WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 6.805
Volume 5, Issue 3, 104-113. Research Article ISSN 2277– 7105

IN SILICO MODELING AND ANALYSIS OF THE LARGEST HEMOLYSIN PROTEIN OF *STAPHYLOCOCCUS HAEMOLYTICUS*

***Bakr M. AL-Alway and Zahra M. AL-Khafaji**

Institute of Genetic Engineering and Biotechnology for Postgraduate Studies/University of Baghdad, Iraq.

Article Received on
27 Dec 2015,

Revised on 16 Jan 2016,
Accepted on 06 Feb 2016

***Correspondence for Author**
Dr. Bakr M. AL-Alway
Institute of Genetic Engineering and Biotechnology for Postgraduate Studies/University of Baghdad, Iraq.

ABSTRACT

Hemolysin protein is cytolytic toxin produced by *Staphylococcus haemolyticus* that effects the human erythrocytes and has nonspecific membrane toxicity to other mammalian cells. It is worth to be inactivated as a virulence factor for bacterial treatment. For the prediction of 3D-Structure of this protein several online modeling and prediction servers were used, the outputs of some servers had shown a poor quality and prediction but, Raptor X server output showed some promising result as it had 97% query coverage. The model was validated by QMEAN score, ProSA server and refined by 3D Refine. Finally the model was revalidated using ProAS server. Significant differences were observed after refinement in the quality of the model according to ProSA server.

KEYWORDS: *Staphylococcus haemolyticus*, *In Silico*, Homology Modeling, Raptor X, University of Baghdad.

INTRODUCTION

Coagulase-negative *staphylococci* (CNS) are a group of bacteria that are increasingly implicated as a cause of hospital-acquired opportunistic and health-care related infections worldwide. *Staphylococcus haemolyticus* strains are the second of the most frequently encountered CNS that have been recognized as an emerging and important human pathogen causing serious infections such as: endocarditis, urinary tract infections, septicemia, peritonitis, wound, bone and joint infections.[1]

Although *S. haemolyticus* strains are among the most common CNS species causing hospital-acquired opportunistic infections, little is known about their virulence-associated properties.[2]

Analyzing the whole genome sequence of human pathogenic strain and reported 57 open reading frames (ORFs) associated with virulence. They identified numerous genes encoded putative enzymes and toxins.

At least three orfs (SH0871, SH1134 and SH2193) showed homology to known staphylococcal hemolysins. A conserved-domain search adapted at NCBI found that SH0871 shared a motif with the *Bacillus cereus* hemolytic protein and it showed 49% amino acid identity to the streptococcal hemolysin III SH1134 shared a motif with the *Aeromonas hydrophila* hemolytic protein and it showed 60% identity to the *Bacillus halodurans* hemolysin. SH2193 also shared a motif for "hemolysin and related molecules," and it was 39% identical to a *Helicobacter pylori* hemolysin.[3]

The aim of this study was to Predict the 3D-Structure of hemolytic protein coding by SH2193 ORF which could lead to the structure function analysis and also help for designing a drug to inhibit one of the most important *S. haemolyticus* virulence factors.

MATERALS AND METHODS

In this study SH2193 was used as a target.

NCBI (www.ncbi.nlm.nih.gov) was used for retrieving the amino acid of the target sequence.

Protein Data Bank (www.rcsb.org/pdb) was used to find PDB file of the protein.

BLASTp.[4] was used to for sequence alignments.

PSIPRED server.[5] was used for prediction of protein secondary structure.

SWISS MODEL[6], Raptor X[7], Phyre 2[8], I-TASSER[9] and LOMETS[10] were used for modeling the 3D-Structure of the protein.

UCSF Chimera software[11] was used for structure visualization.

ProSA server[12] and QMEAN[13] server were used for validation.

3D Refiner server[14] was used for refining the model structure.

RESULTS AND DISCUSSION

FASTA format of amino acids sequence of protein with accession number WP_011276453.1 was retrieved from NCBI as a target sequence which consist of 458 amino acids identify conserved domains feature at NCBI revealed that the protein sequence include three domains "Fig. 1". Searching the RCSB Protein Data Bank confirmed that the 3D structure of this protein was unavailable.

