



Trends in Bioinformatics

ISSN 1994-7941

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>



Research Article

Homology Modeling and Structural Validation of Type 2 Diabetes Associated Transcription Factor 7-like 2 (TCF7L2)

Rajneesh Prajapat, Ijen Bhattacharya and Anoop Kumar

Department of Medical Biochemistry, Faculty of Medical Sciences, Rama Medical College, Rama University, Kanpur, Uttar Pradesh, India

Abstract

Background: New study findings indicate that variation in the transcription factor 7-like 2 (TCF7L2) gene, linked to the pathogenesis of type 2 diabetes. In the present study, the protein structure model of TCF7L2 was generated, to understand the structure, function and mechanism of the action of proteins. The present study was designed to enlist some of the physiochemical and functional properties of TCF7L2 protein and provides information about its three-dimensional structure. **Materials and Methods:** The PDB file of TCF7L2 [CAG38811] was generated by Phyre 2 servers. Model construction and regularization (including geometry optimization) of model were done by optimization protocol in YASARA. The energy of the model was minimized using the standard protocols of combined application of simulated annealing, conjugate gradient and steepest descent. The UCLA-DOE server was used to visual analysis of the quality of a putative crystal structure for protein. The validation for structure models was performed by using PROCHECK. The model was further analyzed by WHATIF, QMEAN and ProSA. **Results:** The model showed good stereo-chemical property in terms of overall G-factor value of -0.64 indicating that geometry of model corresponds to the probability conformation with 67.9% residue in the core region of Ramachandran plot showing high accuracy of model prediction. The Z-score of -6.07 predicted by ProSA represents the good quality of the model. The Z-score also measures the divergence of total energy of the structure with respect to an energy distribution derived from random conformations. The scores indicate a highly reliable structure and are well within the range of scores typically found for proteins of similar size. The energy plot shows the local model quality by plotting knowledge-based energies as a function of amino acid sequence position. **Conclusions:** The generated model could be supportive to understand the functional characteristics of transcription factor 7-like 2 (TCF7L2). The variants in TCF7L2 associated with the risk for type 2 diabetes.

Key words: TCF7L2, diabetes, homology modeling, *in silico*

Received: January 18, 2016

Accepted: May 07, 2016

Published: June 15, 2016

Citation: Rajneesh Prajapat, Ijen Bhattacharya and Anoop Kumar, 2016. Homology modeling and structural validation of type 2 diabetes associated transcription factor 7-like 2 (TCF7L2). Trends Bioinform., 9: 23-29.

Corresponding Author: Rajneesh Prajapat, Department of Medical Biochemistry, Faculty of Medical Sciences, Rama Medical College, Rama University, Kanpur, Uttar Pradesh, India Tel: +91-9460558447

Copyright: © 2016 Rajneesh Prajapat *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Impaired insulin secretion is concomitant with type 2 diabetes, as discussed by Anna *et al.*¹. The transcription factor 7-like 2 (TCL7L2) is a transcription factor whose gene product is concerned for blood glucose homeostasis: A High Mobility Group (HMG), as discussed by Yi *et al.*². Factor TCF7L2 is responsible for type 2 diabetes, whose gene is positioned on chromosome 10q25 and legalizes cell differentiation and proliferation, as discussed elsewhere^{3,4}. The TCF7L2 has recently been implicated in the pathogenesis of type 2 diabetes (T2D) through regulation of pancreatic β -cell insulin secretion, as discussed elsewhere^{5,6}. There are variants of TCL7L2, which intensify the menace for type 2 diabetes, further the variants likely impact both insulin sensitivity and insulin secretion, as discussed by Shu *et al.*⁷. Bioinformatics helps in management of complex biological data, sequence analysis and algorithmic designing, as discussed elsewhere^{8,9}. However, by using the *in silico* analysis we can analyze the protein sequences^{10,11}. Therefore, the present study enlists some of the physiochemical and functional properties of TCF7L2 protein and provides information about its three-dimensional structure.

MATERIALS AND METHODS

Operating system: The present study was conducted by using Intel (R) Core (TM) i3-370 M CPU @ 2.40 GHz and 32 bit operating system (HP ProBook).

