Lattman coefficients from SOLVE and TLS refinement with one TLS group for the whole molecule. Data and refinement statistics are summarized in Table I[25,26,27,28]

Validation and deposition

The quality of the crystal structure was analyzed using the JCSG Quality Control server (http://smb.slac.stanford.edu/jcsg/QC). This server verifies: the stereochemical quality of the model using AutoDepInputTool[29], MolProbity[30], and Phenix[31], the agreement between the atomic model and the data using RESOLVE[21], the protein sequence using CLUSTALW[32], the ADP distribution using Phenix, and differences in R_{cryst}/R_{free}, expected R_{free}/R_{cryst} and various other items including atom occupancies, consistency of NCS pairs, ligand interactions and special positions using in-house scripts to analyze refinement log file and PDB header. Protein quaternary structure analysis was performed using the PISA server[33]. Figure 1B was adapted from an analysis using PDBsum[34] and other figures were prepared with PyMOL[35]. Atomic coordinates and experimental structure factors for PA3611 to 1.60 Å resolution (PDB ID: 3npd) were deposited in the Protein Data Bank (www.wwpdb.org).

RESULTS AND DISCUSSION

Cloning, expression, purification and crystallization of PA3611 were carried out using standard Joint Center for Structural Genomics (JCSG; http://www.jcsg.org) protocols. N-terminal residues 1–19 were excluded from the expression construct due to the prediction of a signal peptide cleavage site. The crystal structure of PA3611 was determined by Multi-wavelength Anomalous Diffraction (MAD) phasing to a resolution of 1.60 Å. Data collection, model and refinement statistics are summarized in Table I [25,26,27,28]. A single PA3611 molecule is present in the crystallographic asymmetric unit. The final model (Figure 1) includes Gly0 (left over after cleavage of the expression and purification tag), residues 20–131 of the 136 residues in the full-length protein, 2 sulfate ions and 4 CAPS molecules from the crystallization reagents, 2 1,2-ethane-diol molecules from the cryoprotectant, and 124 water molecules. The Matthews' coefficient[36] is 1.95 $Å^3$/Da, with an estimated solvent content of ~37 %. The Ramachandran plot produced by MolProbity[30] shows that 100% of the residues are in the favored regions.

PA3611 is comprised of one structural domain with five β-strands (B1- B5) and five α-helices (H1-H5). Analysis of the crystallographic packing of PA3611 using the PISA server[5] indicates that a monomer is the biologically relevant oligomeric state of the protein, consistent with the oligomeric state in solution from SEC. The β-strands form a twisted anti-parallel β-sheet flanked on one side by the helices. A disulfide bond is present between Cys92 and Cys130 in helices H3 and H5, respectively. A residue conservation analysis reveals that the conserved residues Ser38, Arg44, Ile46, Tyr55, Val83, Gln86, Ser90, Asn94, Arg98 and Tyr109 line a groove on the surface of the protein (Figure 2).

A search for other proteins of similar structure was carried out using DALI[37], SSM[38] and FATCAT[39] using default search parameters. The SSM (Secondary-Structure Match) search did not identify any significant match (the highest hit had Q-score of 0.14 and Z-score of

0.6), and the best hit with DALI was with fatty acyl-adenylate ligase/saframycin MX1 synthetase (PDB id 3lnv, 3.9 Å r.m.s.d., 10% sequence identity, 87 aligned Cα atoms, Z-score 5.0), although several other proteins gave hits with lower Z-scores. A search for similar structures using the flexible alignment mode in FATCAT resulted in several hits with significant scores (P-value < 0.05), most of which were to α+β class proteins. The most significant hit was to the C-terminal RNA-binding domain of *Escherichia coli* Era GTPase[40] (score of ~0.0006, r.m.s.d. of 2.7 Å, 5.6% sequence identity, alignment length of 92 Cα atoms, PDB id 1ega), which is involved in maturation of the 30S ribosome by binding to 16S ribosomal RNA[41], and belongs to the alpha-lytic protease prodomain-like fold (SCOP fold 54805) and is a member of the prokaryotic type KH domain superfamily (KH-domain type II, SCOP 54814, Pfam clan CL0007) (Figure 3, 4). The KH domain has been shown to be involved in protein-protein interactions in addition to RNA binding[42]. Numerous lysine and arginine residues on a helix-turn-helix motif in the KH domain of Era (Arg239, Lys243, Lys244, Lys250, Lys253, Lys255, Arg262 and Lys263) are implicated in RNA interactions[40]. Although these residues are not directly conserved in the equivalent positions in PA3611, there are some structural similarities and the basic nature of the helix is conserved: Era residues Arg239, Lys244, Lys255 and Lys263 are close to chemically similar residues Arg69, Arg75, Arg77, Arg84 and Arg98 in PA3611. In addition, PA3611 Arg131, which has no equivalent in Era, contributes to the basic nature of this region, which might be involved in ligand or nuclei acid interactions. Ile254 in Era is part of a hydrophobic core and is equivalent to Ile304 in the KH domain of the protein that is implicated in the fragile X syndrome link to mental retardation[43,44]. In PA3611, Val83 is the corresponding residue and part of the conserved groove described above (Figure 2). Analysis of the electrostatic potential surface (using PDB2PQR[45] and the APBS[46] module in PyMOL) reveals an almost equal distribution of basic and acidic residues on the protein surface (Figure 5).

