Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 3, Issue, 8, pp.027-031, August, 2011

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

INSILICO STRUCTURE PREDICTION OF SaeR PROTEIN FROM *Staphylococcus aureus*

Pitchai Daisy*, Periyasamy Vijayalakshmi and Kandasamy Saipriya

Bioinformatics centre (BIF), PG& Research Department of Biotechnology & Bioinformatics, Holy Cross College (Autonomous), Tiruchirapalli-620002, Tamilnadu, India.

ARTICLE INFO

Article History:
Received 8th April, 2011
Received in revised form
6th May, 2011
Accepted 5th June, 2011
Published online 5th August, 2011

Key words:
S.aureus, saeR, Modeller,
Homology modeling,
Procheck,
Ramachandran plot.

ABSTRACT

Staphylococcus aureus (*S. aureus*) is a common gram-positive pathogen, which predominates as a global cause of bacterial infections. *S. aureus* uses the saeRS- two component system to control the expression of many virulence factors. SaeR the response regulator is one of the potent virulence factor that can cause many bacterial infections. So far the structure of saeR is not predicted. In this paper molecular modeling techniques were done using MODELLER 9v8 software to construct the 3 dimensional structure of SaeR protein with the help of a template which have the sequence identity more than30%. Structural characterization of this protein is important in rational drug design. The stereo chemical quality of the best model is validated by PROCHECK server and the Q site finder is used for locating, delineating and measuring concave surface regions on three-dimensional structure of the protein. Our results provide a framework for understanding the structure and the possible binding sites of saeR protein for drug targeting and the results were found to be reliable.

INTRODUCTION

Bacterial infections are usually a consequence of successful colonization of the host as well as breaching the host's immune defence- system. Gram positive bacterial infections account for ~50% of all reported sepsis and are associated with the dysfunctional production of pro- inflammatory cytokines (Matin,2003, Lappin, 2009, Hotchkiss, 2008, Rittirsch, 2008). *Staphylococcus aureus* (*S.aureus*) is a gram positive bacterium which is an important human and animal pathogen. Human and animals are the natural reservoir of *S.aureus*. It causes wide variety of infections ranging from mild skin infections to, life- threatening diseases such as necrotizing pneumonia and bacteremia. *S.aureus* is a versatile bacterial pathogen that can survive in diverse host environments. This versatility depends on its ability to acquire and utilize nutrients from different sources. Bacteria can sense nutrients and respond by modulating gene expression including the synthesis of virulence factors. Therefore, the metabolic signals encountered by bacteria not only help them in survive in temporary adverse conditions, but also play a crucial role in pathogenesis. *S.aureus* pathogenicity is dependent on the co-ordinated action of a number of virulence factors and the expression of these virulence factors is determined by several global regulators.

Corresponding author: daisylesslie@yahoo.com

The broad range of infections caused by *S.aureus* is, at least in part, a consequence of its many virulence factors. The virulence is generally considered to be multifactorial and due to the combined action of several virulence factors. sae is a key regulatory locus in *staphylococci*, coordinating environmental signals with the internal regulatory circuitry governing virulence and other adaptive processes (Novick, 2003). It was identified on the basis of a Tn551 insertion in saeR (Giraudo, 1994) which profoundly affected the expression of a large set of virulence genes (Giraudo, 1997). The Sae locus contains a classical two-component signalling module (TCS), of which saeS is the receptor kinase and saeR is the response regulator (Giraudo, 1999). SaeR is a 228 amino acid polypeptide with an N-terminal regulatory domain and a C-terminal effector domain with potential DNA binding activity. The saeR protein along with saeS protein plays an important role in *staphylococcal* gene expression and virulence.

