



Original article

<https://doi.org/10.12980/jclm.5.2017J6-222>

©2017 by the Journal of Coastal Life Medicine. All rights reserved.

Homology modeling and structural validation of human amylin: A type 2 diabetes causing protein

Gousia Shaik¹, Janakiram Bobbilipati², Naveena Lavanya Latha Jeevigunta^{1*}¹Department of Biotechnology, Krishna University, Machilipatnam 521001, India²Department of Microbiology, NRI Medical College & General Hospital, Guntur, India

ARTICLE INFO

Article history:

Received 24 Oct 2016

Received in revised form 23 Dec 2016

Accepted 14 Apr 2017

Available online 16 Jun 2017

Keywords:

Human amylin

SOPMA

BLAST

PROCHECK

SWISS MODEL

ABSTRACT

Objective: In the present study, the protein structure model of human amylin was generated to understand the protein's structure, function and mechanism of the action.

Methods: The stereo chemical quality of the protein model was checked by using *in silico* analysis with SWISS MODEL, SOPMA, PROCHECK, ProSA and QMEAN servers.

Results: The 66.7% residues in the core region of Ramachandran plot showing high accuracy of protein model and the QMEAN Z score of -6.57 indicates the overall model quality of amylin protein.

Conclusions: These results may be helpful for further investigation on developing target molecules for amylin proteins that play a key role in type 2 diabetes.

1. Introduction

Amylin [islet amyloid polypeptide (IAPP)] is a peptide hormone that is co-stored and co-secreted with insulin and shares the same processing enzymes[1]. The amylin gene is being transcribed from a 89 amino acid precursor protein that was known to locate on the short arm of chromosome 12. PC1, PC2, and PC3 are the pre-proprotein convertases responsible for the processing of prehormones to the active secreted hormones like insulin and amylin. It is primarily PC2 that is responsible for amylin processing, within the secretory granule and is responsible for converting the prohormone (89 amino acid) to the actively secreted amylin (37 amino acid)[2]. Human amylin shows physicochemical properties inclining the peptide hormone to total and shape amyloid filaments, which may have impact in β -cell decimation in type 2 diabetes[1]. The amyloid fibrils are the β -sheet which exhibits frame from more than 20 distinctive antecedent proteins that create totals in different tissues bringing about pathogenesis. The requested structure of amyloid stores may add to illness pathogenesis through cytotoxicity and cell demise[3]. The amylin hormone has a flag transduction pathway like that of calcitonin (CT), calcitonin gene-related peptide (CGRP), and adrenomedullin, and the specificity of the amylin

receptor has been described. Amylin has been appeared to have restricting destinations inside the renal cortex in the zone of the juxtaglomerular mechanical assembly. Amylin has been appeared to actuate the rennin angiotensin aldosterone framework[2]. In the present study, various bioinformatics tools were employed for characterization of human IAPP that can be utilized for developing target molecules in future.

2. Materials and methods

2.1. BLAST P, multiple sequence alignment

The FASTA sequence of human amylin protein was retrieve from NCBI. The protein sequences are scanned by using the BLAST P algorithm we can obtain the homologous protein sequences from the available protein sequences of various organisms. BLAST search[4] against the primary amino acid sequence contained in the SMTL was performed. An initial HHblits profile has been built using the procedure outlined in the previous studies[5-8], followed by 1 iteration of HHblits against NR20. The obtained profile has then been searched against all profiles of the SMTL. A total of 18 templates were found. For each identified template, the template's quality has been predicted from features of the target-template alignment. The templates with the highest quality have then been selected for model building. Phylogenetic tree was then constructed using phylogeny.fr (<http://www.phylogeny.fr/>) to determine the evolutionary relationships.

*Corresponding author: Naveena Lavanya Latha Jeevigunta, Department of Biotechnology, Krishna University, Machilipatnam 521001, India.

Tel: +919290907081

E-mail: jnlavanyalatha@yahoo.co.in

The journal implements double-blind peer review practiced by specially invited international editorial board members.

2.2. Secondary structure prediction

Secondary structure of calcitonin family protein was predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) tool in ExPasy.