Analysis of potential interacting partners based on genomic context using STRING[47] (http://string.embl.de) indicates that PA3611 interacts with PA3612 (score ~0.8, a 73-residue protein of unknown function classified in PF12843, DUF3820). Also, PA3611 and PA3612 may form a single transcriptional unit based on the prediction that they form an operon according to the Pseudomonas Genome Database. It is also predicted to interact with its adjacent protein PotD (PA3610), which is the polyamine substrate-binding protein in the polyamine uptake system comprised of PotABCD. As polyamine transport has been implicated in quorum sensing and PA3611 was found to be up regulated in quorum sensing, PA3611 (and PA3612) may be involved in quorum sensing via modulation of PotD's function, with implications specific to biofilm formation in Pseudomonas. A computational assessment of PA3611 using a Support Vector Machine method as implemented in VirulentPred[48] (http://203.92.44.117/virulent/index.html) predicts PA3611 as a virulence factor. The structure of PA3611 provides some clues into the potential function of this protein and will serve as a guide for further investigation into its molecular and cellular role.

Acknowledgments

We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of

SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). Genomic DNA from *Pseudomonas aeruginosa* PAO1 (ATCC Number 47085D-5) was obtained from the American Type Culture Collection (ATCC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Grant Sponsor: NIH, National Institute of General Medical Sciences (NIGMS), Protein Structure Initiative; contract grant number: U54 GM094586.

REFERENCES

1. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock- Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of*Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature. 2000; 406:959–964. [PubMed: 10984043]
2. Schuster M, Lostroh CP, Ogi T, Greenberg EP. Identification, timing, and signal specificity of*Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. J Bacteriol. 2003; 185:2066–2079. [PubMed: 12644476]
3. Hentzer M, Wu H, Andersen JB, Riedel K, Rasmussen TB, Bagge N, Kumar N, Schembri MA, Song Z, Kristoffersen P, Manefield M, Costerton JW, Molin S, Eberl L, Steinberg P, Kjelleberg S, Hoiby N, Givskov M. Attenuation of*Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. EMBO J. 2003; 22:3803–3815. [PubMed: 12881415]
4. Folsom JP, Richards L, Pitts B, Roe F, Ehrlich GD, Parker A, Mazurie A, Stewart PS. Physiology of*Pseudomonas aeruginosa* in biofilms as revealed by transcriptome analysis. BMC Microbiol. 2010; 10:294. [PubMed: 21083928]
5. Schuster M. Global expression analysis of quorum-sensing controlled genes. Methods Mol Biol. 2011; 692:173–187. [PubMed: 21031312]
6. Mercier KA, Cort JR, Kennedy MA, Lockert EE, Ni S, Shortridge MD, Powers R. Structure and function of Pseudomonas aeruginosa protein PA1324 (21–170). Protein Sci. 2009; 18:606–618. [PubMed: 19241370]
7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
8. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The Pfam protein families database. Nucleic Acids Res. 2008; 36:D281–D288. [PubMed: 18039703]
9. Nouwens AS, Beatson SA, Whitchurch CB, Walsh BJ, Schweizer HP, Mattick JS, Cordwell SJ. Proteome analysis of extracellular proteins regulated by the las and rhl quorum sensing systems in Pseudomonas aeruginosa PAO1. Microbiology. 2003; 149:1311–1322. [PubMed: 12724392]
10. Yang J, Chen L, Sun L, Yu J, Jin Q. VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. Nucleic Acids Res. 2008; 36:D539–D542. [PubMed: 17984080]
11. Winsor GL, Van Rossum T, Lo R, Khaira B, Whiteside MD, Hancock RE, Brinkman FS. Pseudomonas Genome Database: facilitating user-friendly, comprehensive comparisons of microbial genomes. Nucleic Acids Res. 2009; 37:D483–D488. [PubMed: 18978025]
12. Klock HE, Koesema EJ, Knuth MW, Lesley SA. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins. 2008; 71:982–994. [PubMed: 18004753]
13. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004; 340:783–795. [PubMed: 15223320]
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. [PubMed: 7678431]

15. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens RC. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr. 2002; 35:278–281.
16. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA. 2002; 99:11664–11669. [PubMed: 12193646]
17. Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP. An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot. J Appl Crystallogr. 2002; 2002:720–726.
18. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat. 2002; 9:401–406. [PubMed: 12409628]
19. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallogr. 1993; 26:795–800.
20. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr Sect D Biol Crystallogr. 1999; 55:849–861. [PubMed: 10089316]
21. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr Sect D Biol Crystallogr. 2000; 56:965–972. [PubMed: 10944333]
22. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol. 1999; 6:458–463. [PubMed: 10331874]
23. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
24. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Meth Enzymol. 2003; 374:300–321. [PubMed: 14696379]
25. Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol. 1997; 4:269–275. [PubMed: 9095194]
26. Cruickshank DW. Remarks about protein structure precision. Acta Crystallogr Sect D Biol Crystallogr. 1999; 55:583–601. [PubMed: 10089455]
27. Weiss MS, Hilgenfeld R. On the use of the merging R factor as a quality indicator for Xray data. J Appl Crystallogr. 1997; 30:203–205.
28. Weiss MS, Metzner HJ, Hilgenfeld R. Two non-proline cis peptide bonds may be important for factor XIII function. FEBS Lett. 1998; 423:291–296. [PubMed: 9515726]
29. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Crystallogr Sect D Biol Crystallogr. 2004; 60:1833–1839. [PubMed: 15388930]
30. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007; 35:W375–W383. [PubMed: 17452350]
31. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
32. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947–2948. [PubMed: 17846036]
33. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372:774–797. [PubMed: 17681537]

34. 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. [PubMed: 15608193]
35. DeLano, WL. The PyMOL Molecular Graphics System. Palo Alto, CA, USA: DeLano Scientific LLC; 2008.
36. Matthews BW. Solvent content of protein crystals. J Mol Biol. 1968; 33:491–497. [PubMed: 5700707]
37. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics. 2008; 24:2780–2781. [PubMed: 18818215]
38. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sect D Biol Crystallogr. 2004; 60:2256–2268. [PubMed: 15572779]
39. Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics. 2003; 19:ii246–ii255. [PubMed: 14534198]
40. Chen X, Court DL, Ji X. Crystal structure of ERA: a GTPase-dependent cell cycle regulator containing an RNA binding motif. Proc Natl Acad Sci USA. 1999; 96:8396–8401. [PubMed: 10411886]
41. Sharma MR, Barat C, Wilson DN, Booth TM, Kawazoe M, Hori-Takemoto C, Shirouzu M, Yokoyama S, Fucini P, Agrawal RK. Interaction of Era with the 30S ribosomal subunit implications for 30S subunit assembly. Mol Cell. 2005; 18:319–329. [PubMed: 15866174]
42. Chen T, Damaj BB, Herrera C, Lasko P, Richard S. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. Mol Cell Biol. 1997; 17:5707–5718. [PubMed: 9315629]
43. De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, de Graaff E, Oostra BA, Willems PJ. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nat Genet. 1993; 3:31–35. [PubMed: 8490650]
44. Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G. The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. Cell. 1993; 74:291–298. [PubMed: 7688265]
45. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 2007; 35:W522–W525. [PubMed: 17488841]
46. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA. 2001; 98:10037–10041. [PubMed: 11517324]
47. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009; 37:D412–D416. [PubMed: 18940858]
48. Garg A, Gupta D. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinformatics. 2008; 9:62. [PubMed: 18226234]
49. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010; 38(Suppl):W529–W533. [PubMed: 20478830]

