

ARTICLE

Open Access

DUF3055 from *Staphylococcus aureus* adopts unique strategy for structural distinctiveness



Hyo Jung Kim^{1*}  and Hyoun Sook Kim^{2*}

Abstract

Staphylococcus aureus remains a public health threat with the WHO classifying the pathogen as a high priority in the development of new antimicrobial agents. Whole genome sequencing has revealed a number of conserved genes that may be essential for cell viability and infection. Characterising the structure and function of these proteins will inevitably aid development of new antimicrobials. Therefore, this study elucidated the structure of hypothetical protein DUF3055 from *S. aureus* stain Mu50. The protein possesses an as yet undefined function and a unique fold. The size of DUF3055 made it an ideal candidate for NMR characterisation which in conjunction with circular dichroism revealed the protein to be folded. Crystallisation and structural solution found that the overall dimer fold has a negatively charged surface formed by a β -bulge and tightly crossed α -helices, with a complementary size to a DNA single turn. Our structural observations suggest that hypothetical protein DUF3055 from *S. aureus* has a role in DNA binding and gene regulation.

Keywords *Staphylococcus aureus*, Surface charge, DUF3055, Structure–activity relationship, Hypothetical protein

Introduction

Proteins perform essential functions that facilitate physiological processes; these functions are tied to their structure with surface shape playing a pivotal role. As such, investigations into protein structure can aid us in understanding the function of unknown proteins. Currently, 180,000 protein structures are available in the RCSB PDB database representing less than 0.1% of proteins identified through genome sequencing [1]. Computational methods can assist in the structural analysis of unknown proteins but predictive models produced by PROSITE, SWISS-MODEL, and AlphaFold can be spurious in the case of novel folds, unique structure, or fail to consider

quaternary structure. Therefore, it is crucial to experimentally determine protein structure in order to confidently understand biological systems.

Staphylococcus aureus is a pathogenic bacteria considered to be of high priority to the World Health Organisation (WHO). The pathogen caused over 1 million deaths in 2019 primarily from lower respiratory and bloodstream infections [2]. *S. aureus* can acquire a wide range of antibiotic resistances and currently 99% of *S. aureus* strains are resistant to penicillin [3]. In healthcare settings both vancomycin-resistant (VRSA) and methicillin-resistant strains (MRSA) are frequently observed with some hospitals reporting that >50% of strains exhibit resistance to methicillin [4].

The genome of *S. aureus* strain Mu50 is made up of 2,878,529 nucleotides encoding ~2700 genes that will produce proteins organised into ~2400 families [5, 6]. Presently, >1500 proteins from *S. aureus* have been structurally solved with proteins from the Mu50 strain accounting for less than 200. About 20% of assigned families also include domains of unknown functions (DUFs), which categorize functionally uncharacterized proteins

*Correspondence:

Hyo Jung Kim
hyojungkim@woosuk.ac.kr
Hyoun Sook Kim
hskim@ncc.re.kr

¹ College of Pharmacy, Woosuk University, Wanju 55338, Republic of Korea

² Research Institute, National Cancer Center, Goyang 10408, Republic of Korea

[7]. Although functional information is limited, evolutionary conservation suggests that many DUFs are essential. By converging the database of essential genes on Pfam DUF database, over 300 essential proteins from 16 different bacterial species are found to contain 238 DUFs, and most of which represent single-domain structure [8, 9]. DUF3055 domain possessing proteins are highly conserved in *Alkalihalophilus*, *Bacillus*, *Macrococcus*, and *Staphylococcus*. All proteins classified in the DUF3055 domain-containing proteins are single-domained with less than 100 amino acids. This implies biological essentiality of DUF3055 in bacteria. In the present study, the crystal structure of DUF3055 domain-containing protein from *S. aureus* is determined ($_{SA}$ DUF3055). The unique fold possesses an abundance of glycine residues in both its α -helices and β -strands which typically disfavour secondary structure. Our studies highlight that DUF3055 most likely functions as a DNA mimic and may form one component of a toxin-antitoxin pair.

