RESEARCH ARTICLE

Chinese Journal of Medical Research
Chinese J Med Res 2020; 3(2): 32-36

ISSN(e): 2618-091X
ISSN(p): 2663-8053

# *In Silico* Characterization of Surface Glycoprotein [QHD43416] of Severe Acute Respiratory Syndrome-Coronavirus 2

**Rajneesh Prajapat[1*], Suman Jain[1], Manish K Vaishnav[1], Sonal Sogani[1]**

*1 Department of Biochemistry, Pacific Institute of Medical Sciences, Sai Tirupati University, Umarda, Udaipur, Rajasthan, India*

*Corresponding Author: Dr. Rajneesh Prajapat, Department of Biochemistry, Pacific Institute of Medical Sciences, Sai Tirupati University, Umarda, Udaipur, Rajasthan, India.

Email: prajapat.rajneesh@gmail.com; Tel: +91-7976055027

**Received:** March 23, 2020; **Accepted:** May 21, 2020

## Abstract

The novel coronavirus (SARS-CoV-2) reported from Wuhan, China, that spread rapidly and cause severe acute respiratory syndrome. The disease associated with infection of SARS-CoV-2 that is referred as COVID-19 (Coronavirus Disease 2019). In the present study, the surface glycoprotein [QHD43416] of SARS-CoV-2 was characterized for structure analysis and validation to provide information about its three-dimensional structure by using *in silico* tools and techniques.

The surface glycoprotein [QHD43416] sequence of SARS-CoV-2 was retrieved from NCBI and its PDB file was designed by using phyre$^2$ server. The RAMPAGE and UCLA-DOE (Verify 3D) was used for analysis and validation of structure model of protein. The model quality estimation based on the ProSA.

Alignment of surface glycoprotein [QHD43416], revealed homology (72% identity) with spike protein of bat coronavirus [BM48-31/BGR/2008]. The model corresponding to probability conformation with 90.5% residue of core section, 9.1 % of allowed section and 0.4 % residue of outer section in φ-ψ plot, that specifies accuracy of prediction model. The Verify 3D results shows that 59.53% residues have average 3D-1D score >= 0.2 this determines compatibility of 3D model with its amino acid sequence (1D). ProSA Z-score -11.19 represents the good quality of the model.

The structure and function of coronavirus surface glycoprotein could be predicted by *in silico* modeling studies. The protein model will be further used for designing of vaccine / drug development against coronavirus infection.

**Keywords:** COVID-19, Coronavirus, Surface glycoprotein.

## INTRODUCTION

The first SARS case was reported from China in 2002 [1]. The *Coronaviridae* family viruses identified as the causative agent of disease and nominated as the SARS-associated coronavirus (SARS-CoV) [2,3]. The novel coronavirus (SARS-CoV-2) reported from Wuhan, China, that spread rapidly and cause severe acute respiratory syndrome. The disease associated with infection of SARS-CoV-2, is referred as COVID-19 (Coronavirus Disease 2019). The COVID-19 reported from 213 countries around the world and territories. Till 29 April, according to WHO, confirmed cases were 32,21,617 and more than 2,28,260 deaths reported worldwide. In India the active cases were 23,651 and 1074 death were reported according to COVID-19 Dashboard of India till dated 30 April 2020. The recovery reported in 9,65,197 patients globally and estimated mortality risk is ~2%.

The genome sequencing result of SARS-CoV indicates that it has polyadenylated RNA of 29.7 kb [4]. During clinical trial, after treatment with ritonavir / lopinavir, no or little coronavirus titers were detected in infected patient [5]. Chloroquine phosphate illustrations apparent efficacy against COVID-19 associated pneumonia in multicenter clinical trials [6].

The analysis of sequence, algorithm design and biological data management is possible by *in-silico* methods [7,8,9,10]. In present study, the surface glycoprotein [QHD43416] of SARS-CoV-2 was characterized for structure analysis and validation to provide information about its three-dimensional structure by using *in silico* tools and techniques.