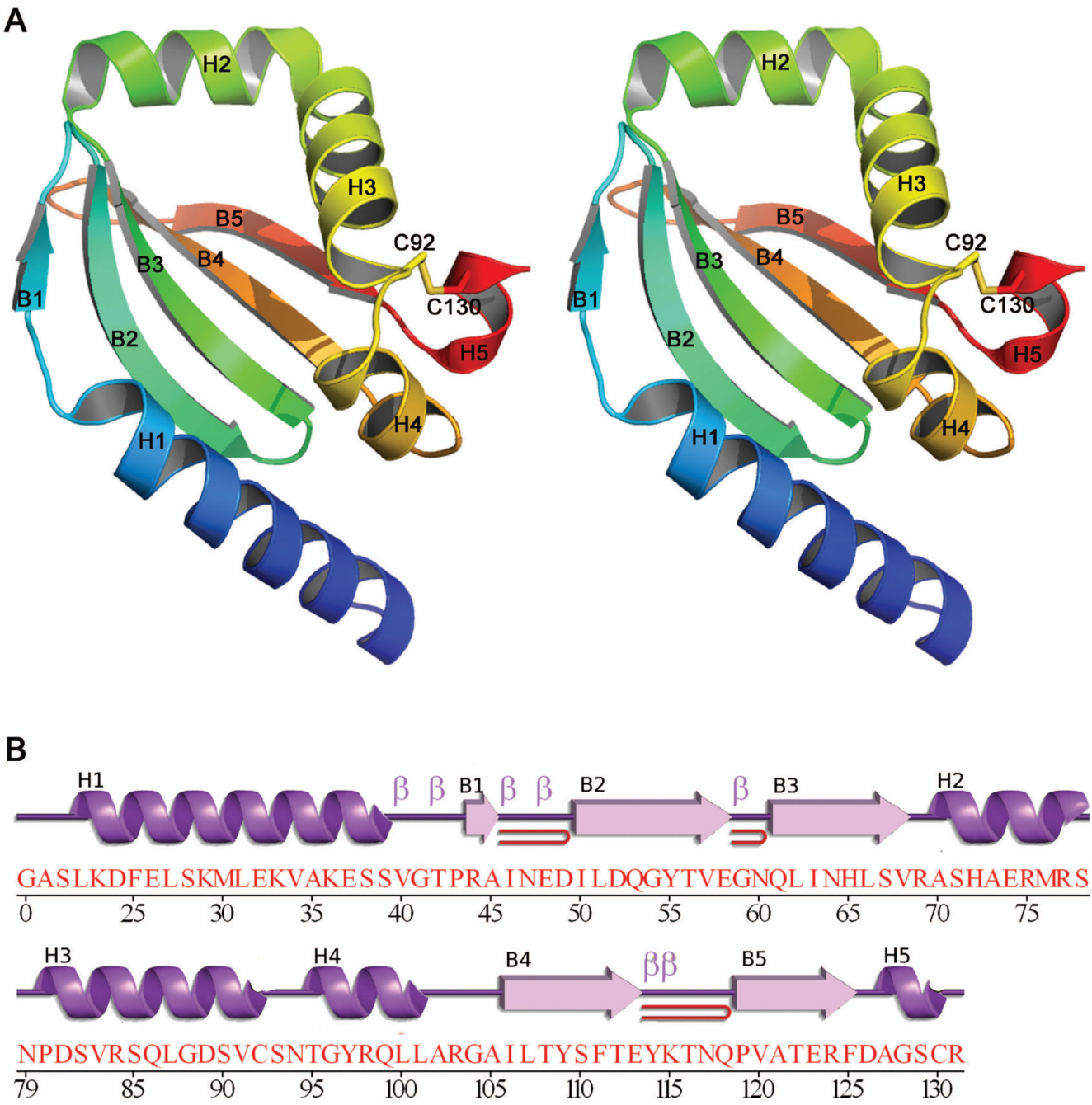

Figure 1. Crystal structure of PA3611 from *Pseudomonas aeruginosa*

(A) Stereo ribbon diagram of PA3611 color-coded from N-terminus (blue) to C-terminus (red). Helices H1-H5, β-strands (B1-B5) and the Cys92-Cys130 disulfide bond are indicated. (B) Diagram showing the secondary structure elements of PA3611 superimposed on its primary sequence, adapted from PDBSum (http://www.ebi.ac.uk/pdbsum). The α-helices (H1-H5), β-strands (B1-B5), β-turns (β) and β-hairpins (red loops) are indicated. The sequence and structure includes Gly0, which remains after removal of the expression and purification tag, and residues 20–131 of the 136 residues in the entire protein.