Proteins are essential to biological processes. They are responsible for catalyzing and regulating biochemical reactions. Protein function can be understood in terms of its structure. Indeed, the three dimensional structure of a protein is closely related to its biological function. Functional characterization of a protein sequence is one of the most frequent problems in biology. In the absence of experimentally determined structure, comparative or homology modeling

often provides a useful 3-D model for protein that is related to atleast one known protein structure (Browne, 1969, Blundell, 1987, Anchez, 1997, Bajorath, 1994, Johnson, 1994, Mart, 2000). Homology modeling combines sequence analysis and molecular modeling to predict three dimensional structures. Homology modeling is also known as comparative modeling of protein refers to constructing an atomic resolution model of the target protein from its amino acid sequence and an experimental 3 dimensional structure of a related homologous protein (template). The quality of the homology model is dependent on the quality of sequence alignment and template structure. The structure of saeR protein has not yet determined experimentally (X-ray or NMR) and therefore models are built following homology modeling protocol through computational methods. In the current study, we predicted the 3D structure of the saeR protein using homology modeling approach which can serve as a potential target for drug development against *S.aureus* in reducing the virulence. In addition, binding site of saeR protein is also predicted in order to eventually identify drug-like molecules that possess enhanced binding to the receptor.

MATERIALS AND METHODS

A. Homology Modeling

The homology modeling was used to build the initial model of SaeR protein. Homology modeling of the Sae R was performed using Modeller 9v8 software. The amino acid sequence of SaeR from *S.aurues* was retrieved from swissprot (Bairoch,2000) protein sequence database (accession number: Q2FIT4). It consists of 228 amino acids. The first step in homology modeling was searching a number of related sequences to find a related protein as a template. The protein sequence was then subjected to a PSI-BLAST search (Altschul,1997) in order to identify the homologous protein from the Brookhaven Protein Data Bank (PDB) (Berman,2003). An appropriate template for SaeR was identified based on the e-value and sequence identity. The template and the target sequences were then aligned using ClustalW (Thompson,2002). Subsequently, homology modeling was carried out for the SaeR protein against the chosen template using Modeller 9v8 (Sali, 1993). The outcomes of the modeled structures were ranked on the basis of an internal scoring function, and those with the least internal scores were identified and utilized for model validation.

B. Validation of the modeled structure

After building the protein 3D structure, in order to assess the overall stereo chemical quality of the modeled protein, Ramachandran plot analysis was performed using the program PROCHECK (Fiser, 2000, Laskowski, 1993, Morris, 1992). Further evaluation of modeled structure was done by using the ERRAT (Ramachandran, 1963). The quality of the consistency between the template and the modeled saeR protein was evaluated using ProSA (Wiederstein, 2007) during which the energy criteria for the modeled structure were compared with the potential mean force obtained from a known protein structure. To assess the reliability of the modeled structure of saeR protein, we calculated the root mean square deviation (RMSD) by superimposing it on the template structure using a Chimera.

C. Protein structure accession number

The refined homology model of 3D structure of SaeR protein from *Staphylococcus aureus* was submitted to PMDB (http://mi.caspur.it/PMDB/) (Castrignano, 2006) and the same was assigned the identifier PM0077444.

D. Active Site Identification

Sites of binding in proteins usually lie in cavities or on the polar surface. The size and shape of protein surface dictates the three- dimensional geometry of ligand that must binds due to intra atomic forces between them. The binding of a ligand typically serves as a mechanism for chemical modification or conformational change of protein (Ramachandran, 1963)**.** Identifying the location of ligand binding sites on a protein is of fundamental importance for a range of applications including molecular docking, *de novo* drug design and structural identification and comparison of functional site (Laurie, 2005). So the binding pockets of SaeR protein from *S.aureus* were identified using Q Site Finder. This program is used for indentifying and characterizing the protein active sites, binding sites, and functional residues located on protein surfaces and voids, buried in the interior of protein by measuring concave surface regions on three dimensional structures of proteins. It also measures the area and volume of the pockets.

RESULTS AND DISCUSSION

A. Sequence Alignment between Target and Template

Homology modeling produced high quality structural models when the target and template are closely related. Protein structure modeling is undertaken in this present work. The main criteria in homology modeling are template selection and sequence alignment between the target and the template. BLAST searching result shows that, the sequence identity between the SaeR and the template (PDB ID:2OQR) is 36% which allows for rather straightforward sequence alignment. The protein sequences of target (SaeR) and template (PDB ID-2OQR) were aligned and the result of alignment is shown in figure 1. The asterisk showed the identity of amino acids present in two protein sequences.

B. Homology modeling of SaeR protein and its evaluation

Total 5 models were generated after performing homology modeling with modeller 9v8. Dope scores of the generated models were calculated using the command model-single.py. The model TvLDH.B99990001.pdb**,** having minimum dope score was considered as the best model of protein SaeR (Table 1).