Methods/experimental

Cloning, expression and purification

The SAV0927 gene coding for $_{SA}$ DUF3055 was amplified from the genomic DNA of *S. aureus* Mu50 by polymerase chain reaction using 5'-GAACATATGATTGATATGTATTTATATGAT-3' and 5'-GATCTCGAGATGAATGACTTCATTTAAATA-3' as the forward and reverse primers, respectively. The gene for $_{SA}$ DUF3055 was cloned into the *NdeI* and *XhoI* restriction site of the pET21a (+) vector. The resulting construct has additional residues (LEHHHHHH) that encodes a C-terminal histidine tag. The sequence of the cloned vector was confirmed by DNA sequencing (data not shown). The recombinant plasmid was transformed into *E. coli* BL21 (DE3). Cells were grown with ampicillin (50 μ g/ml) in LB medium for circular dichroism and crystallography and M9 medium supplemented with ^{15}N - NH_4Cl for NMR spectroscopy. The cultures were incubated at 37 °C with shaking at 180 rpm until OD_{600} reached 0.5 and expression was induced with 0.5 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation at 4500 \times g at 4 °C after 4 h of additional growth. For purification, the cell pellet was resuspended in buffer (50 mM Tris-HCl, pH 7.5, and 500 mM NaCl) and disrupted using an Ultrasonic processor (Cole-Parmer, U.S.A.) at 4 °C. After centrifugation at 20,000 \times g for 1 h at 4 °C, the supernatant was purified by binding to His-Trap affinity column (Cytiva, England) and eluted with elution buffer containing 300 mM imidazole. The purified protein was analyzed with >99% purity by SDS-PAGE and was concentrated to 10 mg/ml by ultrafiltration in 3,000 Da molecular-mass cutoff spin columns (Millipore, U.S.A.).

Circular dichroism and NMR spectroscopy

Circular dichroism experiments were carried out on a Circular Dichroism Spectropolarimeter J815 (JASCO, Japan). $_{SA}$ DUF3055 was diluted to 20 μ M in 50 mM sodium phosphate (pH 7), 100 mM NaCl at 20 °C. A 1 mm path quartz cuvette was used for measurements. The far-UV CD spectra were recorded from 195 to 260 nm with a scan speed of 2 cm/min. The spectra were plotted as a mean ellipticity θ ($\text{degree} \times \text{cm}^2 \times \text{dmol}^{-1}$) versus wavelength λ (nm). Thermal unfolding scan experiments were conducted by increasing temperature from 20 °C to 95 °C in steps of 0.2 °C.

For NMR spectroscopy, 0.5 mM ^{15}N labeled protein was prepared in 20 mM sodium phosphate (pH 6), 100 mM NaCl, 1 mM dithiothreitol, in 90% (v/v) H_2O and 10% (v/v) D_2O buffer. The NMR spectrum was recorded at 298 K on a JEOL 600 MHz NMR spectrometer equipped with a cryoprobe. The 2D- ^{1}H - ^{15}N HSQC spectrum was acquired for uniformly ^{15}N labelled $_{SA}$ DUF3055. The spectrum was processed with NMRPipe and analyzed using NMRview [10, 11].

Crystallization and data collection

$_{SA}$ DUF3055 samples for crystallization were diluted to 10 mg/ml with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. Initial crystallization screening was established in 96-well sitting drop (MiTeGen, U.S.A.) at 20 °C. $_{SA}$ DUF3055 was crystallized in a variety of conditions, which were further improved by hanging-drop vapor diffusion methods. For the optimal growth of $_{SA}$ DUF3055 crystals, 1 μ l of precipitant solution (1 M $(\text{NH}_4)_2\text{SO}_4$, 1% (w/v) PEG3350, 100 mM Bis-Tris pH 5.5) and 1 μ l of protein solution was mixed on a siliconized cover slip and equilibrated against a 1 ml reservoir of precipitant solution. This condition yielded block-shaped crystals in 5 days which were cryo-protected in artificial mother liquor supplemented with 20% (w/v) glycerol. Diffraction data was collected at 100 K to a resolution of 3.0 Å on beamline 5C at the Pohang Light Source, Korea. Data were processed and scaled through HKL2000 program suite [12]. Further structural analysis was performed using the CCP4 suite [13]. The $_{SA}$ DUF3055 crystals belong to space group P6 $_5$ 22 and contained 6 monomers (3 dimers) per asymmetric unit.