Sequence retrieval, alignment and homology modeling: The FASTA sequence of transcription factor 7-like 2 (TCF7L2 [CAG38811]) protein was retrieve from NCBI. The PDB file of TCF7L2 [CAG38811] protein was generated by Phyre 2 servers by using its FASTA sequence. In order to build a model of protein domain, multiple sequence alignment was performed between full length TCF7L2 protein sequence and another protein sequences in this database. To build the model of the TCF7L2 protein with more homology, structure of TCF7L2 protein model in 3D-JIGSAW server was selected as template. Model construction and regularization (including geometry optimization) of model were done by optimization protocol in YASARA. The energy of the model was minimized using the standard protocols of combined application of simulated annealing, conjugate gradient and steepest descent.

Model reputation: The UCLA-DOE server provides a visual analysis of the quality of a putative crystal structure for protein.

Verify 3D expects this crystal structure to be submitted in PDB format, as discussed by Luthy *et al.*¹². The validation for structure models was performed by using PROCHECK, as discussed elsewhere^{13,14}, which provides satisfactory results suggesting reliability of the model, as discussed by Sehgal *et al.*¹⁵. The model was selected on the basis of various factors such as overall G-factor, number of residues in core that fall in generously allowed and disallowed regions in Ramachandran plot. The model was further analyzed by WHATIF, as discussed elsewhere^{16,17} QMEAN^{18,19} and ProSA, as discussed by Wiederstein and Sippl²⁰. The ProSA was used for the display of Z-score and energy plots.

RESULTS AND DISCUSSION

Building of protein model: The basic principle of homology modeling is the selection of template and sequence alignment between the target and the template, as discussed by Chhabra and Dixit²¹. Sequence alignment of TCF7L2 protein by using the phyre 2 server, revealed sequence homology with catenin binding domain (ID = 99%), which was selected as template for the model building of TCF7L2 protein. Total 41 residues (7% of query sequence) have been modelled with 99% confidence by the single highest scoring template, as discussed by Pitchai *et al.*²². To build the model, PSI-BLAST was done with the maximum E-value allowed for template being 0.005. Using catenin binding domain sequence modeling of TCF7L2 protein domains was done with the help of YASARA (Fig. 1).

Model reputation: The model showed good stereo chemical property in terms of overall G-factor value of -0.64 indicating that geometry of the model corresponds to the probability conformation with 67.9% residues in the core region of Ramachandran plot showing high accuracy of model predicted, as discussed elsewhere²³. The number of residues in allowed and generously allowed region was 20.5% and 11.5%, respectively and none of the residues were present in the disallowed region of the plot (Fig. 2). The above results indicate that the protein model is reliable, as discussed by Sahu and Shukla²⁴.

The verify 3D graph illustrate the compatibility of an atomic model (3D) with amino acid sequence, as discussed by Biswas²⁵ and score profile access the quality of the model, as discussed by Sahu and Shukla²⁴. The high score of 0.28 indicates that environment profile of the model is good (Fig. 3). Profile score above zero in the verify 3D graph, as discussed elsewhere^{12,26} corresponds to acceptable