Table 1: Dope Score of successfully produced models

S.NO	File Name	DOPE Score
1	TvLDH.B99990001.pdb	-22877.26953
2	TvLDH.B99990002.pdb	-22769.83594
3	TvLDH.B99990003.pdb	-22650.24023
4	TvLDH.B99990004.pdb	-22706.46289
5	TvLDH.B99990005.pdb	-22555.52344

```
Target      ---MTHLLIVDDEQDIVDICQTYFEYEGYKVTTTTSGKEAISLLS-NDIDIMVLDIMMPE 56
Template    GAMATSVLIVEDEESLADPLAFLLRKEGFEATVVTDGPAALAEFDRAGADIVLLDLMLPG 60
                 * :***:**:.:.*      :. **::.*..*.*  *:: :.  . **::**:*:*

Target      VNGYDIVKEMKRQKLDIPFIYLTAKTQEHDTIYALTLGADDYVKKPFSPRELVLRINNLL 116
Template    MSGTDVCKQLR-ARSSVPVIMVTARDSEIDKVVGLELGADDYVTKPYSARELIARIRAVL 119
            :.* *: *::: : .:*.* :**: .* *.: .* *******.**:*.***: **. :*

Target      TRMKKYHHQPVEQLSFDELTLINLS-KVVTVNGHEVPMRIKEFELLWYLASRENEVISKS 175
Template    RRGGDDDSEMSDGVLESGPVRMDVERHVVSVNGDTITLPLKEFDLLEYLMRNSGRVLTRG 179
             *  . . :  : :  .  . :::. :**:***. :.: :***:** **  ....*:::.

Target      ELLEKVWGYDYYEDANTVNVHIHRIREKLEKESFTTYTITTVWGLGYKFERSR 228
Template    QLIDRVWGADYVGDTKTLDVHVKRLRSKIEADPANPVHLVTVRGLGYKLEG-- 230
            :*:::*** **  *::*::**::*:*.*:* :. ..  :.** ******:*
```

Fig 1. Alignment between Target and Template sequence obtained from Clustal w

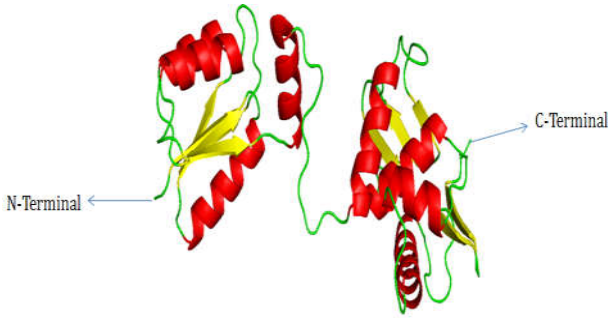

Fig 2. Shows the modelled structure of SaeR protein obtained from Modeller9v8.Red colour indicate alpha helices, yellow colour indicate the beta sheet and green colour indicate the loops

Fig 3. Superimposition of Modeled SaeR protein (Target) and Template 2OQR using Chimera. In this picture pink colour represents target and white represents Template

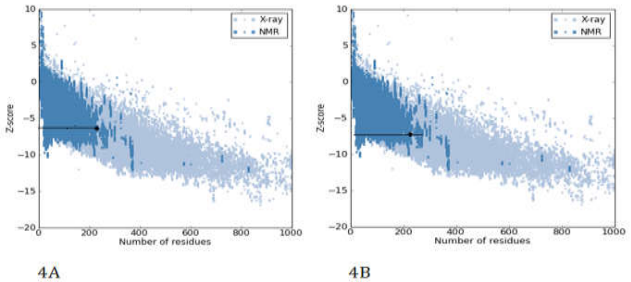

Fig 4. 4A) This plot that contains the Z-scores of experimentally determined protein chain in the current PDB that have been solved by either X-ray diffraction or NMR. The Z-score of -6.36 indicated on this graph represents the overall quality of the modeled 3D structure of SaeR. This plot can be used to check whether the Z-score of this 3D structure is within the range of scores typically found for native proteins of a similar size. (4B) The Z-score of −7.24 indicated on this graph represents the overall quality of the template's 3D structure (PDB ID 2OQR). When compared to Figure. 4A, it was found that the overall quality of the 3D structure of the template and that of the target are very similar in terms of their Z-scores.