The structure was determined by molecular replacement technique using Phaser with the coordinates of a model predicted by program AlphaFold from Google Deepmind as a starting model [14]. Refinement was continued with Refmac5 and model was built using Coot [15, 16]. 5% of data was set aside for the refinement calculations of R_{free} , and the final crystallographic statistics are summarized in Table 1.

Table 1 Crystallographic data collection and refinement statistics

| | DUF3055 (8XFU) |
|---|--|
| Data collection | |
| Beamline | BL-17A |
| Wavelength (Å) | 0.99 |
| Resolution range (Å) ^a | 50.06–2.99 |
| Space Group | P6522 |
| Unit cell parameters (Å) | a = 106.08 b = 106.08 c = 454.03 |
| Observations (total/unique) | 54,753/30,588 |
| Completeness (%) | 95.9 (82.5) |
| R _{merge} | 0.35 (0.85) |
| CC _{1/2} | 0.96 (0.10) |
| Redundancy | 10.1 (4.4) |
| I/sigma | 12.9 (4.2) |
| Refinement | |
| R _{work} ^b / R _{free} ^b (%) | 27.15/33.80 |
| Protein atoms | 4258 |
| Average B value (Å ²) | 20.71 |
| r.m.s.d. bond (Å) | 0.008 |
| r.m.s.d. angle (°) | 1.603 |

^a Numbers in parentheses indicate the statistics for the last resolution shell

^b R_{work} = $\sum(|F_{obs} - F_{calc}|) / \sum F_{obs}$, where F_{obs} = observed structure factor amplitude, and F_{calc} = structure factor calculated from model. R_{free} is computed in the same manner as R_{work}, but from a test set containing 5% of data excluded from the refinement calculation

Accession number

The coordinates of *s*_ADUF3055 was deposited in the PDB database as entry 8XFU.

Result

Physical stability of *s*_ADUF3055

Initial characterisation of *s*_ADUF3055 focused on thermodynamic stability and secondary structure analysis using CD [17]. Our far UV spectra showed two distinguishable peaks with the lowest negative peak near 220 nm and an adjacent minor peak at 212 nm, which indicated the protein had a predominantly β -sheet secondary structure. However, the spectra was slightly shifted to lower wavelengths compared to a theoretical β -sheet indicative of a smaller proportion of α -helix. Conversely, secondary structure prediction using K2D2 and BeStSel suggested that *s*_ADUF3055 had a greater α -helix content of 60% [18, 19]. Given the lack of signal for random coil it is unlikely that the protein contains unstructured regions. Temperature scans at 220 nm show that *s*_ADUF3055 is remarkably stable at all temperature below 70 °C, exhibiting a melting transition (T_m) at 72 °C (Fig. 1A). To study the changes in protein conformation and structural determination, protein NMR spectroscopy was conducted. Attempts to produce ¹H-¹⁵N-¹³C triple-labeled *s*_ADUF3055 were not successful since inefficient amount of cells were grown using ¹³C-glucose as a carbon source. ¹H-¹⁵N 2D NMR HSQC spectrum showed a few dispersed broad peaks with around 50 counts. This confirmed the protein was folded and supports our CD observations. However, it is also apparent that many residues are showing dynamics in different timescales because of the uneven intensity distribution (Fig. 1B). As such, for structural elucidation crystallography was performed.

Structural analysis of DUF3055

The crystal structure of *s*_ADUF3055 was determined to a 3.0 Å resolution with clear density from residue 2 to the C-terminal histidine tags (at residue 92). The monomer