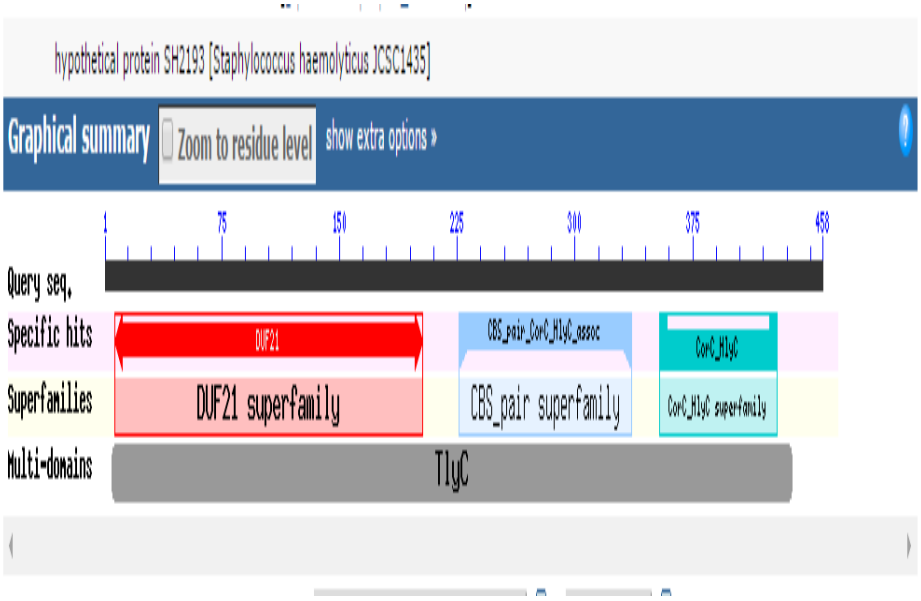

Fig. 1: Conserved domains result from NCBI.

2 D-Protein Structure

In the PSIPRED server and Phyre 2 results shown approximately the same result; Alpha helix (53%), Beta strand (16%) and Transmembrane helix (23%) of the sequence under investigation "Fig. 2".

The secondary structures of proteins are the regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the alpha-helix and beta-sheet that consider as an evidence helps to pick the reliable 3D-Structure of protein.