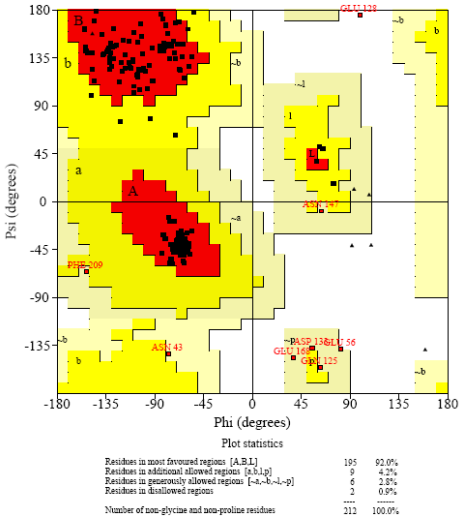

Fig 5. Ramachandran plot of SaeR protein obtained from PROCHECK

The best ranked model (Figure. 2) was then subjected to evaluation to check the quality of the generated model. The modeled SaeR protein was used for further optimization and validation. The calculated root mean square deviation between the target and template structure was found to be 0.506 Å (Figure. 3).

Table 2: Comparative analysis of Ramachandran plot statistics of all the five predicted models

Protein Structure	Ramachandran Statistics			
	No of Residues in %			
	No. of Residues in (%) Most favored region	Additional allowed Region	Generously allowed Region	Disallowed Region
TvLDH.B99990001.pdb	92	4.2	2.8	0.9
TvLDH.B99990002.pdb	92	5.2	0.9	1.9
TvLDH.B99990003.pdb	91	6.6	1.4	0.9
TvLDH.B99990004.pdb	91	6.1	1.4	1.4
TvLDH.B99990005.pdb	90.1	4.2	2.8	0.9

Furthermore, the quality of the structure derived from homology modeling was validated by calculating the ProSA Z-score, which gives the overall model quality based on the Cα positions. The Z-scores of the target and template were -6.36 and −7.24. Hence, the quality of the model obtained was validated using the ProSA score. Both the target and the template models show similar profiles, as seen in

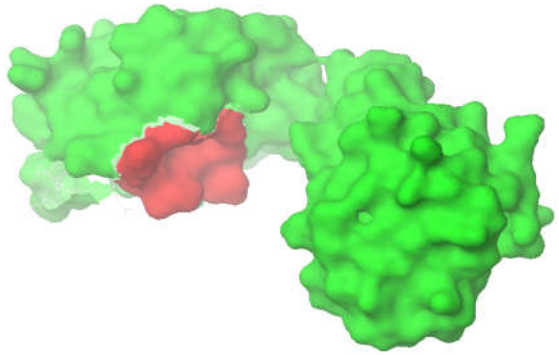

Fig 6. Red colour indicate the active site region of modelled protein

Figure 4A &4B. Geometric evaluations of the modeled 3D structure of SaeR were performed using PROCHECK by calculating the Ramachandran plot (Fig. 5) & (Table 2). This plot represents the distribution of the phi and psi angles for the amino acid residues. The percentage of phi and psi angles that occur in the allowed region was 92 % for residues in the core region, whereas this percentage was 0.9% for the residues located in the disallowed regions. So, the quality of Ramachandran plots is acceptable and the stereo chemical quality of the model was found to be satisfactory**.**

C. Active Site Analysis

Binding sites of the target protein were predicted using Q-site finder. (Active site prediction tool) The residues in receptor sites are summarized as residue name and corresponding residue numbers in the 3-D protein structure. The residues are identified on the protein receptor sites as LYS 142, VAL 143, VAL 144, HIS 149, GLU 150, VAL 151, PRO152, MET 153, ARG 154, ILE 155, PHE 158, with hydrophobic nature on the surface of the protein is shown in Figure 6.

Conclusion

The Regulation of virulence gene in *Staphylococcus aureus is very* complex and several regulator involved in determining the expression. In this paper, 3D structure of SaeR is predicted from know Crystal structure of 2OQR.using modeler 9v8 software. The energy minimization was used to refine the structure. Active site of the protein was indentified using Q site finder. This novel predicted structure of the SaeR protein place a good foundation to gain insight in designing antimicrobial drug by inhibiting the expression of SaeR protein.

