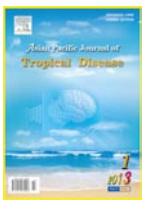
Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Disease

journal homepage: www.elsevier.com/locate/apjtd



Document heading doi:

Homology modeling and validation of SAS2271 transcriptional regulator of AraC family in *Staphylococcus aureus*

Money Gupta[1], Gulshan Wadhwa[2], Sanjeev Kumar Sharma[1], Chakresh Kumar Jain[1*]

[1]*Department of Biotechnology, Jaypee Institute of Information Technology, A-10, Sector-62, NOIDA, UP, 201301, India*

[2]*Department of Biotechnology (DBT), Ministry of Science & Technology, New Delhi-110003, India*

PEER REVIEW

Peer reviewer

Prof. Dr. R. N. Padhy, PhD, Post-Doc (ICP, Brussels), CSIR Scientist, Central Research Laboratory, IMS & Sum Hospital, Siksha 'O' Anusandhan University, K-8, Kalinga Nagar, Bhubaneswar 751003, Odisha, India.
Tel/Fax: +916742432034
Email: rnpadhy54@yahoo.com

Comments

This is a good study. The use of high-end techniques in establishing the 3D structures of virulent proteins would help immunologists and infection biologists to understand the pathogenicity of virulent bacteria, as done herein. It would encourage other clinicians and researchers to do such studies with other pathogens.
(Details on Page 4)

ABSTRACT

Objective: To predict three dimensional structure of AraC Family transcription regulator protein in *staphylococcus aureus* involved in causing virulence. **Methods:** Evolutionary dynamics of *S. aureus* reveled that several mutations preceding virulence lead to truncated proteins that plays an important role in virulence. The Structural templates are identified using homology search and then homology modelling is used to get the 3-D structure of the protein. **Results:** The 3-D structure of SAS2271 transcriptional factor of AraC family in MSSA476 strain of *S. aureus* was modelled and validated using the Ramachandran plot. **Conclusions:** The knowledge of 3-D structure of the protein will be helpful in identifying its biochemical function along with its regulatory mechanism in causing virulence.

KEYWORDS

AraC family transcriptional regulator (AFTRs), Modeller, PROCHECK.

1. Introduction

Staphylococcus aureus (S. aureus) is a well known opportunistic bacteria which causes disease in humans. *S. aureus* is a major human pathogen that causes cutaneous infection, food poisoning and dangerous septicaemia[1]. The whole genome sequencing of *S. aureus* revealed the evolutionary dynamics associated with the causes of bacterial disease. Several mutations or alterations are observed in the genome of *S. aureus* from nasally carried bacteria to the bloodstream bacteria which is predicted to have introduced some virulence factor in bacteria. The debate on this transition is still on but it's clear that the ability of *S. aureus* to cause virulence depends upon its ability to modulate the gene expression according to change in the environmental factors around the organism[2]. Various comparative genomics approaches are being used to explore the mechanisms of evolution of *S. aureus* genomes, so as to identify the regions affecting virulence and drug resistance. Along with several other mutations one is found in the AraC family transcriptional

*Corresponding author: Dr. Chakresh Kumar jain, Department of Biotechnology, Jaypee Institute of Information Technology, A-10, Sector-62, NOIDA, UP, India.
Tel: +91-120 2400973
Fax: +91-120 2400986
E-mail: ckj522@yahoo.com

Article history:
Received 1 Nov 2012
Received in revised form 8 Nov, 2nd revised form 13 Nov, 3rd revised form 17 Nov 2012
Accepted 28 Dec 2012
Available online 28 Feb 2013