2.3. Homology modeling

The sequence of human amylin was downloaded from the universal protein resource[9](Uniprot KB) (<http://www.uniprot.org/>) (entry ID: P10997). The PDB file of human amylin protein was generated by SWISS MODEL servers by using its FASTA sequence. Multiple sequence alignment was performed between full length of human amylin protein sequence and template protein sequences in order to build a model of protein domain, in this database. The suitable template for homology modeling was identified through searching human amylin on PDB using the BLAST P algorithm[10]. The 3D structure of human amylin was downloaded from PDB (PDB

ID: 2kb8.1.A) as the template structure.

2.4. Model reputation

The structure validation for models was performed by utilizing PROCHECK[11,12] which gives palatable results proposing dependability of the model[13]. The model was chosen on the premise of different components, for example, general G-calculate, number of buildups in center that fall in liberally permitted and refused locales in Ramachandran plot (Figure 1). The model was further investigated by QMEAN[14,15] and ProSA[16]. ProSA was utilized for the show of Z-score and energy plots.

3. Results

Scanning of protein sequence databases using BLAST P with the FASTA sequence obtained from NCBI was used to obtain the MSA with an entry ID: P10997 which showed that the sequence is an islet

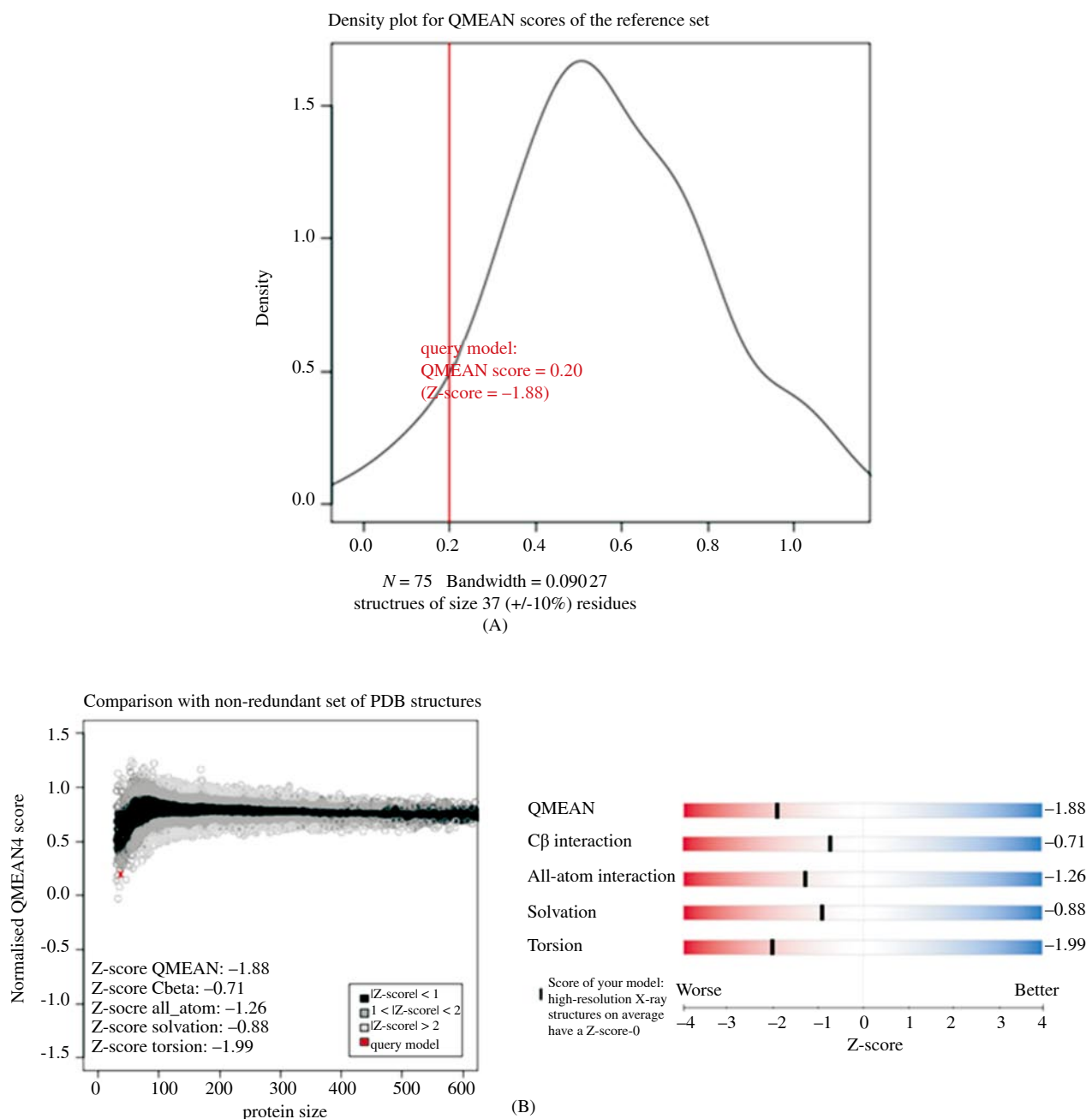


Figure 1. The density plot (A) for QMEAN showing the value of Z-score and QMEAN score (B) plot showing the QMEAN value as well as Z-score.

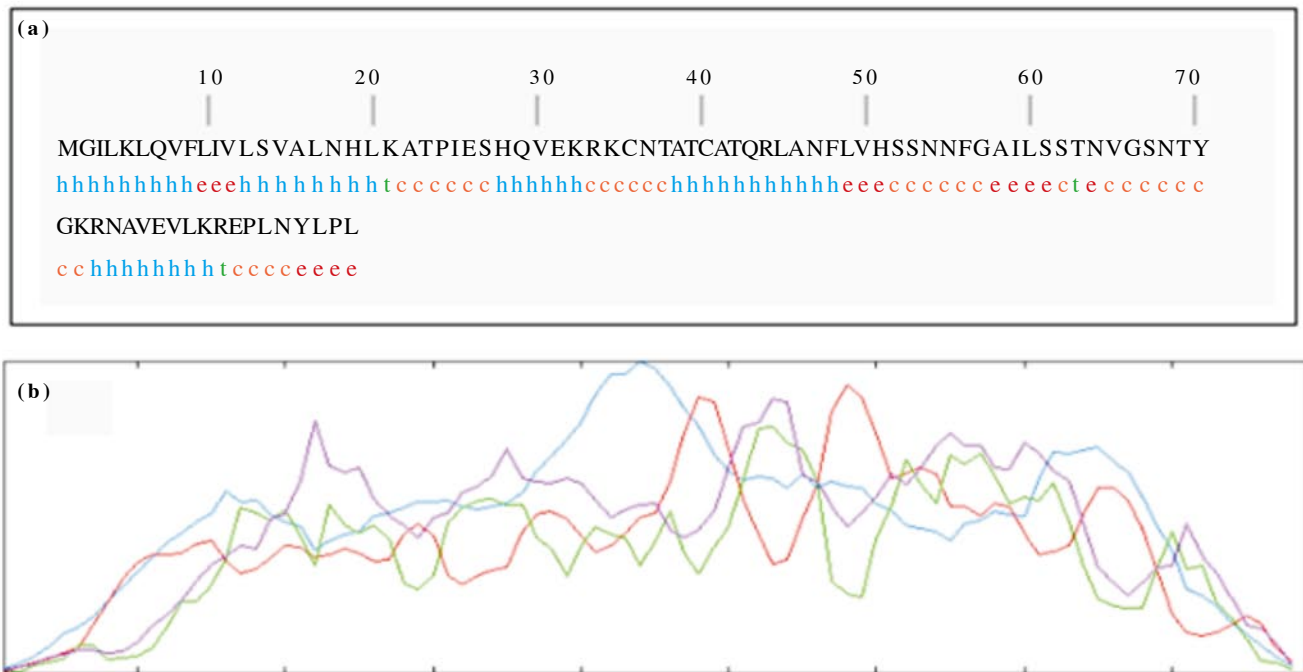


Figure 2. Secondary structure of human amylin. (a) Sequence length: 89; Alpha helix (Hh): 41 is 46.07%; Extended strand (Ee): 15 is 16.85%; Beta turn (Tt): 3 is 3.37%; Random coil (Cc): 30 is 33.71%. (b) Distribution of secondary structure elements of human amylin. Blue line: Alpha helix; Red: Extended strand; Green: Beta turn; Orange: Random coil.