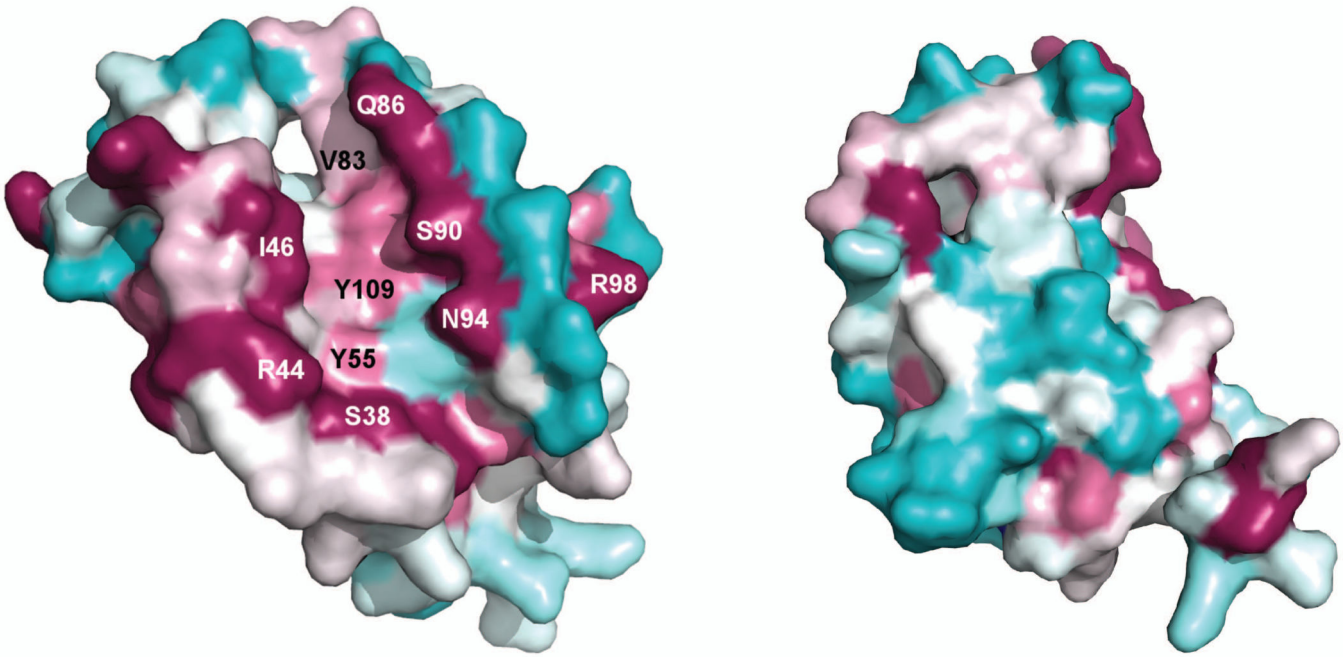

Figure 2. Sequence conservation in PA3611

Residue conservation analysis performed using CONSURF[49] of PA3611(UniProt Q9HY15) homologs included in the analysis (with UniProt id codes and sequence identities to PA3611 in parentheses): PSPPH_3802 from *Pseudomonas syringae pv. phaseolicola* str. 1448A (Q48F97, 70%), PPUT_4084 from *Pseudomonas putida* str F1 (A5W7U8, 57%), PFL01_1180 from *Pseudomonas fluorescens* str Pf0-1 (Q3KH33, 59%), PFLU_1192 from *Pseudomonas fluorescens* str SBW25 (C3KDG3, 52%). The surface representation color gradient goes from cyan (most variable) to magenta (most conserved). Most of the conserved residues in PA3611 and its homologs are located on one side of the protein and are located on the surface or surround a prominent groove in the protein. They include Ser38, Arg44, Ile46, Tyr55, Val 83, Gln86, Ser90, Asn94, Arg98 and Tyr109 (labeled in left panel; orientation is similar to that in Fig. 1). Right panel: ~180° rotation around a vertical axis compared to the left panel.