Acknowledgement

This project is funded by BTIS (Biotechnology information system), DBT (Department of Biotechnology) and Ministry of Science & Technology, Government of India, India.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs *Nucleic Acids Research*, 25**:**3389-402.

Anchez, R.S., Sali, A., 1997. Advances in comparative protein-structure modelling *Current Opinion Structural Biology*, 7**:** 206-214.

Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000 *Nucleic Acids Research,* 28: 45-8.

Bajorath, J., Stenkamp, R., and Aru, A., 1994. Knowledge-based model building of proteins: concepts and examples *Protein Science*, 2: 1798-810.

Berman, H., Henrick, K., and Nakamura, H., 2003. Announcing the worldwide Protein Data *Bank Nature Structural Biology*, 10: 980.

Blundell, T.L., Sternberg, M.J.E., Sibanda, B.l. and Thornton, J.M .,1987. Knowledge-based prediction of protein structures and the design of novel molecules *Nature*, 326: 347-352.

Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. and Hill, R.C **1969**. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme *Journal of Molecular Biology*, 42: 65-86.

Castrignano, T., De Meo, P.D., Cozzetto, D., Talamo, I.G. and Tramontano, A., 2006. The PMDB Protein Model Database *Nucleic Acids Research,* 34: D306-309

Fiser, R.K., Sali, A., 2000. Modeling of loops in protein structures *ProteinScience*, 9: 1753-73.

Giraudo, A.T., Cheung, A.L. and Nagel, R., 1997. The sae locus ofStaphylococcus aureus controls exoprotein synthesis at the transcriptionallevel *Archives of Microbiology,* 168: 53-58.

Giraudo, A.T., Calzolari, A., Cataldi, A.A., Bogni, C. and Nagel, R., 1999. The sae locus of Staphylococcus aureus encodes a twocomponentregulatory system. *FEMS Microbiolog. Letters.* 177: 15-22.

Giraudo, A.T., Raspanti, C.G., Calzolari, A. and Nagel, R., 1994a *Canadian Journal of Microbiology* 40: 677-681

Hotchkiss, R.S., Coopersmith, C.M., McDunn, J.E. and Ferguson, T.A., 2009. The sepsis seesaw: tilting toward imunosuppression *Nature Medicine* 15: 496-497.

Johnson, M.S., Srinivasan, N., Sowdhamini, R. and Blundell, T.L., 1994. Knowledge-based protein modeling *CRC Critical Reviews in Biochemistr. MolecularBiol.,*29: 1-68.

Lappin, E., Ferguson, A.J., 2009. Gram-positive toxic shock syndromes *Lancet Infectious Disease,* 9: 281-290.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., 1993 PROCHECK: a program to check the stereo chemical quality of protein structures *Journal of Applied Crystallography,* 26: 283-291.

Laurie, A.T., Jackson, R.M., 2005. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites *Bioinformatics,* 21: *1908-16.*

Mart, M.A., Renom, A., Stuart, A., Fiser, R.,. Anchez, S., Melo, F. and Sali, A., 2000.Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Bimolecular Structure*, 29: 291-325.

Matin, G.S., Mannino DM, Eaton S, and Moss M 2003. The epidemiology of sepsis in the United States from 1979 though 2000 *New England journal. Medicine* 348 : 1546-1554.

Morris, A.L., MacArthur,M.W., Hutchinson, E.G. and Thornton,J.M., 1992. Stereochemical quality of protein

structure coordinates *Proteins: Structure, Function, and Bioinformatics* 12:345-364.

Novick, R.P., Jiang, D., 2003. The staphylococcal saeRS systemcoordinates environmental signals with agr quorum sensing *Microbiology,* 149: 2709-2717.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations *Journal of Molecular Biology***,** 7: 95-99.

Rittirsch, D., Flierl, M.A., and Ward, P.A., 2008. Harmful molecular mechanisms ofsepsis *Nature Reviews Immunology* 8: 776-787.

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints *Journal of Molecular Biology.* 234:779-815.

Thompson, J.D., Gibson, T.J. and Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX *Current. Protocols in Bioinformatics,* 2.3

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins *Nucleic Acids Research,* 35: W407-410.

****