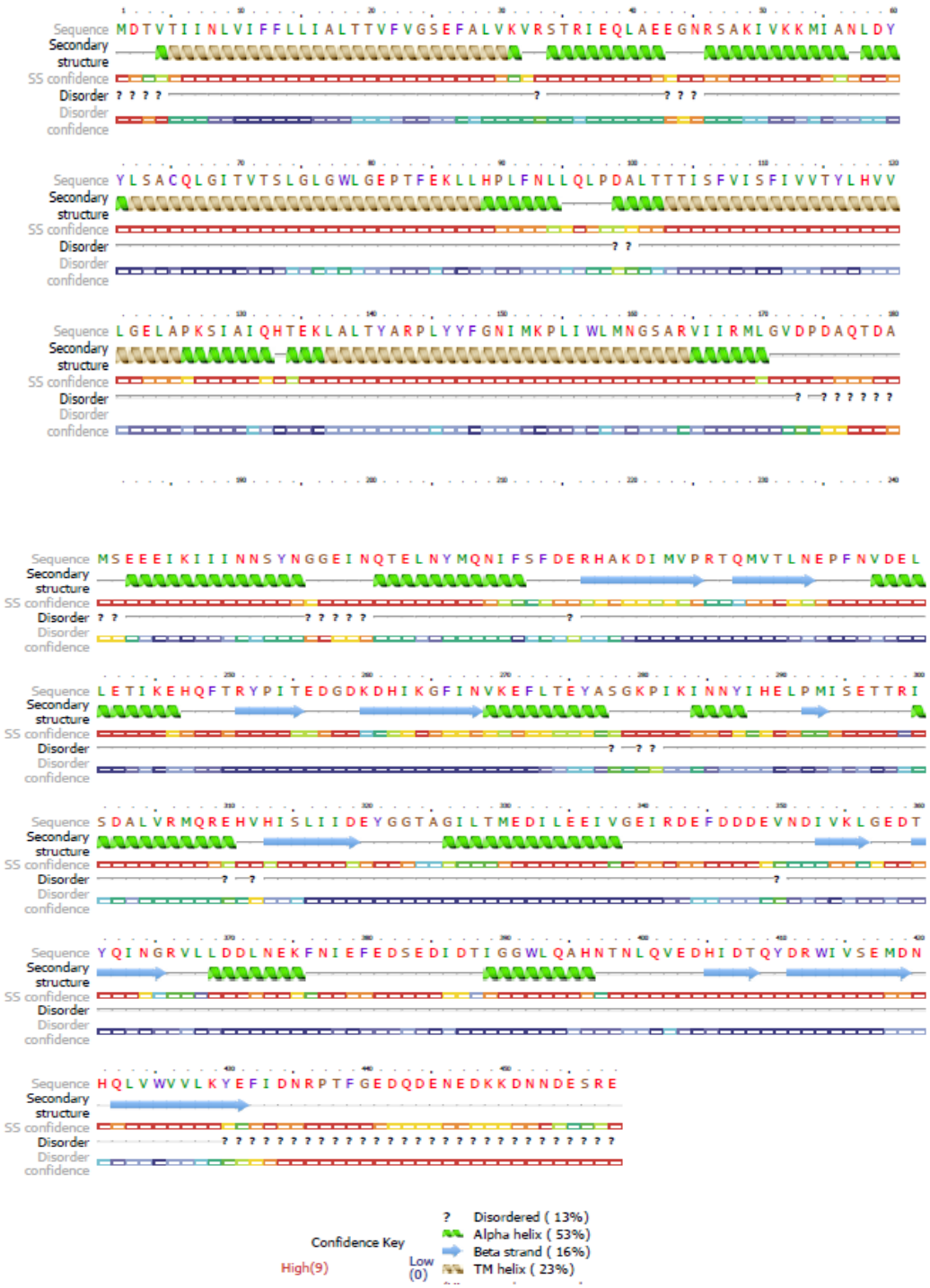

Fig. 2: Secondary structure of Phyre 2.

3D-Protein Structure

The majority of the current approved small molecule drugs target proteins like enzymes and receptors in the human body.[15] Therefore, structural understanding of these targets are highly beneficial for *In Silico* drug design. The primary and ideal resource of protein structural information is X-ray crystallographic experiments. However, the experimental determination of proteins are often time consuming and expensive. In circumstances where

the experimental structure of target protein is unavailable, computational method can be helpful. Computational techniques have been developed to model proteins based on evidences from biophysical knowledge of proteins as well as sequence specific information of the target protein under study.[16] The most popular method of protein modeling is by homology modeling (comparative modeling) in which the structural elements are borrowed from already experimentally solved homologous protein structures.[17] The 3D protein structure was predicted using SWISS MODEL, Raptor X, Phyre 2, I-TASSER and LOMETS. Twenty models were obtained visualized under UCSF Chimera software and then submitted to 3D-Structure validation and refinement. The 3D-Structure of Raptor X model is given below "Fig. 3".