## MATERIALS AND METHODS

**Homology modeling:** The surface glycoprotein [QHD43416] sequence of SARS-CoV-2 was retrieved from NCBI. The phyre$^2$ [Protein Homology/AnalogY Recognition Engine] server used to generate PDB of surface glycoprotein [QHD43416]. In phyre$^2$ analysis, 82% of sequence was exhibited with 100 % assurance by single uppermost recording template. Simulated annealing protocol used for energy minimization of model. A perfect model of protein domain could be created by sequence alignment between sequence of surface glycoprotein [QHD43416] and database sequence.

**Model reputation:** The UCLA-DOE server provides quality analysis of protein crystal structure and it requires structure in PDB format [11]. PROCHECK server used for validation of structure model [12,13] and its results suggesting reliability of model [14]. The overall G-factor, residue positions in φ-ψ plot regions and WHATIF analysis was used

for the selection of suitable model [15,16] and QMEAN (https://swissmodel.expasy.org/qmean/version 3.1.0) [17,18] and ProSA [19]. The protein stability analysed by using ProSA and QMEAN Z-score.

## RESULTS

**Protein Model Building:** The alignment between target and template was performed by using homology modeling [20]. The sequence alignment of surface glycoprotein [QHD43416], revealed sequence homology (72% identity) with spike protein of bat coronavirus [BM48-31/BGR/2008]. Alignment results correlate origin of SARS-CoV-2 surface glycoprotein [QHD43416] with spike protein of bat coronavirus. The ribbon model of surface glycoprotein [QHD43416] was generated by using RaptorX structure prediction server (Fig. 1).

**Model reputation:** The overall G-factor -0.62 indicates good stereo chemical property of model and represents that model geometry resembles to conformation with 90.5% residues in core section of φ-ψ plot [21]. Resulted percentage of residues in allowed and outer section was 9.1% and 0.4% respectively (Fig. 2a, b). The above results indicate reliability of protein model [21].

The compatibility and score profile of (3D) amino acid atomic model illustrated by verify 3D graph [22]. The high score of 0.75 indicates good quality of model (Fig. 3). Profile score beyond zero of verify 3D indicates acceptable model output [23].

The 59.53% of graph residues displayed an average 3D-1D rating > = 0.2 and less than 80% scored > = 0.2.

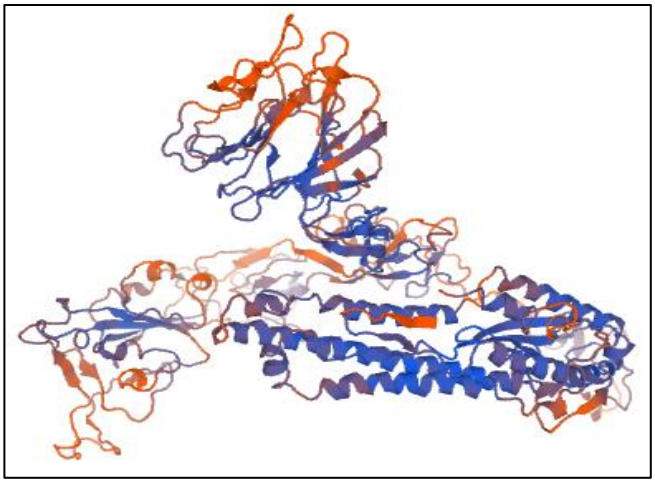

**Fig. 1:** Surface glycoprotein [QHD43416] ribbon model generated using RaptorX structure prediction server.

**Validation of Model:** ProSA was used to figure out potential errors in 3D model of surface glycoprotein [QHD43416]. The archived ProSA Z-score score -11.19 indicates two aspects: overall model quality and energy deviation of surface glycoprotein [QHD43416] (Fig. 4). The values of Z-score thus predicted indicates less erroneous structures [24]. Reliability of projected model based on scoring function of QMEAN that stated as 'Z-score' (Fig. 5) [25].