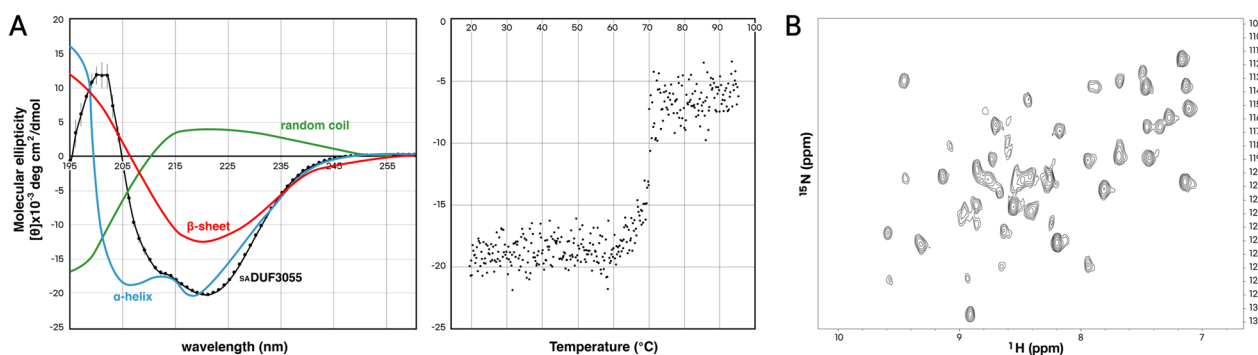


Fig. 1 Protein stability of *s*_ADUF3055. **A** (left) Circular dichroism spectrum in the form of molecular ellipticity against wavelength (nm) for *s*_ADUF3055 (colored in black). The experiment was measured from 195 to 260 nm in Tris-HCl buffer, pH 6.0 at 20 °C. Standard CD spectra are drawn as references (blue, α -helix; red, β -sheet; green, random coil). Values are the mean of three separate determinations. (Right) Thermal stability of *s*_ADUF3055 was analyzed using circular dichroism at 222 nm. **B** NMR contour plots for the *s*_ADUF3055 ¹H-¹⁵N HSQC experiments conducted at 600 MHz frequency

shows an antiparallel β -sheet and 3 consecutive α -helices with $\beta 1$ - $\beta 2$ - $\beta 3$ - $\beta 4$ - $\alpha 1$ - $\alpha 2$ - $\alpha 3$ topology (Fig. 2A and B). The $_{SA}$ DUF3055 crystal belongs to the $P6_522$ spacegroup with unit cell dimensions 106, 106, 454 Å, that harbors 3 dimers in an asymmetric unit. In contrast to our previously characterised crystal system which adopted a linear polymer this new crystal form has molecules packed in a ball-like state with substantial solvent volume (43%) [20]. Although the dimerization is mainly dependent upon β -sheet backbone hydrogen bonds, some interactions are observed between side chains. 25 out of 87 residues (30%) participate in hydrogen bonding or salt bridge interactions with the neighbouring monomer with a large 1900 Å² contact surface formed by the long $\beta 1$ strand. A small repeating sequence FVGFVG in the strand supports dimer formation through a π - π stack between Phe20 and Tyr5 and π -cation interaction between Phe17 and Arg36. We surmise that the stability observed in our CD experiments is largely granted by the tight bonding between two monomers (Fig. 2C).

Additionally, the $\beta 1$ strand in both monomers are distorted and irregular. When the standard β -strands are exposed to solvent area, they become aggregation-prone structure and favors to form insoluble fibrils. Therefore, natural proteins adopt alternative β -strand architecture rather than ordinary shape to protect β -strand on the edge (β -edge). Therefore, natural

proteins adopt alternative β -strand architecture rather than ordinary shape to protect β -strand on the edge (β -edge) [20, 21]. In contrast to a typical β -strand, the $\beta 1$ strand of $_{SA}$ DUF3055 shows a loss in planarity, featuring a $\sim 115^\circ$ bend with Glu12 serving as the apex of the protrusion. The bulged β -strand also exhibits twisted edges, incorporating atypical glycine residues for β -strands (Gly19 and Gly52) [22]. Sheltered beneath a canopy, three consecutive α -helices occupy the pocket area, establishing secure interactions with bulged β -strands. In addition to the β -strands, the α -helices exhibit distinctive characters. Despite glycine being one of the most destabilizing residues for an α -helix due to the wider range of backbone dihedral angles, it is frequently observed in membranous proteins, especially in the region of helix crossings [23–25]. This allows helix packing, providing favorable hydrophobic interaction in the confined space between adjacent helices. In the structure of $_{SA}$ DUF3055, Gly76 on $\alpha 3$ is close proximity to $\alpha 2$, enables helix crossing (Fig. 2B). Through this arrangement, three α -helices are tightly packed into a narrow space with the charged side chains of helices oriented toward the solvent region. Structural analysis by the DALI server revealed a very low Z-score (less than 4.3 for top scored structure), indicating that $_{SA}$ DUF3055 has a unique fold [26].