Model_07	MGILKLVFLIVLSVALNHLKATPIESHQVEKRKCNTATCATQRLANFLVHSSNFGAILSSTNVGSNTY	75
2kb8.1.A	----- KCNTATCATQRLANFLVHSSNFGAILSSTNVGSNTY -----	37
Model_07	VEVLKREPLNYLPL	89
2kb8.1.A	-----	

Figure 3. Target template.

amyloid polypeptide. A phylogram constructed based on multiple sequence alignment using phylogeny.fr revealed that calcitonin was closely related to a conserved islet amyloid polypeptide.

2^o structure of the target protein was predicted by using SOPMA tool in ExPasy (Figure 2). The results indicate that calcitonin has 46.07%, α -helix thus making it stable for homology modeling[17].

The crystal structure of islet amyloid polypeptide (2kb8.1.A) with a sequence identity of 100% to the target sequence was selected based on BLAST P search against PDB database (Table 1). The sequence alignment between the template (2kb8.1.A) and the target was shown in Figure 3.

3.1. Model reputation

The 3D structure of modeled h IAPP was shown in Figure 4 as generated by Swiss Model server. The number of residues in allowed region and disallowed region was 27.3% and 6.1%, respectively, and none of the residues were present in the generously allowed regions of the plot (Figure 5). The above results indicate that the protein model is reliable[18].



Figure 4. 3D structure of human amylin generated by using Jmol-SWISS MODEL server.

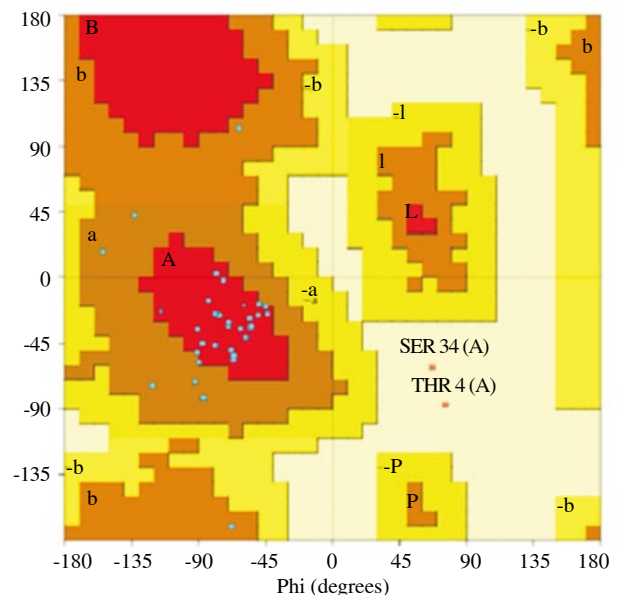


Figure 5. Ramachandran plot analysis of human amylin protein. Total number of residues were 37 with 66.7% in most favored regions (A, B, L), 27.3% in additional allowed regions (a, b, l, p), 0.0% in generously allowed regions and 6.1% in disallowed regions.