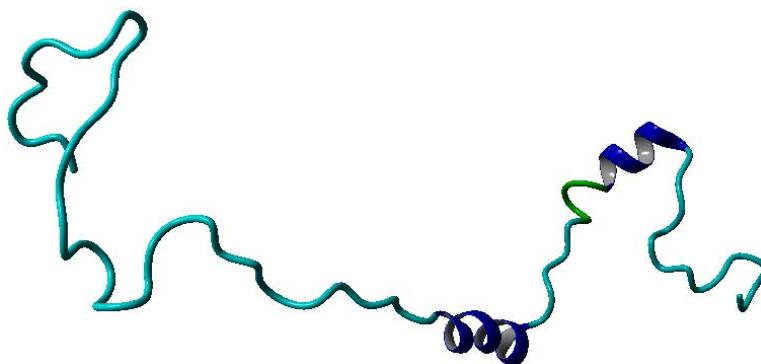


Fig. 1: TCF7L2 protein ribbon model generated using YASARA

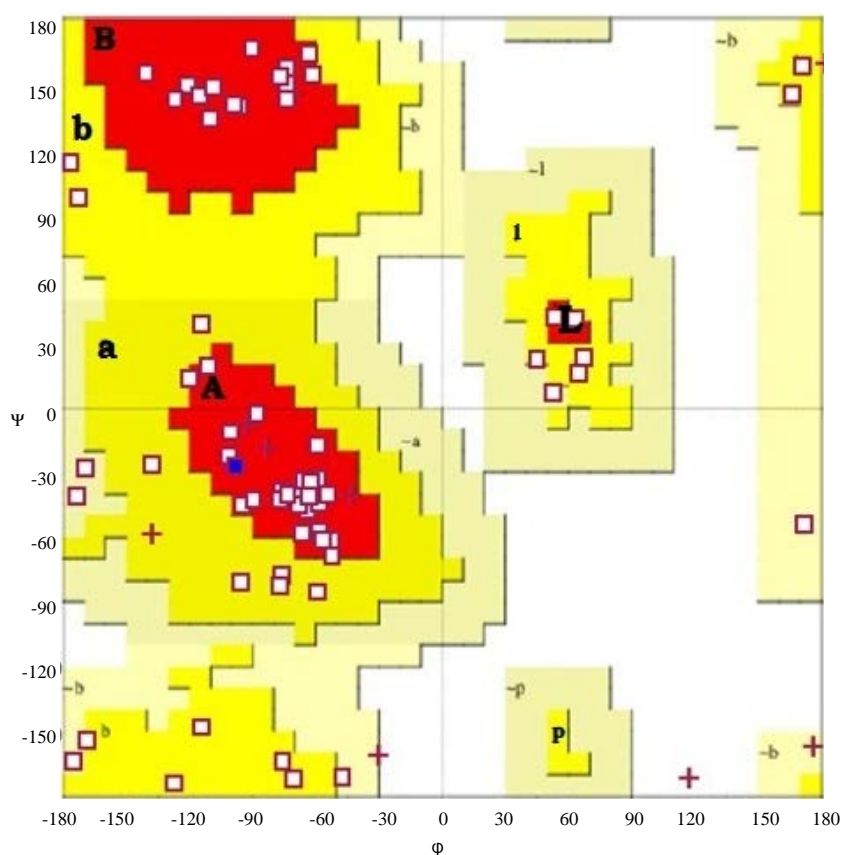


Fig. 2: Ramachandran plot analysis of TCF7L2 protein. Total number of residues were 156 with 67.9% in most favored regions [A, B, L], 20.5% in additional allowed regions [a,b,l,p], 11.5% in generously allowed regions and 0% in disallowed regions

environment of the model. In verified 3D plot, 17.50% of the residues had an averaged 3D-1D score ≥ 0.2 .

Model validation: ProSA was used to check the three-dimensional model of TCF7L2 proteins for potential errors. The ProSA Z-score of -6.07 indicates the overall model quality and measures the deviation of the total energy of

TCF7L2 protein (Fig. 4). The predicted value of Z-score -6.07 was in a range characteristic of native proteins indicating very less erroneous structures, as discussed by Mustufa *et al.*²⁷.

The quality of estimated model is based on the QMEAN scoring function were normalized with respect to the number of interactions, as discussed by Benkert *et al.*²⁸. The QMEAN score of the model was 0.189 and the Z-score was -4.16, which