regulator (AFTRs) (MSSA476 SAS2271) which led to pre-mature stop codon in the protein sequence thus truncating it from 701 to 76 amino acids[3]. Studies have shown the involvement of AFTRs in regulating carbon metabolism, stress response and virulence factor[4]. Thus, in order to completely understand the mechanism of virulence in *S. aureus*, further investigations are needed to carry out on AFTRs. The knowledge of its three dimensional structure will be helpful in determining the basics of its biochemical functions along with the cause and the treatment of diseases. Here, we have used the homology modelling to determine the 3-dimensional structure of the protein. Homology modelling is used to predict the 3 dimensional structures of proteins with unknown 3D structure, using solved homologous proteins as templates. Homologues can be determined using various tools such as Basic Local Alignment Search Tool (BLAST)[5]. It identifies the region of local similarity between the query and database sequences and calculates the statistical importance of the matches. It is useful for inferring functional and evolutionary relationships between sequences. PSI-BLAST derives position-specific scoring matrices (PSSM) from multiple sequence alignment. Since, homology modelling depends upon the sequence alignment of target sequence with the template sequence, there many tools available to check the alignment. Needleman wunsch and smith waterman algorithm can be used for pair-wise sequence alignment. They perform global and local alignment respectively to align protein and nucleotide sequences and generate a similarity matrix. Multiple sequence alignment is used to align three or more biological sequences (protein or nucleic acid) of similar length. The output of the alignment can be used to infer homology and the evolutionary relationships between sequences. From bioinformatics prospective, there are several tools available for multiple sequence alignment such as ClustalW, Clustal Omega and T-coffee.

2. Material and methods

2.1. Query sequence from NCBI

AraC family transcriptional regulator (MSSA476 SAS2271) amino acid sequence has been retrieved from the NCBI-Protein. This protein database is a collection of sequences from various sources, including SWISSPROT, Protein Information Resource(PIR), Protein Data Bank (PDB), Genpept and RefSeq. This sequence acted as query which is to be modelled based upon templates taken from the homology search.

2.2. Homology search and modelling

Template sequences for modelling are obtained from BLASTP (Basic Local Alignment Search Tool-Protein) for homology search against PDB (Protein Data Bank)[6,7]. The templates are used to predict the three dimensional structure using modeller 9v11. The modelling involves four basic steps, first searching structures showing homology with the target, then selecting a best template having maximum identity with the target sequence which follows its alignment with the target and modelling the structure using MODELLER. The modelled structure is then evaluated using PROCHECK.

3. Results

AraC transcription factor is subjected to homology search in blastp against pdb to determine its structural template. Homology search showed maximum 32% identity with the 1T8B, a transport protein in Escherichia coli with the target sequence since important difference found (P=0.0072) found, we used multiple structural templates for the homology modelling in MODELLER9v11[8]. It is a molecular graphics visualization computer program primarily used for exploration of biological macromolecule structures such as of proteins, as shown in Figure 1. The structure is evaluated using the Ramachandran plot shown in Figure 2.

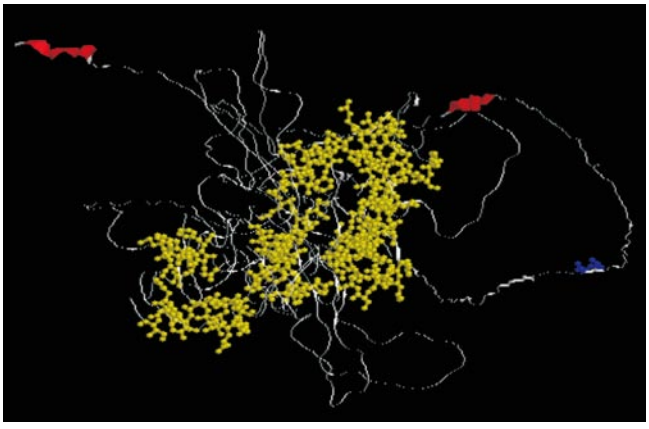


Figure 1. Modelled structure showing helices (yellow, ball and stick), sheet region (red, ribbon) and mutated position, 77th amino acid (blue).