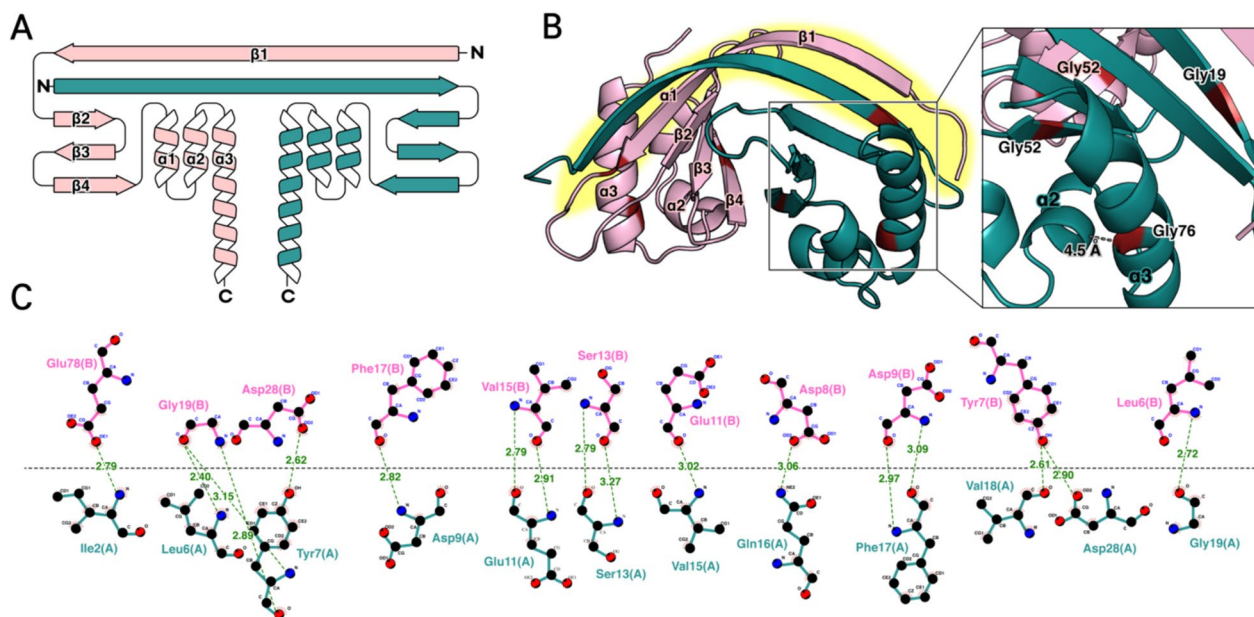


Fig. 2 Structural overview and secondary structure topology of $_{SA}$ DUF3055. **A** Secondary-structure diagram of the $_{SA}$ DUF3055 dimer. The two monomers are colored light and dark cyan. **B** Crystal structure of the $_{SA}$ DUF3055 dimer in ribbon representation. The glycine residues that are involved in secondary structure elements are indicated as red and showed in an enlarged inset. The two α -helices, $\alpha 2$ -helix and $\alpha 3$ -helix are separated by a short distance of 4.5 Å. **C** Visualization using LIGPLOT depicts the amino acid residues at the dimer interface, with green dashed lines indicating hydrogen bonds

Structure and function relationship of *S_A*DUF3055

Our previous investigation asserted that *S_A*DUF3055 may form a long chain polymer as part of its function, owing to observations from the crystal packing. However, the crystallographic oligomer does not correlate with the solution state. The altered packing observed in our new crystal form suggests *S_A*DUF3055 most likely functions as a dimer. Indeed, analysis using PISA confirms the likely solution state is dimer with a calculated ΔG of -23.5 , -21.2 , -23.3 kcal/mol, for monomer pairs AB, CD and EF respectively.