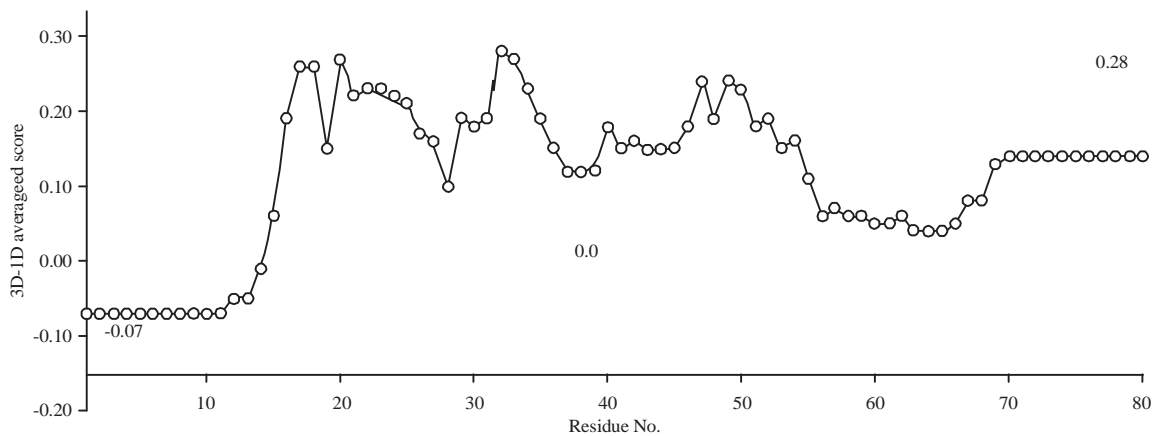


Fig. 3: Verified 3D graph of TCF7L2 protein [CAG38811]