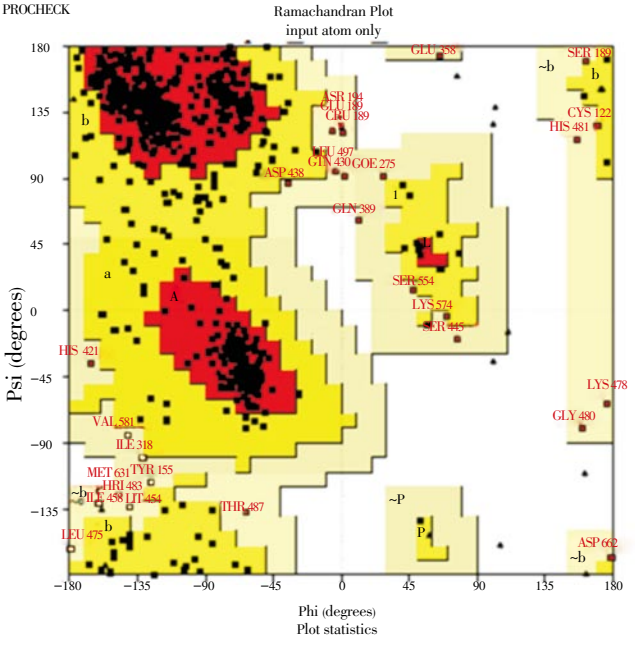


Figure 2. Ramachandran plot of the modelled structure.

The modelled structure in Figure 3 shows the presence of Helix turn helix, DNA binding motif from 150 to 248 amino acid as identified using PROSITE[15]. As expected, HTH motif was absent in truncated 76 amino acid structure Figure 4.

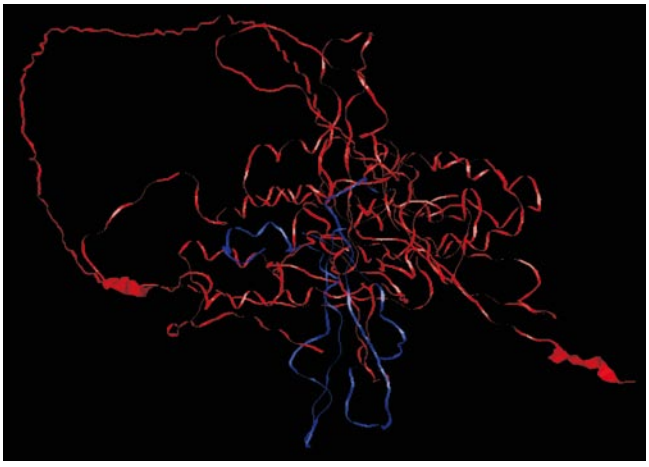


Figure 3. Modelled structure showing Helix Turn Helix DNA binding motif (Blue).

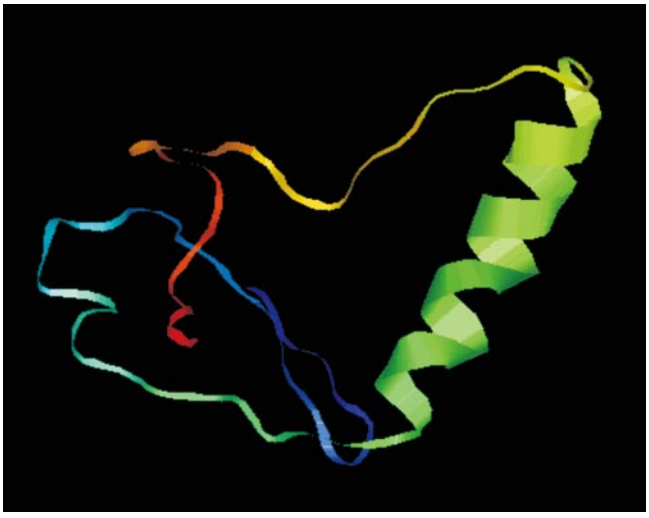


Figure 4. Truncated protein structure exhibits no DNA binding motif.