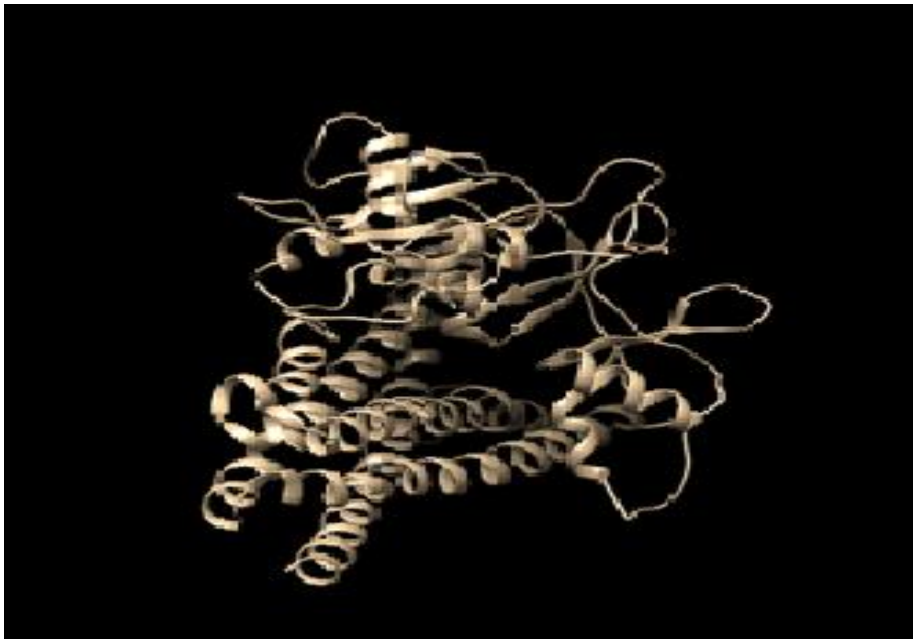

Fig. 3: 3D-Structure of Raptor X model.

Structure Validation

The availability of a structural model of a protein is one of the keys for understanding biological processes at a molecular level. The recent advances in experimental technology have led to the emergence of large-scale structure determination pipelines aimed at the rapid characterization of protein structures. The resulting amount of experimental structural information is enormous. The application of computational methods for the prediction of unknown structures adds another plethora of structural models. The assessment of the accuracy and reliability of experimental and theoretical models of protein structures is a necessary task that needs to be addressed regularly, it is essential for maintaining integrity,

consistency and reliability of public structure repositories.[18] The geometrical and structural consistencies of both modeled and template proteins were evaluated by different approaches. The structural validation was carried out by QMEAN server and ProSA server. The output models from each server were submitted to the QMEAN server for scoring, the best model was of Raptor X because it covered 97% of the protein sequence this percent was obtained after initiation of pairwise alignment between the protein sequence and the resulted Raptor X model sequence using BLASTp "Fig. 4", In addition it has the best Z-score (-3.44) and QMEAN score (0.476) shown in "Fig. 5".

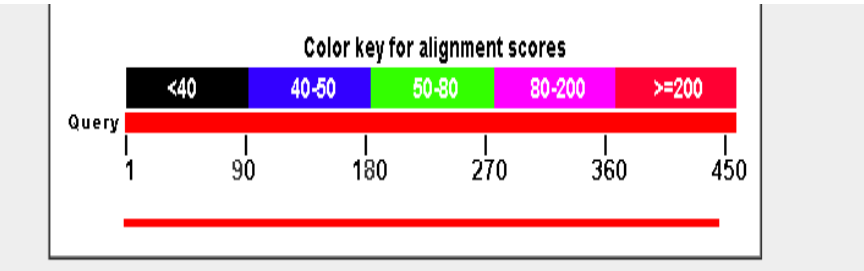

Fig. 4: Pairwise alignment output of hemolysin sequence (WP_011276453.1) and Raptor X model sequence.