Analysis of the surface area shows a prominent negative charge due to solvent exposed aspartate and glutamate residues. In total, all 17 negatively charged residues (7 aspartates and 10 glutamates) have side chains displayed on the protein's exterior. Comparing the sequence of *S_A*DUF3055 to homologs in other bacteria show that

these negatively charged residues are well conserved and aligned (Fig. 3A). Moreover, the residues occur as consecutive pairs in the amino acid sequence or are localised within the 3D structure. These characteristics are often observed in nucleic acid related functional proteins [27]. This observation, together with the curved shape afforded by the β -bulge would allow for the protein to mimic the DNA major groove, showing 34 Å of width (Fig. 3B). As the dimer imparts a symmetry operator onto the protein known DNA binding motifs such as a zinc finger and helix-turn-helix would be accommodated by *S_A*DUF3055. Using ClusPro, a reasonable binding model between *S_A*DUF3055 and a helix-turn-helix transcription factor can be generated [32]. Substantial hydrogen bonds and ionic interactions are found in the proposed representation, suggesting potential protein-protein interaction (Fig. 4).

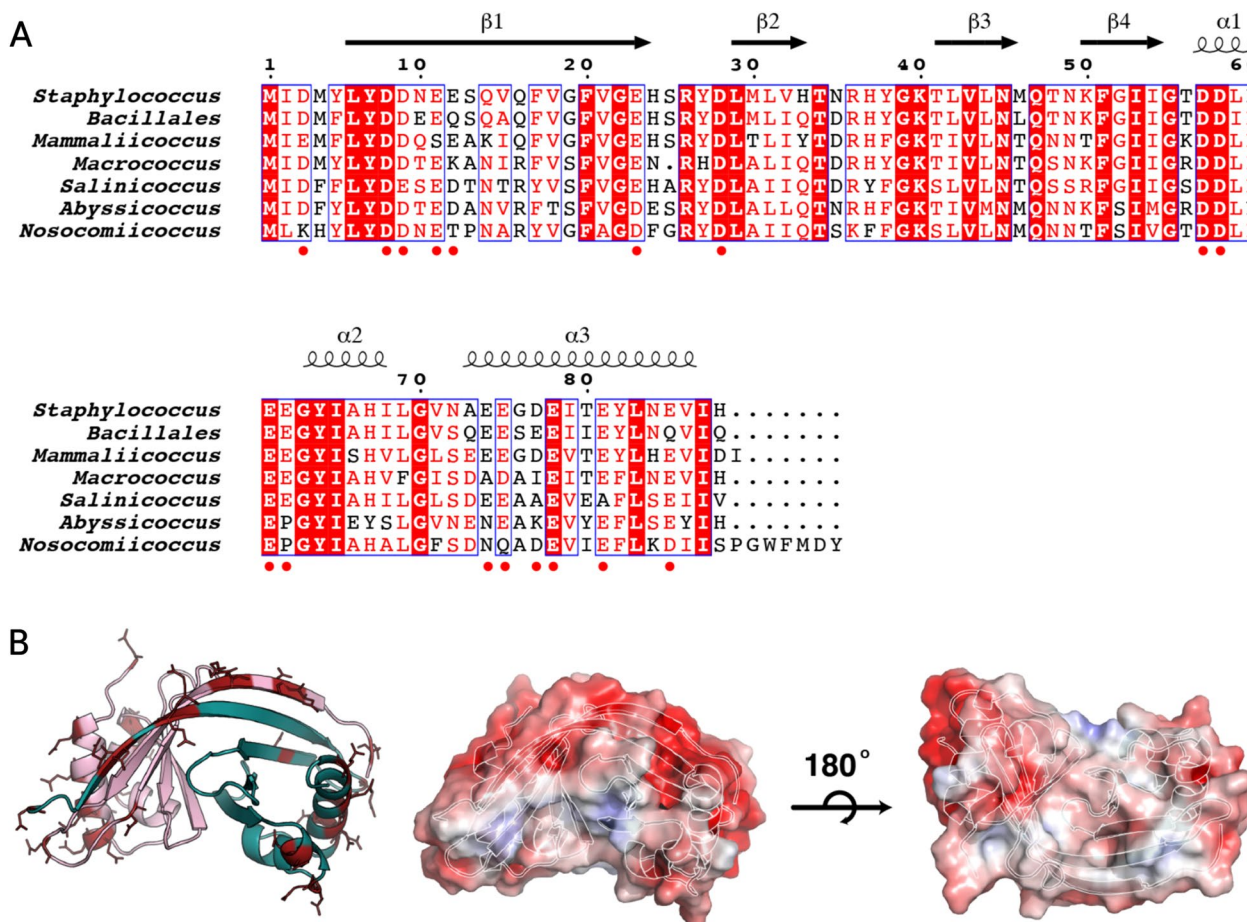


Fig. 3 Sequence alignment and protein electrostatics. **A** Sequence alignment of DUF3055 from different bacterial species. Identical residues are colored white on a red background and similar residues are red on white background. Secondary structure elements (springs, α -helices; arrows, β -strands) are shown above. Negatively charged residues in *S_A*DUF3055 is denoted as red dots. Sequence alignments were performed using ClustalX and visualized by ESPript [30, 31]. **B** Side chains for residues denoted as red dots in **A** are represented as sticks. All shown side chains are heading toward the solvent area. Corresponding electrostatic potential at the molecular surface of *S_A*DUF3055. The negatively charged surface is shown as front and flipped view (by 180° rotation)