4. Discussion

Ramachandran plot shows that the predicted model has 100% amino acids in allowed region after loop refinement, hence confirming the reliability of the structure. Discrete Optimized Protein Energy (DOPE) is used by the MODELLER to assess the quality for all the predicted models. It is used to assess the homology models by calculating statistical potential. The predicted structure has an important role in determining the biological importance of this protein and its importance in causing virulence. The rapid increase in the protein numbers has made it impossible to determine each 3-D structures experimentally, therefore the role of *in silico* modelling has become important, it is generated without or with very limited experimentally compulsion, it mainly depends on the various hypotheses included in the modelling process, on the force-field used in the simulations as well as on the quality of the scientific computing tools. Homology modelling has failed to determine good quality models with low identity homologue templates such in this case, MSSA476 SAS2271 has maximum of 32% identity with 1T8B thus generating a low quality model even after using multiple structural templates. The quality of an experimentally predicted model can be accessed directly against the experimental data, but the quality of an *in silico* model is more subjective and ultimately defined through the usefulness of the model. Although experimental methods both X-ray crystallography and NMR spectroscopy are also prone to errors in validation the model and requires proper validation methods in structural biology[9]. The structure of AraC family transcriptional regulator will be of immense help in studying the basics of its regulation in causing virulence. There are several possibilities of further refinement of this protein structure in the near future for improving its quality. Considering the need of hour it is important to functionally characterize this protein and for that matter knowledge of 3-D structure is the best option. We can structurally analyze its active sites and binding pattern with other proteins involved in transcription or directly with the DNA. Comparative genomic studies has reveled the evolution of drug resistance in *S. aureus* such as in community acquired Meticillin-resistant Staphyloccocus aureus (MRSA) which will make it more difficult to curb diseases caused by drug resistant strains of *S. aureus*. Therefore, it is an urgency to identify the virulent factor in these organisms and inhibit them at the molecular level. AFTRs are one of the largest groups of transcription regulatory proteins found in bacteria[10]. Function of many AFTRs are identified in various bacteria involved in causing pathogenesis such as in *Mycobacterium tuberculosis* where AraC family transcriptional regulator Rv1931 experimentally verified to be involved in causing virulence[11]. Rsp and Rbf are AraC family regulator present in *S. aureus* MW2 strain are involved in biofilm formation. Rbf is known to form biofilm formation under the control of glucose and NaCl[12]. And Rsf on the contrary represses the biofilm formation because of the repression of the FnbA[13]. Formed as a result of mutation in the gene, thus making it inactive as a transcriptional regulatory protein. The structure of a protein determines its biological improce making it, important to predict the structure of MSSA476 SAS2271[14]. Homology modelling is based on the fact that related amino acids have similar 3-D structure. Its popularity is growing keeping in mind the large amount of proteins being identified and the limitations of X-ray Laboratories to resolve 3-D structures[14]. Several automated homology modelling servers are available, in SWISS-MODEL template selection, alignment and model building are done completely automated by the server[15]. Model validation is done using the PROCHECK, which checks the stereochemical quality of a protein structure, producing a number of PostScript plots analyzing its overall and residue-by-residue geometry along with the Ramachandran plot.

There have been great advancements in bioinformatics over the last decade and it has certainly improved the sensitivity of the tools used in homology modelling. But improving the quality of your predicted structure is still a tough task and requires further studies so as to match the experimentally verified 3-D

models using X-ray and NMR. The above work is entirely based on *in silico* methods and can be useful in curbing virulence and destroying its cause. Since, quality of the homology model depends upon the percentage sequence identity with the template as the identity decreases the quality of the model also decreases. Thus, for better model we should have at least 60% identity with the template. MSSA476 SAS2271 structure with maximum 32% identity only opens the sources of errors and the possibilities of further methodological improvements in the future.

Conflict of interest statement

We declare that we have no conflict of interest.

Comments

Background

S. aureus is considered as the "superbug" of the health domain, as its drug-resistant strains (MRSA and VRSA) have escalated throughout the world in both nosocomial and community settings. Studies are required to link between the drug resistance and the virulent factors of bacteria. So, this would help the infection biologists to find suitable options for controlling the "superbug" with help of OMICS.

Research frontiers

The use of tools of bioinformatics in locating the 3D structure of a protein responsible for virulence of a critical bacterium is one of the praise worthy approaches.