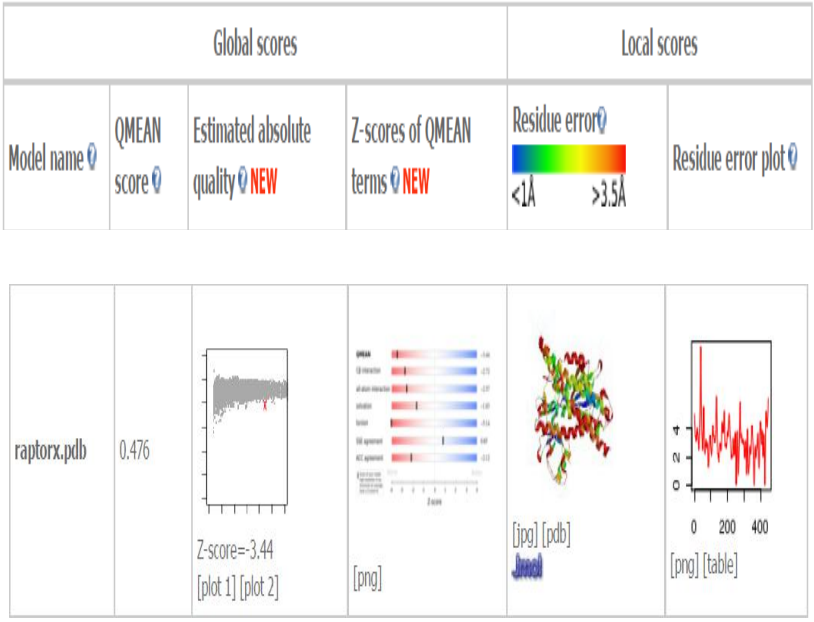

Fig. 5: QMEAN Server result of Raptor X model.

The SWISS MODEL server gave two models with Z-score and QMEAN score higher than the Raptor X but they had 50% coverage to the protein sequence as shown in "Fig. 6".

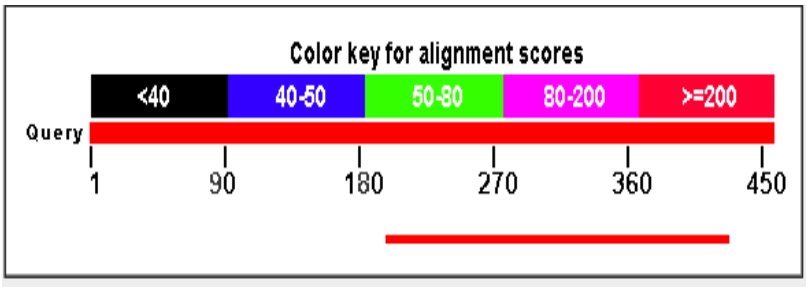

Fig. 6: Pairwise alignment output of hemolysin sequence (WP_011276453.1) and model sequence modeled by SWISS MODEL server.

The other models had lower Z-scores and QMEAN scores worst than the Raptor X model, the selected model was validated furthermore by ProSA server, the overall quality score calculated by ProSA for this model was displayed in a plot that shows the scores of all experimentally determined proteins currently available in the Protein Data Bank (PDB), The model had Z-score (-6.2) with an excepted local energy folding "Fig. 7".