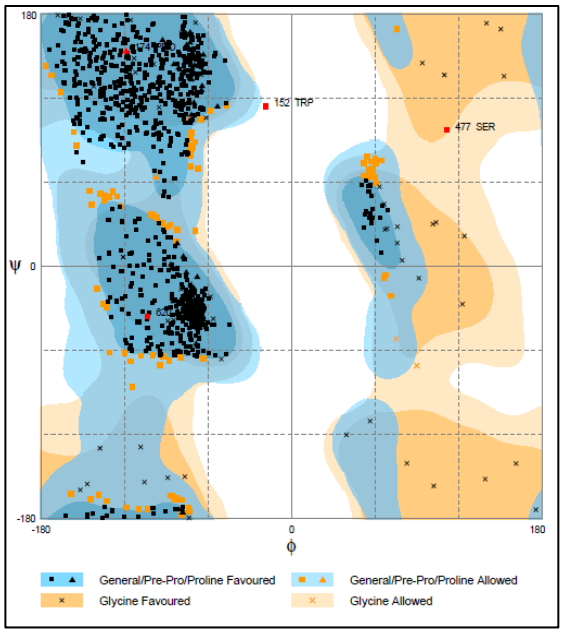


**Fig 2a.** φ-ψ plot of surface glycoprotein [QHD43416]. Total number of residues were

953 (90.5%) in favoured [A, B, L], 96 (9.1%) in allowed [a,b,l,p] and, 4 (0.4%) in outlier regions.

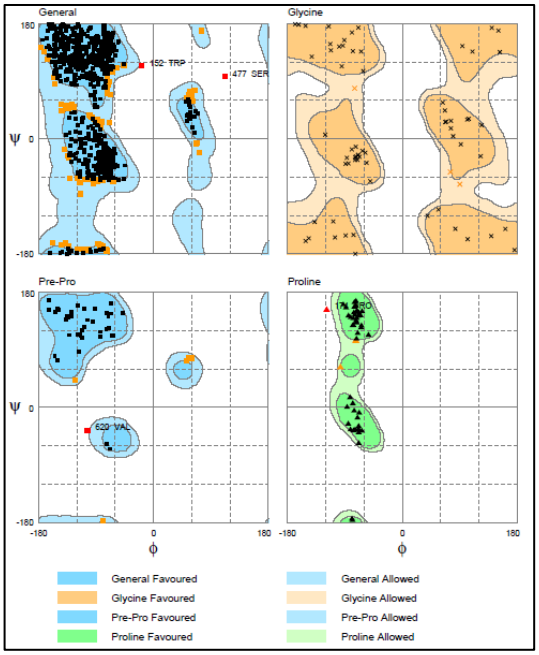


**Fig. 2b:** The non-glycine and non-proline residue regions of φ-ψ plot

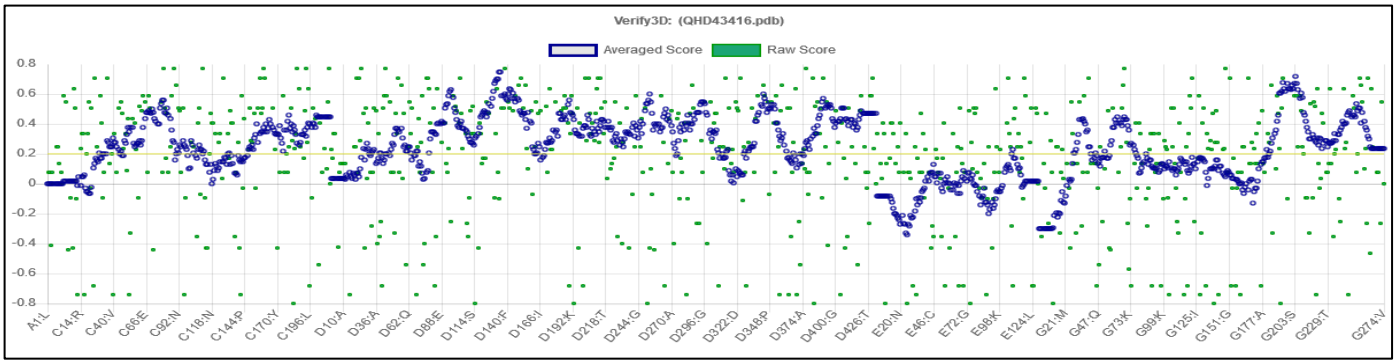


**Fig. 3:** The Verified 3D graph of surface glycoprotein [QHD43416]. 59.53% of graph residues displayed an average 3D-1D rating > = 0.2 and less than 80% scored > = 0.2.