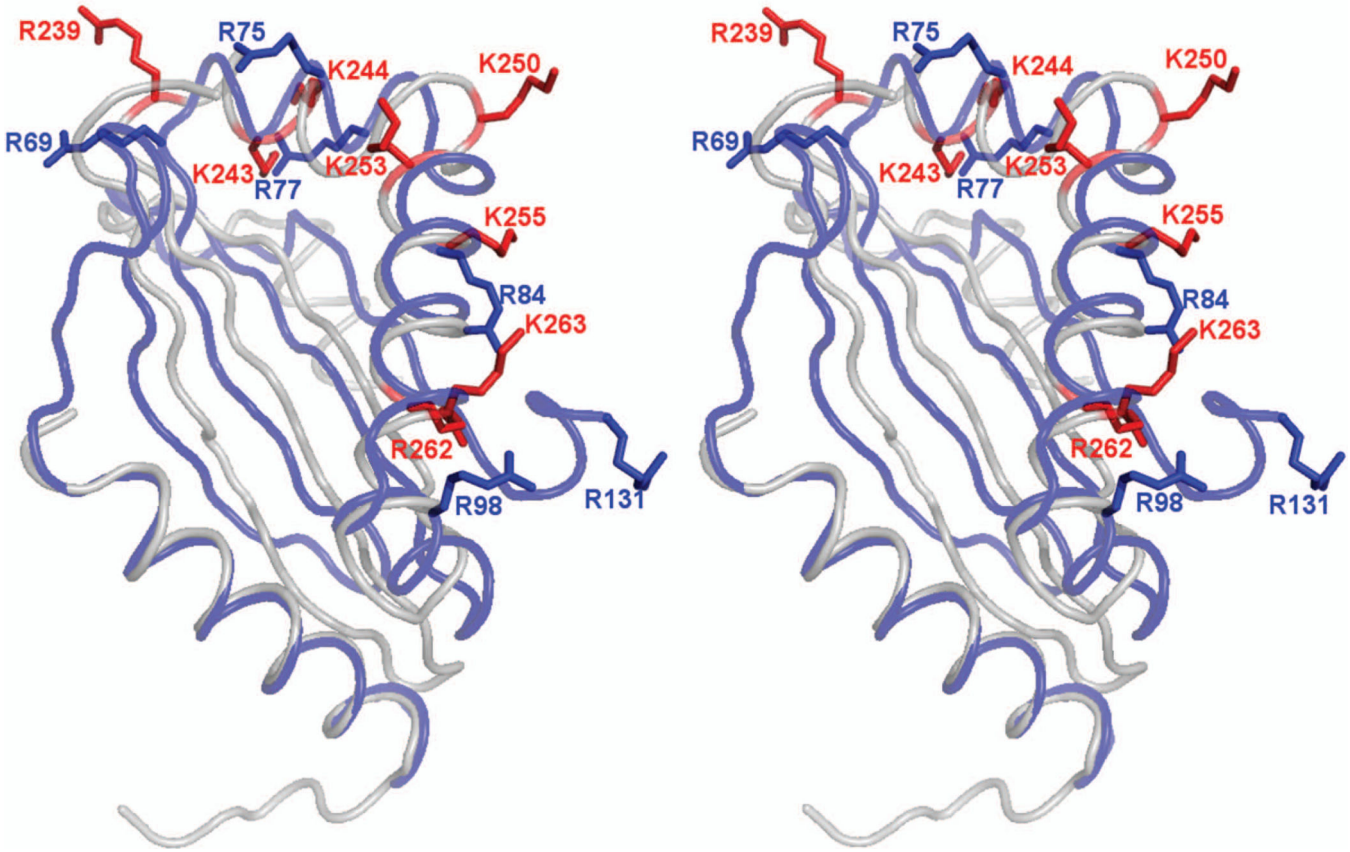

Figure 3. Structural comparison of PA3611 with the RNA-binding domain of Era

PA3611 (blue) is structurally similar to the C-terminal RNA-binding domain of *E. coli* Era GTPase (grey). Some of the Era residues (red sticks) implicated in RNA binding (Arg239, Lys244, Lys255 and Lys263) approximate the location of PA3611 residues (blue sticks) Arg69, Arg75, Arg77, Arg84 and Arg98, respectively. Although, only Arg98 is conserved in PA3611 homologs, the similar chemical nature of the residues leads to a similar basic region on both proteins. PA3611 Arg131, which is not a counterpart of any Era residues, also contributes to the basic nature of this region and might be involved in ligand or nucleic acid interactions.