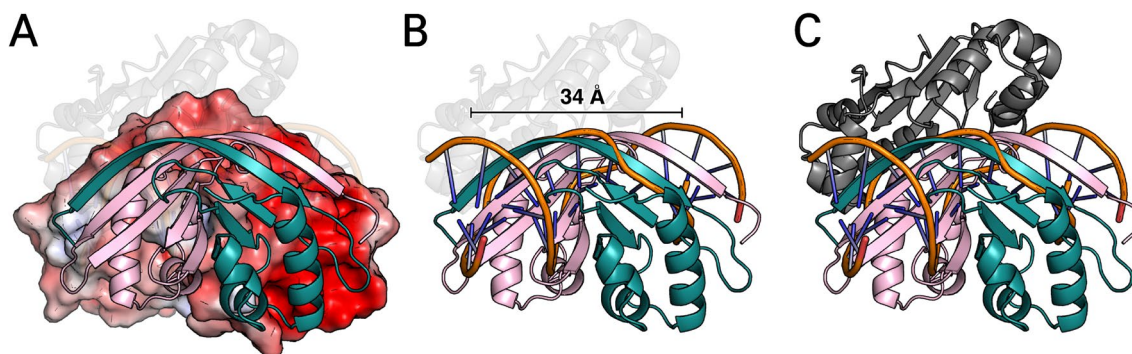


Fig. 4 Proposed function of $_{SA}DUF3055$. **A** Surface representation of $_{SA}DUF3055$. The negative charges are distributed on the β -bulge region. **B** An alignment of typical B-DNA with $_{SA}DUF3055$ β 1-strand. **C** Refined ClusPro docking simulation with a common transcription factor from our previous study [32]. Favored docking model interacts directly with $_{SA}DUF3055$

Discussion

$_{SA}DUF3055$ is a well-conserved hypothetical protein present throughout pathogenic *staphylococcus* with an as yet uncharacterised role. Our previous study elucidated the structure with an alternative packing state that suggested the protein formed a linear polymer [28]. However, the packing adopted by our new structure strongly suggests the native form of the protein is likely a dimer. Our structural analysis shows $_{SA}DUF3055$ has a highly negative electrostatic surface which is conserved in closely related homologs. From its shape, size, and charge, we surmise that $_{SA}DUF3055$ interacts with DNA binding proteins. $_{SA}DUF3055$ acts as a DNA-mimetic, with the length of the major beta strand ~ 34 Å, mirroring that of a B-DNA single turn. Further studies such as pull-down assay to find partner molecule, DNA dissociation activity test to compare the binding affinity, or RNA sequencing understand its network are expected to confirm $_{SA}DUF3055$ function.

Staphylococci represent a threat to public health with species such as *S. aureus*, *S. epidermis*, or *S. haemolyticus* gaining multi-drug resistances [29]. $_{SA}DUF3055$ is highly conserved in many species of *Staphylococci*, implying this would perform a key role in its pathogenicity. Whilst the structure has elucidated the likely function, a subsequent microbiological investigation would be required to determine the protein's impact on virulence phenotypes. This study presents structural analysis, which can widen our understandings regarding bacterial physiology.