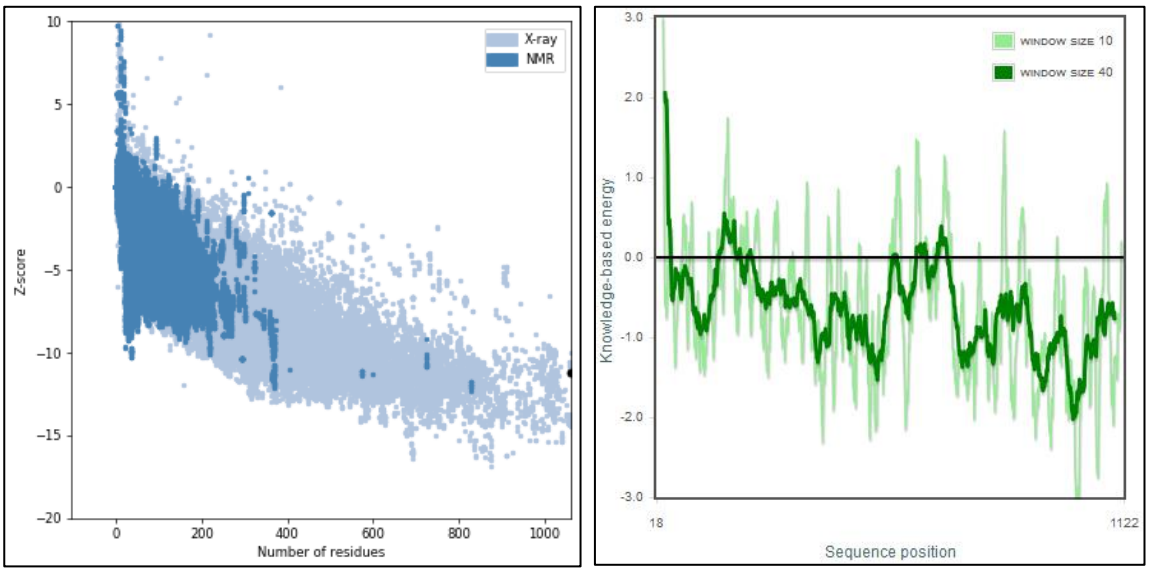


**Fig. 4:** ProSA service examination of surface glycoprotein [QHD43416] overall model quality (a) and local model quality (b)

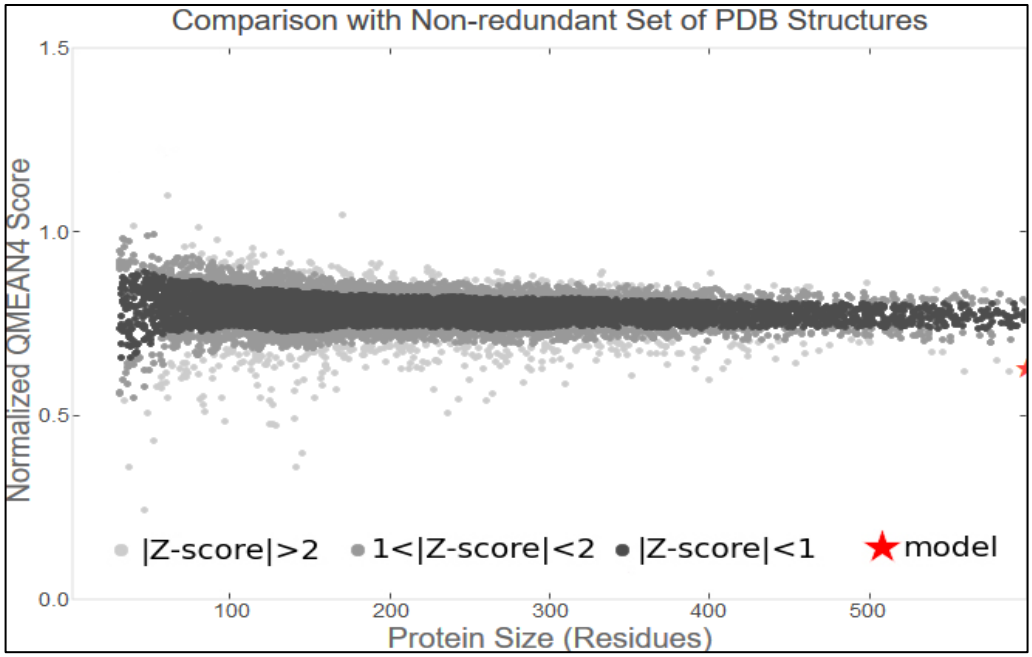


**Fig 5:** QMEAN scores for biological unit reference set. Plot showing Z-score.

The Local Distance Difference Test (lDDT) was performed [http://swissmodel.expasy.org/lddt] for evaluating local accuracy and stereochemical plausibility of models. The QMEAN Z-score -3.67, which was very close to 0 and its illustrations acceptable value [26].