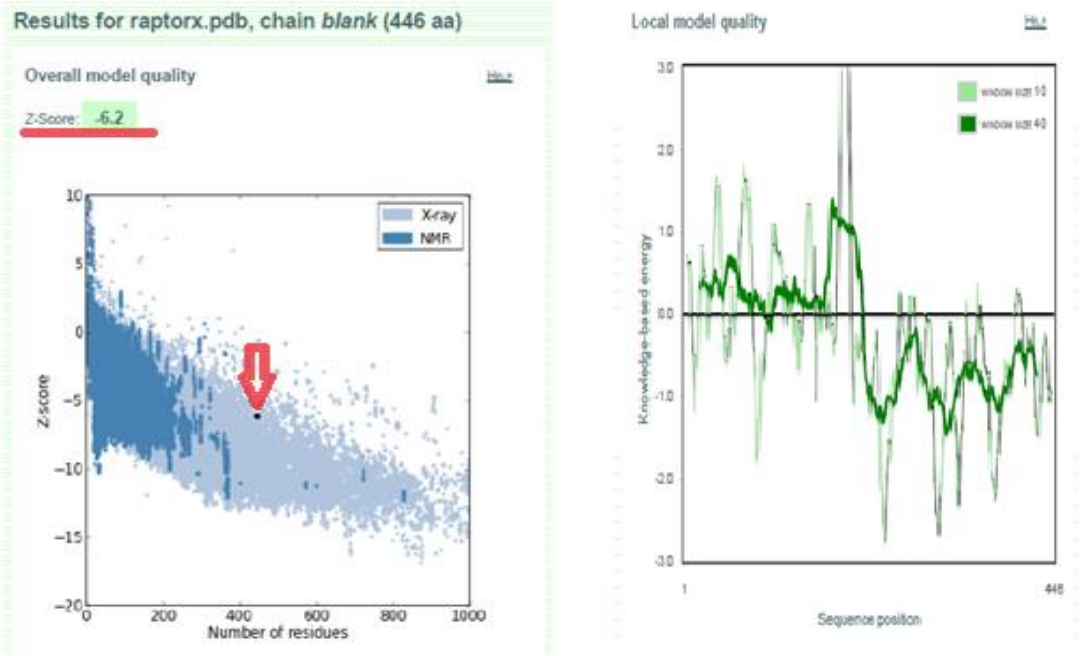

Fig. 7: Validation of model by ProSA server.

Structure Refinement

The introduction of multiple templates has certainly enhanced the accuracy of structure prediction by bringing the model closer to the native structure than using a single template. Despite having largely correct backbone conformations, these models sometimes still have poor structural qualities, including irregular hydrogen bonding network, steric clashes, unphysical bond length, unrealistic bond angles, torsion angles and side-chain χ angles. Thus,

direct refinement of the predicted models from their coordinate information alone with the goal of detection and correcting the errors is an essential part of computational protein structure prediction process.[10]

The selected model refined with 3D Refiner server that follow protocol based on two steps of refinement process, first step is based on optimization of hydrogen bonding (HB) network and the second step is applies atomic-level energy minimization on the optimized model using a composite physics and knowledge-based force fields.

The refinement step decrease the value of the Z-Score which reflects the reality enhancement of the model and reduces the energy on the energy chart which Reflects an increase in the stability of of the structure that is obvious in the result of revalidation using ProSA server as shown in " Fig. 8" in compared with "Fig. 7".

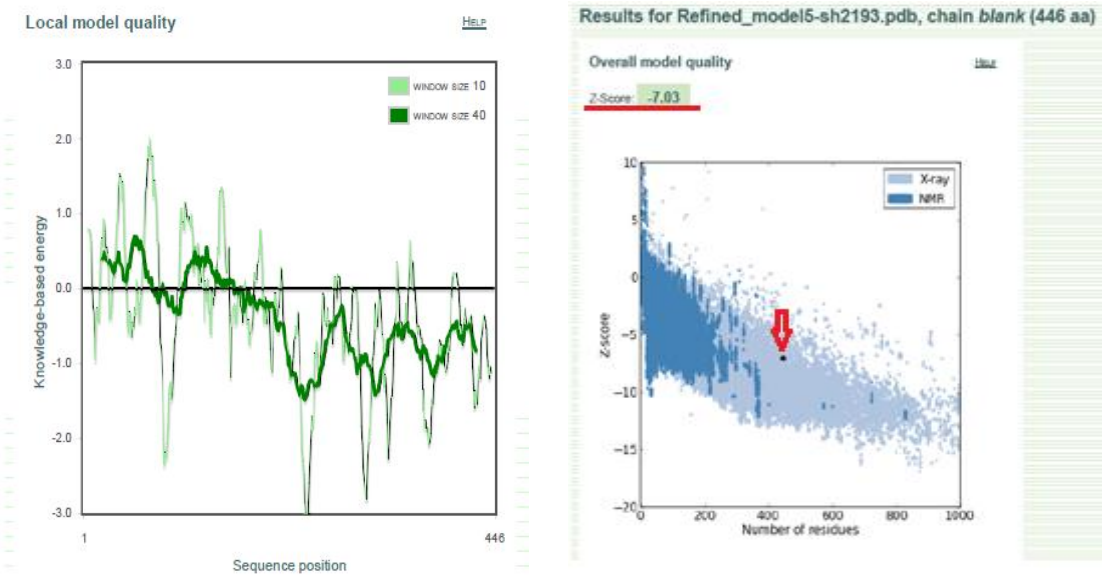

Fig. 8: Revalidation result by ProSA server of the Raptor X model after refinement.