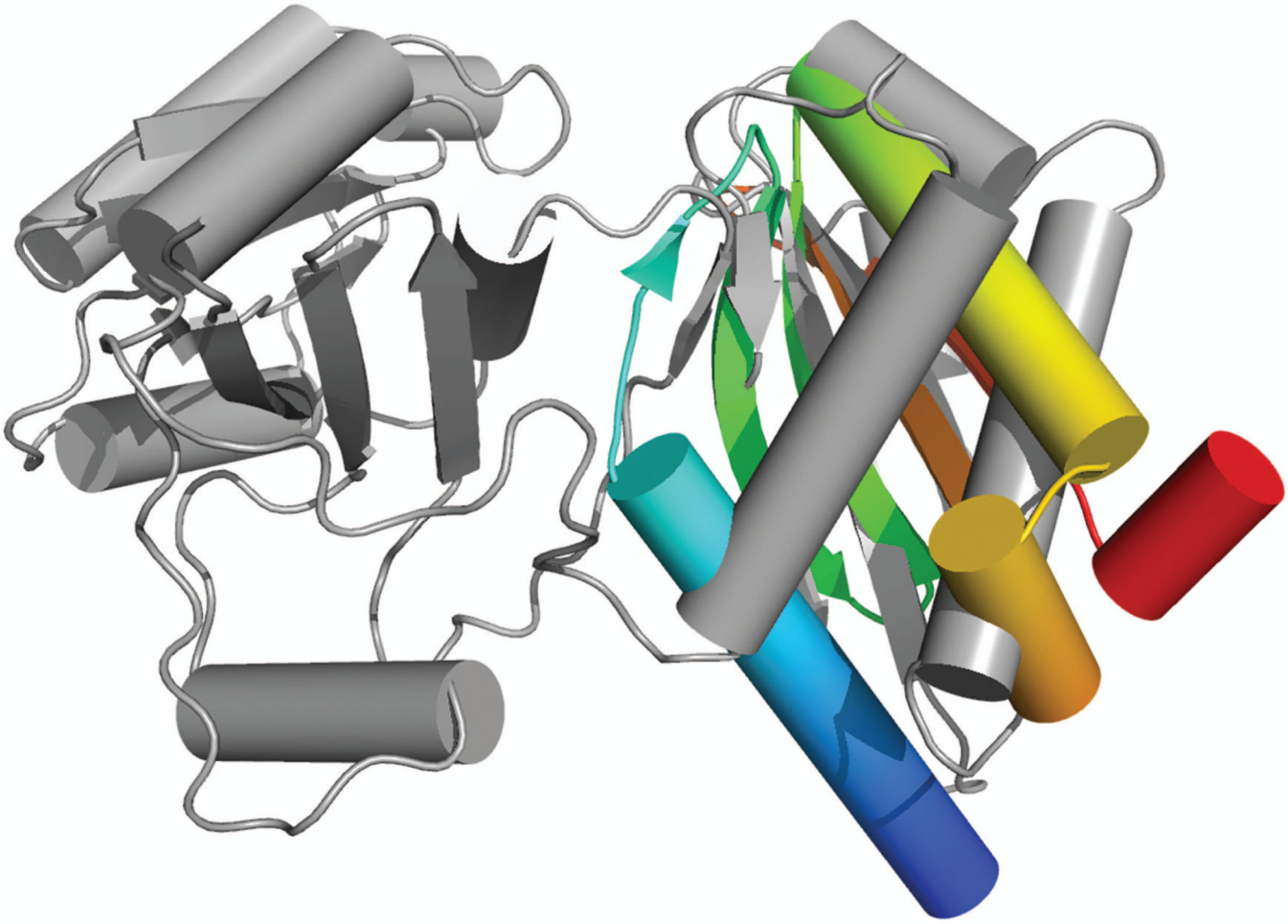

Figure 4. Structural alignment of PA3611 Era

Superimposition of PA3611 (in blue to red from N- to C-terminus, similar orientation as in Figure 3) with the *E. coli* Era GTPase (grey) using DaliLite[37] (Z-score=3.2, r.m.s.d of 3.2 Å over 71 Cα residues) highlighting the significant structural differences in helix positions but the FATCAT flexible alignment mode is still able to identify the overall structural similarity between PA3611 and the C-terminal RNA-binding domain of Era.