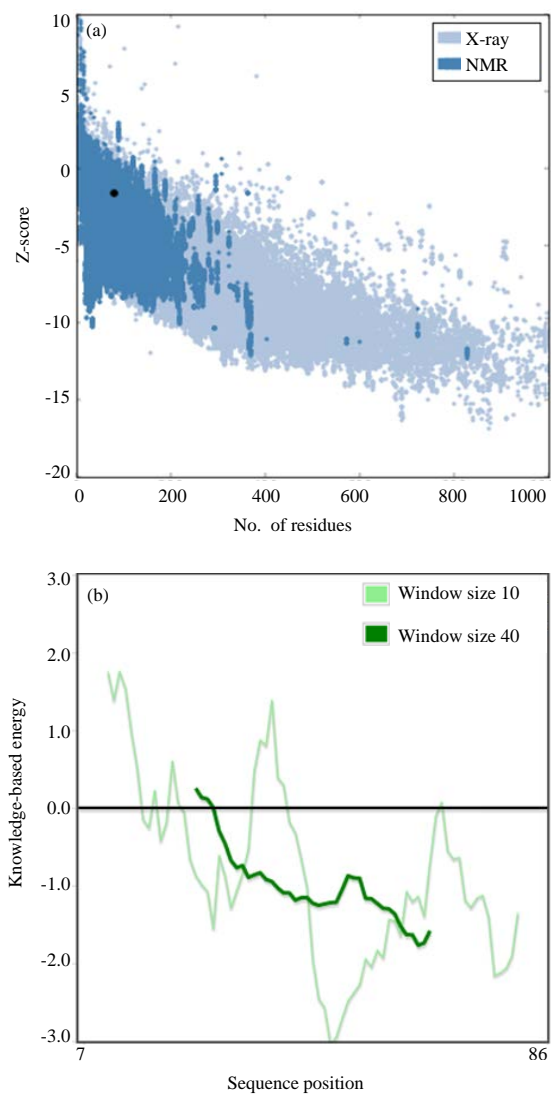


Fig. 4(a-b): ProSA web service analysis of TCF7L2 protein model

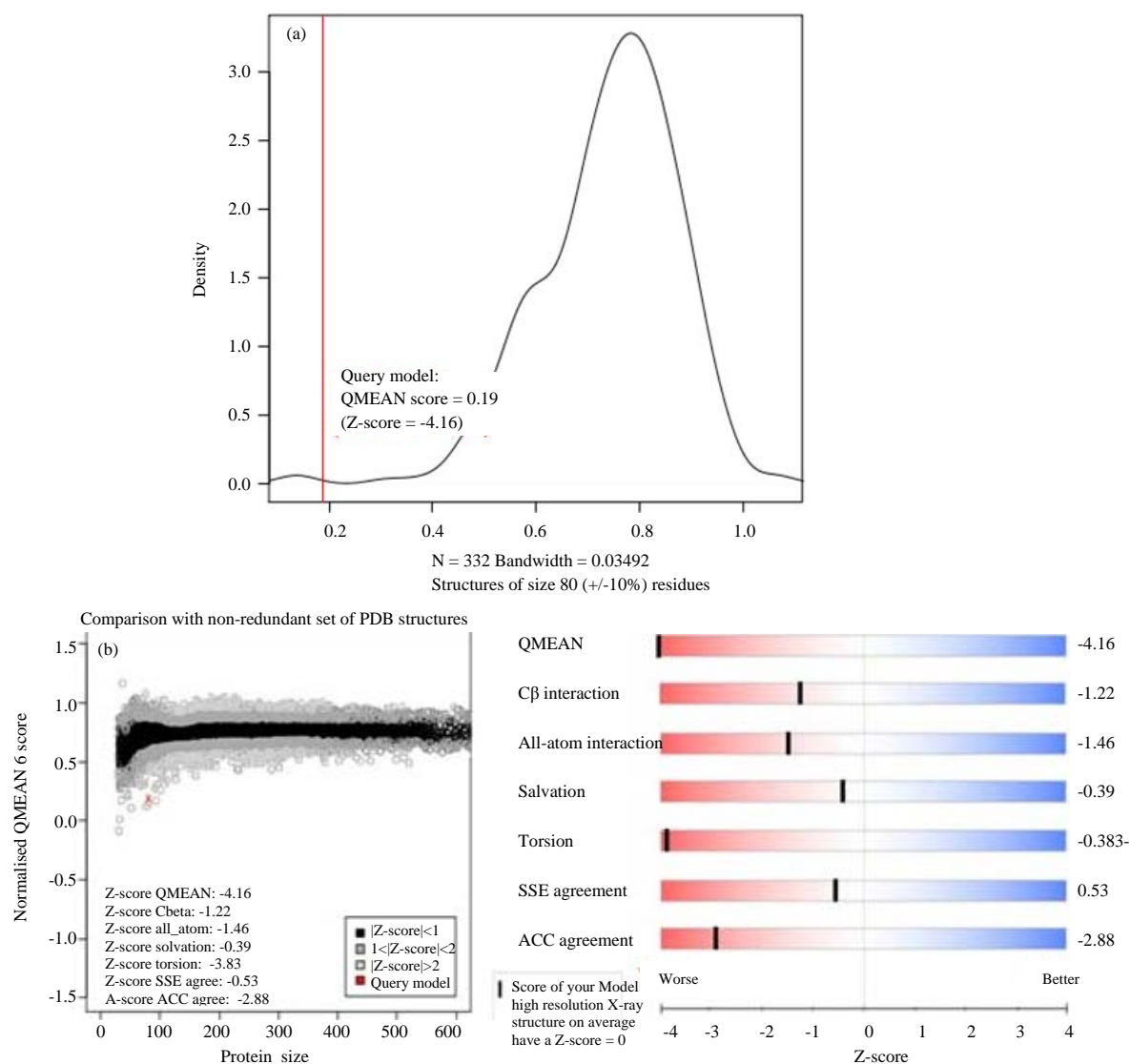


Fig. 5(a-b): (a) Density plot for QMEAN showing the value of Z-score and QMEAN score and (b) Plot showing the QMEAN value as well as Z-score

was very close to the value of 0 and this shows the fine quality of the model, as discussed elsewhere^{29,30} because the estimated reliability of the model was expected to be in between 0 and 1 and this could be inferred from the density plot for QMEAN scores of the reference set (Fig. 5a). A comparison between normalized QMEAN score (0.40) and protein size in non-redundant set of PDB structures in the plot revealed different set of Z-values for different parameters such as C-beta interactions (-1.22), interactions between all atoms (-1.46), solvation (-0.39), torsion (-3.83), SSE agreement (-0.53) and ACC agreement (-2.88) (Fig. 5b). The Z-score measures the total energy deviation of the TCF7L2 protein structure with

respect to an energy distribution derived from random conformations, as discussed by Rekić *et al.*³¹.

CONCLUSION

The generated model could be supportive to understand the functional characteristics of transcription factor 7-like 2 (TCF7L2). The variants in TCF7L2 associated with the risk for type 2 diabetes. The *in silico* molecular modeling and validation studies is helpful to understand the structure, function and mechanism of proteins action. The structure validation of generated model was done by

using WHATIF, PROCHECK, ProSA and QMEAN confirmed the reliability of the model.