Assessed validity of model predictable among 0 and 1, that could be concluded from the density plot locus set for QMEAN score. Figure 5 illustrations QMEAN scores for biological unit reference set that used as a tool for oligomeric protein assessment. The QMEAN value (-3.67) comparison with the non-redundant protein collection revealed different set of Z-values for different parameters. The diversion of total energy of surface glycoprotein [QHD43416] was measured by using Z-score [27].

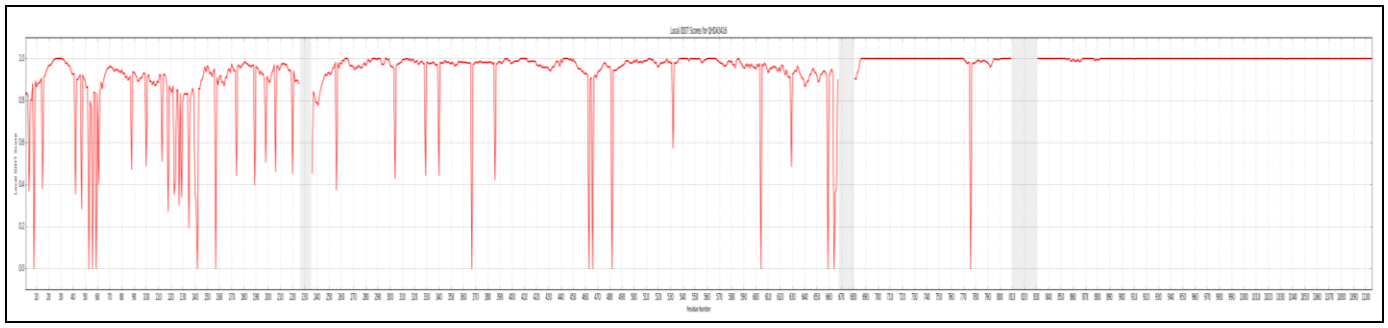

**Fig. 6:** Local lDDT score of surface glycoprotein [QHD43416]

The Z-score also tests variance of total structural energy with respect to the energy dispersal resulting from random conformation. Local Distance Difference Test (lDDT) score 0.9387 indicates a highly reliable structure (Fig. 6). The lDDT evaluates validation of stereochemical plausibility and local distance variances of atoms in model [28].

## DISCUSSION

The surface glycoprotein [QHD43416], revealed homology (72% identity) with spike protein of bat coronavirus [BM48-31/BGR/2008]. Model configuration resembles 90.5% residue conformation of probability in the core section of φ-ψ plot. It shoes the reliability of projected model. ProSA Z-score -11.19

represents the good quality of model. QMEAN Z-score shows absolute quality of theoretical model and measure experimental structures with significant errors.

The Z-score also tests total structural energy variance in relation to a spectrum of energy derived from arbitrary conformations. The scores showed a very stable structure found in proteins of similar size.

The energy plot suggested model utility in function of the location of amino acid sequence. The lDDT score used to evaluat protein structure models with respect to a reference structure and to evaluate the consistency of the model in context of domain movements.

## CONCLUSION

The functional characteristics of surface glycoprotein [QHD43416] could be predicted by the generated mode. The structure, function and mechanism of proteins action can be studied through *in silico* modeling techniques. Methods ProSA, QMEAN, and PROCHECK build model reliability. The research findings useful for vaccine development as well as understand the infection mechanism of SARS-CoV-2.

**Conflict of Interest**

Here by, we have no conflict of Interest.

**Financial Support**

None declared

**ACKNOWLEDGEMENTS**

The authors are thankful to Dr. B. L. Kumar (Principal, Pacific Institute of Medical Sciences, Udaipur, Rajasthan (India) for their precious support and to biochemistry research group members for technical support.