CONCLUSION

The purpose of this study was to minimize the gap between *In Silico* and wet lab prediction of 3D-Structure of a protein by molecular modeling. The 3D-Structure model of hemolysin protein was stable proved reliable using the ProSA server and the QMEAN server. The overall results provided the evidences that the predicted 3D-Structure of hemolysin protein by Raptor X is acceptable and of good quality. The predicted structure for hemolysin protein will give an idea of its active site and the active site residues which can be farther analyzed

for designing inhibitors to inactive one of the most important virulence factor of S. *haemolyticus* (result have not being published yet).

REFERENCES

1. Piette A, Verschraegen G (Role of coagulase-negative *staphylococci* in human disease). *Vet Microbiol*, 2009; 134(1-2): 45–4.
2. Fumihiko TA, Shinya WA, Tadashi BA, Harumi YU, Teruyo IT, Yuh MO, Makoto KU, Longzhu CU, Mikio TA, Akiho AN, Shin-ichi BA, Shigehiro FU, Jean C. Lee and Keiichi.(Whole genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species). *J Bacteriol*, 2005; 187(21): 7292-308.
3. Krzyminska SY, Ewa SZ, Kinga DU, Adam KA. (Virulence and the presence of aminoglycoside resistance genes of Staphylococcus haemolyticus strains isolated from clinical specimens). *Anton Leeuw Int J G*, 2015; 107: 857–68.
4. Altschul, SF, Gish W, Miller W, Myers EW, Lipman DJ. (Basic local alignment search tool). *J. Mol. Biol.*, 1990; 215: 403-10.
5. McGuffin LJ, Bryson KJ. (The PSIPRED protein structure prediction server). *Bioinformatics*., 2000; 16(4): 404-5.
6. Schwede T, Kopp J, Guex N, Peitsch MC. (SWISS-MODEL: an automated protein homology-modeling server). *Nucleic Acids Res*., 2003; 31(13): 3381-5.
7. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. (Template-based protein structure modeling using the Raptor X web server). *Nat Protoc*., 2012 19; 7(8): 1511-22.
8. Kelley LA, Sternberg MJ. (Protein structure prediction on the Web: a case study using the Phyre server). *Nat Protoc*., 2009; 4(3): 363-71.
9. Zhang Y, (I-TASSER server for protein 3D structure prediction). *BMC Bioinformatics*., 2008; 23(9): 40.
10. Wu S, Zhang Y. (LOMETS: a local meta-threading-server for protein structure prediction). *Nucleic Acids Res*., 2007; 35(10): 3375-82.
11. Pettersen EF1, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. (UCSF Chimera--a visualization system for exploratory research and analysis). *J Comput Chem*., 2004; 25(13): 1605-12.
12. Wiederstein M, Sippl MJ. (ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins). *Nucleic Acids Res.*, 2007; 35(Web Server issue): 407-10.

13. Benkert, Michael K, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res*., 2009; 1(37): 510–14.
14. Bhattacharya D, Cheng J. (3Drefine: consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization). *Proteins*., 2013; 81(1): 119-31.
15. Imming PE, Sinning CH, Meyer AC. (Drugs, their targets and the nature and number of drug targets). *Nat Rev Drug Discov*, 2006; 5: 821-34.
16. Lee D, Redfern O, a Orengo C. (Predicting protein function from sequence and structure). *Nat Rev Mol Cell Biol.*, 2007; 8(12): 995-5.
17. Sali A, Blundell T. (Comparative protein modeling by satisfaction of spatial restraints*). J Mol Biol*., 1993; 234(3): 779-15.
18. Berman HM, Burley SK, Chiu W, Sali A, Adzhubei A, Bourne PE, Bryant SH, Dunbrack RL, f Fidelis K, Frank J, Godzik J, Henrick J, Joachimiak J, Heymann B, Jones D, Markley JL, Moult J, Montelione GT, Orengo C, Rossmann MG, Rost B, Saibil H, Schwede T, Standley DM, Westbrook GD. (Outcome of a workshop on archiving structural models of biological macromolecules). *Structure*., 2006; 14(8): 1211–17.