The model showed good stereo-chemical property in terms of overall G-factor value of -0.64 indicating that geometry of model corresponds to the probability conformation with 67.9% residue in the core region of Ramachandran plot showing high accuracy of model prediction. The Z-score of -6.07 predicted by ProSA represents the good quality of the model. The Z-score also measures the divergence of total energy of the structure with respect to an energy distribution derived from random conformations. The scores indicate a highly reliable structure and are well within the range of scores typically found for proteins of similar size. The energy plot shows the local model quality by plotting knowledge-based energies as a function of amino acid sequence position.

ACKNOWLEDGMENTS

The authors are thankful to Dr. M.R. Jape (Principal, Rama Medical College, Ghaziabad Uttar Pradesh, India) for their precious support and approval of present study project. The thanks also to the Bioinformatics and Biochemistry Laboratory research group members for technical support.

SIGNIFICANCE STATEMENT

- As we know that the transcription factor 7-like 2 (TCF7L2) recently implicated in the pathogenesis of type 2 diabetes (T2D). The present study was designed to enlist some of the physiochemical and functional properties of TCF7L2 protein and provides information about its 3D structure
- The generated model could be supportive to understand the characteristics of TCF7L2. The *in silico* molecular modeling and validation studies of TCF7L2 is helpful to understand the structure, function and mechanism of proteins action
- The predicted model of TCF7L2 is useful for finding interactions with other proteins involved in type 2 diabetes (T2D). These findings can be used to future investigation about the molecular basis of type 2 diabetes (T2D) and new drug designing for treatment

REFERENCES

1. Anna, L.G., M. Braun and P. Rorsman, 2009. Type 2 diabetes susceptibility gene TCF7L2 and its role in β -cell function. *Diabetes*, 58: 800-802.

2. Yi, F., P.L. Brubaker and T. Jin, 2005. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by β -catenin and glycogen synthase kinase-3 β . *J. Biol. Chem.*, 280: 1457-1464.

3. Damcott, C.M., T.I. Pollin, L.J. Reinhart, S.H. Ott and H. Shen *et al.*, 2006. Polymorphisms in the Transcription Factor 7-Like 2 (*TCF7L2*) gene are associated with type 2 diabetes in the Amish: Replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes*, 55: 2654-2659.

4. Shah, M., R.T. Varghese, J.M. Miles, F. Piccinini and C.D. Man *et al.*, 2016. TCF7L2 genotype and α -cell function in humans without diabetes. *Diabetes*, 65: 371-380.

5. Grant, S.F., G. Thorleifsson, I. Reynisdottir, R. Benediktsson and A. Manolescu *et al.*, 2006. Variant of Transcription Factor 7-Like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.*, 38: 320-323.

6. Ng, M.C.Y., C.H.T. Tam, V.K.L. Lam, W.Y. So, R.C.W. Ma and J.C.N. Chan, 2007. Replication and identification of novel variants at *TCF7L2* associated with type 2 diabetes in Hong Kong Chinese. *J. Clin. Endocrinol. Metab.*, 92: 3733-3737.

7. Shu, L., N.S. Sauter, F.T. Schulthess, A.V. Matveyenko, J. Oberholzer and K. Maedler, 2008. Transcription factor 7-like 2 regulates β -cell survival and function in human pancreatic islets. *Diabetes*, 57: 645-653.

8. Naulaerts, S., P. Meysman, W. Bittremieux, T.N. Vu, W.V. Berghe, B. Goethals and K. Laukens, 2015. A primer to frequent itemset mining for bioinformatics. *Briefings Bioinform.*, 16: 216-231.

9. Rasouli, H. and B. Fazeli-Nasab, 2014. Structural validation and homology modeling of lea 2 protein in bread wheat. *Am.-Eurasian J. Agric. Environ. Sci.*, 14: 1044-1048.

10. Pevzner, P. and R. Shamir, 2011. *Bioinformatics for Biologists*. 1st Edn., Cambridge University Press, Cambridge, UK., ISBN-13: 978-1107648876, Pages: 394.

11. Prajapat, R., A. Marwal, Z. Shaikh and R.K. Gaur, 2012. *Geminivirus Database (GVDB): First database of family geminiviridae and its genera Begomovirus*. *Pak. J. Biol. Sci.*, 15: 702-706.