Acknowledgements

We express gratitude to the beamline staff at Photon Factory, Japan, for their assistance during X-ray diffraction experiments.

Author contributions

H.J.K. and H.S.K. designed, conducted experiments, analyzed the data, and wrote the manuscript.

Funding

This paper was supported by Woosuk University.

Availability of data and materials

The protein coordinate and structure factor have been deposited to the RCSB PDB under the code 8XFU.

Declarations

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Received: 21 March 2024 Accepted: 12 June 2024

Published online: 02 July 2024

References

- UniProt C (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 51:D523–D531
- Collaborators, G. B. D. A. R (2022) Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 400:2221–2248
- Vestergaard M, Frees D, Ingmer H (2019) Antibiotic resistance and the MRSA problem. *Microbiol Spectr* 7:10
- Hasan R, Acharjee M, Noor R (2016) Prevalence of vancomycin resistant *Staphylococcus aureus* (VISA) in methicillin resistant *S. aureus* (MRSA) strains isolated from burn wound infections. *Ci Ji Yi Xue Za Zhi* 28:49–53
- Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K (2001) Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* 357:1225–1240
- Suzuki H, Lefebvre T, Bitar PP, Stanhope MJ (2012) Comparative genomic analysis of the genus *Staphylococcus* including *Staphylococcus aureus* and its newly described sister species *Staphylococcus simiae*. *BMC Genomics* 13:38
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–301
- Goodacre NF, Gerloff DL, Uetz P (2013) Protein domains of unknown function are essential in bacteria. *MBio* 5:e00744–e813

10. Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614
11. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
12. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326
13. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242
14. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674
15. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132
16. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355–367
17. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* 1:2876–2890
18. Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C (2012) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* 80:374–381
19. Micsonai A, Moussong E, Wien F, Boros E, Vadaszi H, Murvai N, Lee YH, Molnar T, Refregiers M, Goto Y, Tantos A, Kardos J (2022) BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res* 50:W90–W98
20. Siepen JA, Radford SE, Westhead DR (2003) Beta edge strands in protein structure prediction and aggregation. *Protein Sci* 12:2348–2359
21. James LC, Jones PC, McCoy A, Tennent GA, Pepys MB, Famm K, Winter G (2007) Beta-edge interactions in a pentadecameric human antibody V kappa domain. *J Mol Biol* 367:603–608
22. Craveur P, Joseph AP, Rebehmed J, de Brevern AG (2013) beta-Bulges: extensive structural analyses of beta-sheets irregularities. *Protein Sci* 22:1366–1378
23. Javadpour MM, Eilers M, Groesbeek M, Smith SO (1999) Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. *Biophys J* 77:1609–1618
24. Otaki JM, Tsutsumi M, Gotoh T, Yamamoto H (2010) Secondary structure characterization based on amino acid composition and availability in proteins. *J Chem Inf Model* 50:690–700
25. Dong H, Sharma M, Zhou HX, Cross TA (2012) Glycines: role in alpha-helical membrane protein structures and a potential indicator of native conformation. *Biochemistry* 51:4779–4789
26. Holm L, Laiho A, Toronen P, Salgado M (2023) DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci* 32:e4519
27. Chou CC, Wang AH (2015) Structural D/E-rich repeats play multiple roles especially in gene regulation through DNA/RNA mimicry. *Mol Biosyst* 11:2144–2151
28. Jeong S, Kim HJ, Ha NC, Kwon AR (2019) Crystal structure of SAV0927 and its functional implications. *J Microbiol Biotechnol* 29:500–505
29. Ben Zakour NL, Guinane CM, Fitzgerald JR (2008) Pathogenomics of the staphylococci: insights into niche adaptation and the emergence of new virulent strains. *FEMS Microbiol Lett* 289:1–12
30. 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 (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
31. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324
32. Richardson W, Kang GW, Lee HJ, Kwon KM, Kim S, Kim HJ (2021) Chasing the structural diversity of the transcription regulator *Mycobacterium tuberculosis* Hg A2. *IUCrJ* 8:823–832

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.