## REFERENCES

1. Zhong NS. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February. Lancet 2003;362:1353-1358.
2. Peiris JS. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 2003;361:1319-1325.
3. Drosten C. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. 2003;348:1967-1976.
4. Chow KYC, Hon CC, Hui RK, Wong RTY, Yip CW, Zeng F, *et al*. Molecular Advances in Severe Acute Respiratory Syndrome-associated Coronavirus (SARS-CoV). Genomics Proteomics Bioinformatics 2003;1(4): 247–262.
5. Lim J, Jeon S, Shin HY, Kim MJ, Seong YM, Lee WJ *et al*. Case of the Index Patient Who Caused Tertiary Transmission of COVID-19 Infection in Korea: The Application of Lopinavir/Ritonavir for the Treatment of COVID-19 Infected Pneumonia Monitored by Quantitative RT-PCR.J Korean Med Sci. 2020; 35(6):e79.
6. Gao J, Tian Z, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci Trends. 2020.doi: 10.5582/bst.2020.01047.
7. Naulaerts S, Meysman P, Bittremieux W, Vu TN, WV Berghe, Goethals B, Laukens K. A primer to frequent itemset mining for bioinformatics. Briefings Bioinform. 2015;16:216-231.
8. Rasouli H, Fazeli-Nasab B. Structural validation and homology modeling of lea 2 protein in bread wheat. Am.-Eurasian J. Agric. Environ. Sci., 2014;14:1044-1048.
9. Pevzner P, Shamir R. Bioinformatics for Biologists. 1st Edn., Cambridge University Press, Cambridge, UK., 2011; 394. ISBN-13: 978-1107648876
10. Prajapat R, Marwal A, Shaikh Z, Gaur RK. Geminivirus Database (GVDB): First database of family geminiviridae and its genera Begomovirus. Pak. J. Biol. Sci. 2012;15:702-706.
11. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature, 1992;356:83-85.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J. Applied Cryst. 1993;26:283-291.
13. Vriend G. WHAT IF. A molecular modeling and drug design program. J. Mol. Graphics; 1990;8:52-56.
14. Sehgal SA, Tahir RA, Shafique S, Hassan M, Rashid S. Molecular modeling and docking analysis of CYP1A1 associated with head and neck cancer to explore its binding regions. J. Theoret. Comput. Sci. 2014, 1.
15. Agrawal P, Thakur Z. Kulharia M. Homology modeling and structural validation of tissue factor pathway inhibitor. Bioinformation; 2013; 9:808-812.
16. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Struct. Funct. Bioinform. 2008;71:261-277.
17. Benkert P, Kunzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res. 2009; 37:W510-W514.
18. Novotny WF, Girard TJ, Miletich JP, Broze GJ. Platelets secrete a coagulation inhibitor functionally and antigenically similar to the lipoprotein associated coagulation inhibitor. Blood. 1988;72:2020-2025.
19. 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:W407-W410.
20. Prajapat R, Marwal A, Gaur RK. Recognition of errors in the refinement and validation of three-dimensional structures of AC1 proteins of begomovirus strains by using ProSA-web. J. Viruses; 2014, 6. doi.10.1155/2014/752656.
21. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991; 253:164-170.
22. Mustufa MMA, Chandra S, Wajid S. Homology modeling and molecular docking analysis of human RAC-alpha serine/threonine protein kinase. Int. J. Pharma Bio Sci. 2014; 5:1033-1042.
23. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 2011; 27:343-350
24. Rekik I, Chaabene Z, Grubb CD, Drira N, Cheour F, Elleuch A. *In silico* characterization and molecular modeling of double-strand break repair protein MRE11 from Phoenix dactylifera v deglet nour. Theor. Biol. Med. Model 2015;12:23. doi: 10.1186/s12976-015-0013-2.
25. Wiederstein M, Sippl MJ. Protein sequence randomization: Efficient estimation of protein stability using knowledge-based potentials. J. Mol. Biol. 2005; 345:1199-1212.
26. Agrawal P, Thakur Z, Kulharia M. Homology modeling and structural validation of tissue factor pathway inhibitor. Bioinformation 2013; 9(16):808-812.
27. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics.2011;27(3):343-350
28. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics 2003; 29(21):2722-2728.