Related reports

A review by, Coburn *et al.* 2008, describes the role of AraC-Type transcriptional regulator protein in the increase of virulence and pathogenicity, by escaping the intracellular phagocytosis in *E. faecalis*, another Gram-positive pathogen.

Innovations & breakthroughs

"*In silico*" modeling of proteins, though may not be accurate in predicting the structure, and it may require further techniques like X-ray crystallography, or NMR technique to authenticate the findings, but still it is of worth for researchers to understand the infection mechanism of a bacterium. Hence, this work is of great clinical importance and such work should become the base for further studies on infection dynamics of pathogenic bacteria.

Applications

Ara-C is one of the largest groups of transcription regulatory proteins found in bacteria related to their expression of virulence. The comparative genomic studies would help in establishing links with drug resistance genes in the bacterial transformation of commensal to a virulent pathogen. This organism was originally a commensal and now it has become a ghoulish, intractable pathogen all over. This study and some other related studies suggest the role of this protein in biofilm production. Therefore, it becomes mandatory to find the role of this protein in other pathogenic bacteria by establishing such 3D '*in silico*' structures.

Peer review

This is a good study. The use of high-end techniques in establishing the 3D structures of virulent proteins would help immunologists and infection biologists to understand the pathogenicity of virulent bacteria, as done herein. It would encourage other clinicians and researchers to do such studies with other pathogens.

References

[1] Yue-Qing Zhang, Shuang-Xi Ren, Hua-Lin Li, Yong-Xiang Wang, Gang Fu, Jian Yang, et al. Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228). *Molecular Microbiol* 2003; **49**(6): 1577-1593.

[2] Yang J, Tauschek M, Robins-Browne RM. Control of bacterial virulence by AraC like regulators that respond to chemical signals. *Trends Microbiol* 2011; **19**: 128-135.

[3] Younga BC, Golubchikb T, Battyb EM, Funga R, Svenssond HL, Votintsevaa AA, et al. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *PNAS* 2012; **109**(12).

[4] Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol* 1997; **61**: 393-410.

[5] Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Thomas L. Madden NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008; **36**: 5-9.

[6] Boeckmann B, Blatter C, Famiglietti L, Hinz U, Lane L, Roechert B, et al. *Comptes Rendus Biol* 2005; **328**: 882-899.

[7] Altschul S, Madden L, Schaffer A, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**(17): 3389-3402.

[8] Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Meth Enzymol* 2003; **374**: 461-491.

[9] Luccio E, Koehl P. A quality metric for homology modelling: the *H*-factor. *BMC Bioinformatics* 2011; **12**: 48.

[10] Yang J, Tauschek M, Roy M. Robins-Browne control of bacterial virulence by AraC-like regulators that respond to chemical signals. *Trends Microbiol* 2011; **19**(3); 128-135.

[11] Frota CC, Papavinasasundaram KG, Davis EO, Colston MJ. The AraC family transcriptional regulator Rv1931c plays a role in the virulence of *Mycobacterium tuberculosis*. *Am Soc Microbiol* 2004; **72**(9): 5483-5486.

[12] Lim Y, Jana M, Luong TT, Lee CY. Control of glucose- and NaCl-induced biofilm formation by rbf in *Staphylococcus aureus*. *J Bacteriol* 2004; **186**: 722-729.

[13] Lei MG, Cue D, Roux CM, Dunman PM, Lee CY. Rsp inhibits attachment and biofilm formation by repressin fnbA in *Staphylococcus aureus* MW2. *J Bacteriol* 2011; **193**: 5231-5241.

[14] Bhagavathi S, Wadhwa G, Prakash A. *In silico* modelling and validation of differential expressed proteins in lung cancer. *Asian Pac J Trop Dis* 2012; **2**(suppl2): S524-S529.

[15] R Rodriguez, Chinea G, Lopez N, Pons T, Vriend G. Homology modelling, model and software evaluation: three related resources. *Bioinformatics* 1998; **14**(6): 523-528.

[16] Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res* 2009; **37**: 387-392.