12. Luthy, R., J.U. Bowie and D. Eisenberg, 1992. Assessment of protein models with three-dimensional profiles. *Nature*, 356: 83-85.

13. Laskowski, R.A., M.W. MacArthur, D.S. Moss and J.M. Thornton, 1993. *PROCHECK*: A program to check the stereochemical quality of protein structures. *J. Applied Cryst.*, 26: 283-291.

14. Vriend, G., 1990. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graphics*, 8: 52-56.

15. Sehgal, S.A., R.A. Tahir, S. Shafique, M. Hassan and S. Rashid, 2014. Molecular modeling and docking analysis of CYP1A1 associated with head and neck cancer to explore its binding regions. *J. Theoret. Comput. Sci.*, Vol. 1.

16. Agrawal, P., Z. Thakur and M. Kulharia, 2013. Homology modeling and structural validation of tissue factor pathway inhibitor. *Bioinformation*, 9: 808-812.
17. Benkert, P., S.C.E. Tosatto and D. Schomburg, 2008. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Struct. Funct. Bioinform.*, 71: 261-277.
18. Benkert, P., M. Kunzli and T. Schwede, 2009. QMEAN server for protein model quality estimation. *Nucleic Acids Res.*, 37: W510-W514.
19. Novotny, W.F., T.J. Girard, J.P. Miletich and G.J. Broze, 1988. Platelets secrete a coagulation inhibitor functionally and antigenically similar to the lipoprotein associated coagulation inhibitor. *Blood*, 72: 2020-2025.
20. Wiederstein, M. and M.J. Sippl, 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, 35: W407-W410.
21. Chhabra, G. and A. Dixit, 2013. Structure modeling and antidiabetic activity of a seed protein of *Momordica charantia* in Non-Obese Diabetic (NOD) mice. *Bioinformation*, 9: 766-770.
22. Pitchai, D., V. Gopalakrishnan, V. Periyasamy and R. Manikkam, 2012. *In-silico* modeling and docking studies of AA_{2b}R with catechin to explore the anti diabetic activity. *Int. J. Pharm. Pharmaceut. Sci.*, 4: 328-333.
23. Prajapat, R., A. Marwal, V. Bajpai and R.K. Gaur, 2011. Genomics and proteomics characterization of alphasatellite in weed associated with begomovirus. *Int. J. Plant Pathol.*, 2: 1-14.
24. Sahu, R. and N.S. Shukla, 2014. *In-silico* analysis of different plant protein and their essential compound with sulfonyleurea binding protein of β -cells of homo sapiens for curing diabetes mellitus type II disease. *Eur. Chem. Bull.*, 3: 568-576.
25. Biswas, P., 2014. *In silico* approach to develop structure and functional analysis of response receiver regulator protein of the strain pseudomonas fulva 12x. *IOSR J. Pharm. Biol. Sci.*, 9: 79-86.
26. Bowie, J.U., R. Luthy and D. Eisenberg, 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253: 164-170.
27. Mustufa, M.M.A., S. Chandra and S. Wajid, 2014. Homology modeling and molecular docking analysis of human RAC- α serine/threonine protein kinase. *Int. J. Pharma Bio Sci.*, 5: 1033-1042.
28. Benkert, P., M. Biasini and T. Schwede, 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27: 343-350.
29. Prajapat, R., A. Marwal and R.K. Gaur, 2014. Recognition of errors in the refinement and validation of three-dimensional structures of AC1 proteins of begomovirus strains by using ProSA-web. *J. Viruses*, Vol. 2014. 10.1155/2014/752656.
30. Wiederstein, M. and M.J. Sippl, 2005. Protein sequence randomization: Efficient estimation of protein stability using knowledge-based potentials. *J. Mol. Biol.*, 345: 1199-1212.
31. Rezik, I., Z. Chaabene, C.D. Grubb, N. Drira, F. Cheour and A. Elleuch, 2015. *In silico* characterization and molecular modeling of double-strand break repair protein *MRE11* from *Phoenix dactylifera* v deglet nour. *Theor. Biol. Med. Model.*, Vol. 12. 10.1186/s12976-015-0013-2.