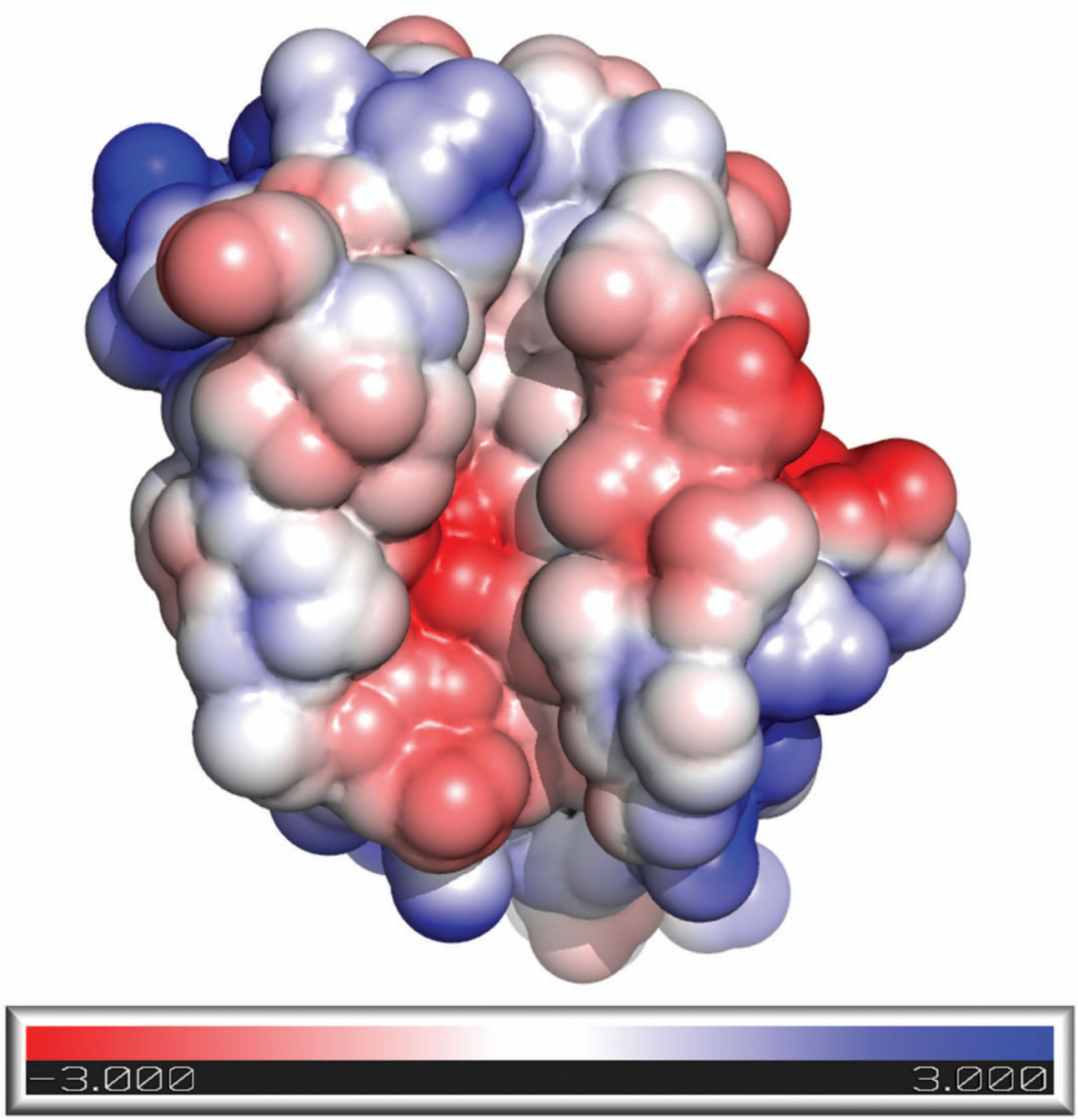

Figure 5. Surface charge analysis

The electrostatic surface representation of PA3611 shows an almost equal distribution of positively- and negatively-charged residues (blue and red, respectively) on the protein surface. The basic region (blue) is primarily made up of residues described in Figure 3. The molecular orientation is similar to that in Figure 1. The color scale is in units of kT/e from −3 to +3.

TABLE I

Summary of crystal parameters, data collection and refinement statistics for PDB 3npd

Space group	P 2_1 2_1 2_1		
Unit cell parameters	a = 34.99 Å, b = 51.32 Å, c = 56.90 Å		
Data collection	λ_1 MAD Se	λ_2 MAD Se	λ_3 MAD Se
Wavelength (Å)	0.91837	0.97941	0.97925
Resolution range (Å)	28.9–1.60	28.9–1.60	28.9–1.60
Highest resolution shell (Å)	1.64–1.60	1.64–1.60	1.64–1.60
Number of observations	49,327	49,052	49,176
Number of unique reflections	14,032	14,047	14,051
Completeness (%)	99.6 (99.3)	99.6 (99.4)	99.6 (98.9)
Mean I/σ (I)	10.9 (2.1)	11.1 (2.1)	10.5 (2.0)
R_{merge} on $I^{\dagger}$ (%)	7.4 (64.9)	7.5 (62.6)	8.5 (64.8)
R_{meas} on $I^{\ddagger}$ (%)	8.7 (76.6)	8.9 (73.8)	9.9 (76.4)
$R_{p.i.m.}$ on $I^{\ddagger\ddagger}$ (%)	4.5 (40.0)	4.6 (38.5)	5.1 (39.8)
Model and refinement statistics			
Resolution range (Å)	28.9–1.60	Data set used in refinement	λ_1
No. of reflections (total)	13,997[a]	Cutoff criteria	|F|>0
No. of reflections (test)	700	R_{cryst} ¶	0.160
Completeness (% total)	99.5	R_{free} ¶	0.205
Stereochemical parameters			
Restraints (RMSD observed)			
Bond angle (°)	1.7		
Bond length (Å)	0.016		
Average isotropic B-value†† / Wilson plot B-value (Å^2)	22.7 / 17.8		
ESU based on R_{free}	0.096		
Protein residues/ atoms (Å)	113 / 868		
Water/ solvent molecules	124 / 8 (SO4=2, CAPS=4, EDO=2)		
Ramachandran plot: residues (%) in favored / allowed	100 / 100		

Values in parentheses are for the highest resolution shell.

[a] Typically, the number of unique reflections used in refinement is slightly less than the total number that were integrated and scaled. Reflections are excluded due to systematic absences, negative intensities, and rounding errors in the resolution limits and cell parameters ESU = Estimated overall coordinate error [26]

$^{\dagger} R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\,\Sigma_i(hkl)$.

$^{\ddagger} R_{meas} = \Sigma_{hkl}[N/(N\text{-}1)]^{1/2}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$ [25].

$^{\ddagger\ddagger} R_{p.i.m.}$ (precision-indicating R_{merge}) $= \Sigma_{hkl}[(1/(N\text{-}1)]^{1/2}\,\Sigma_i|I_i\,(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i\, I_i(hkl)$ [27,28].

[¶] $R_{cryst} = \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. R_{free} is the same as R_{cryst} but for 5.0% of the total reflections chosen at random and omitted from refinement.

[††] This value represents the total *B* that includes TLS and residual *B* components.