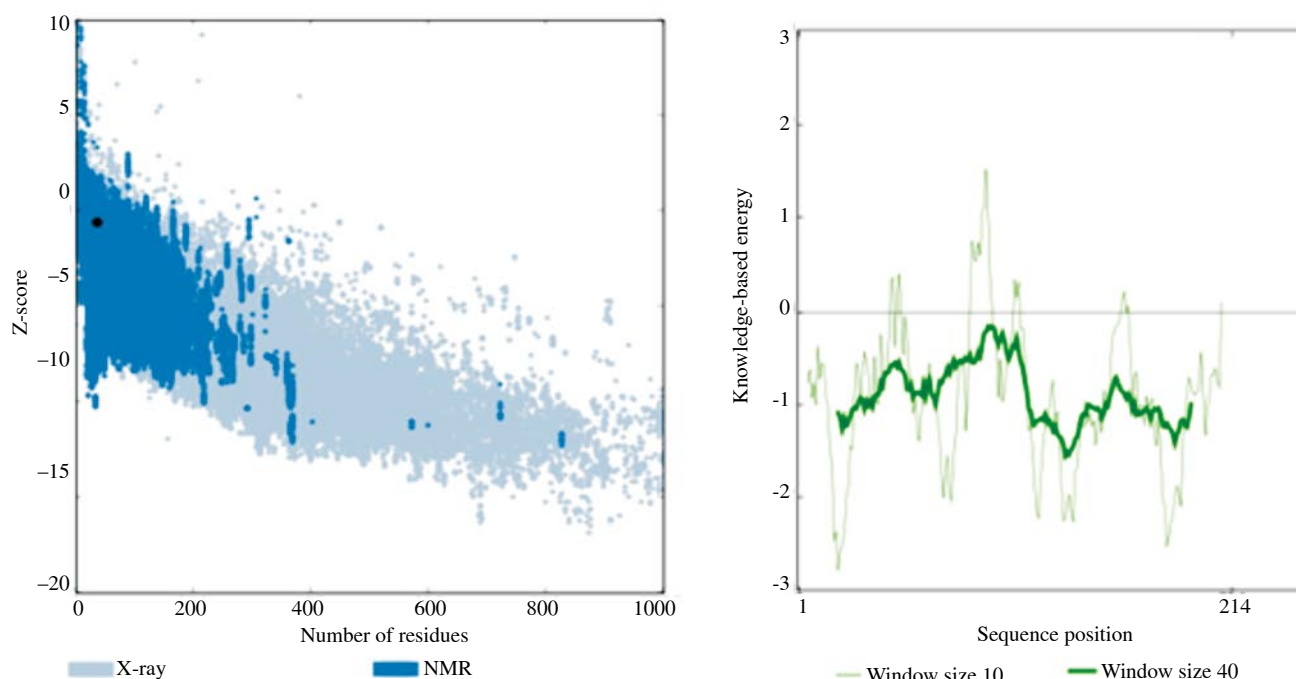
3.2. Model validation

ProSA was used to check the three-dimensional model of human amylin proteins for potential errors. The ProSA Z-score of -6.57

Table 1

Showing overall 19 templates.

Template	Seq. Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Coverage	Description
3g7v.1.A	90.24	Homo-dimer	HHblits	X-ray	1.86Å	0.57	0.46	Maltose-binding periplasmic protein, Islet amyloid polypeptide fusion protein
2l86.1.A	97.37	Monomer	HHblits	NMR	NA	0.59	0.43	Islet amyloid polypeptide
2kb8.1.A	100.00	Monomer	HHblits	NMR	NA	0.61	0.42	Islet amyloid polypeptide
2g48.1.B	100.00	Hetero-oligomer	HHblits	X-ray	2.60Å	0.61	0.42	Islet amyloid polypeptide
2g48.2.B	100.00	Hetero-oligomer	HHblits	X-ray	2.60Å	0.61	0.42	Islet amyloid polypeptide
2l86.1.A	100.00	Monomer	BLAST	NMR	NA	0.61	0.42	Islet amyloid polypeptide
3g7v.1.A	100.00	Homo-dimer	BLAST	X-ray	1.86Å	0.61	0.42	Maltose-binding periplasmic protein, Islet amyloid polypeptide fusion protein
2kj7.1.A	81.58	Monomer	HHblits	NMR	NA	0.54	0.43	Islet amyloid polypeptide
2kj7.1.A	83.78	Monomer	BLAST	NMR	NA	0.55	0.42	Islet amyloid polypeptide
3lu2.1.A	14.00	Homo-dimer	HHblits	X-ray	2.20Å	0.27	0.56	Lmo2462 protein
2l7s.1.A	15.91	Monomer	HHblits	NMR	NA	0.30	0.49	Adrenomedullin
2glg.1.A	31.25	Monomer	HHblits	NMR	NA	0.34	0.36	Calcitonin-1
1fb9.1.A	31.25	Monomer	HHblits	NMR	NA	0.34	0.36	Calcitonin analogue
2glh.1.A	31.25	Monomer	HHblits	NMR	NA	0.34	0.36	Calcitonin-1
1byv.1.A	28.13	Monomer	HHblits	NMR	NA	0.33	0.36	Protein (calcitonin)
1bku.1.A	28.13	Monomer	HHblits	NMR	NA	0.33	0.36	Calcitonin
1bzb.1.A	28.13	Monomer	HHblits	NMR	NA	0.33	0.36	Protein (calcitonin)
3g7w.1.A	100.00	Monomer	BLAST	X-ray	1.75Å	0.62	0.25	Maltose-binding periplasmic protein, Islet amyloid polypeptide fusion protein
2jxz.1.A	25.00	Monomer	HHblits	NMR	NA	0.31	0.36	Calcitonin

**Figure 6.** ProSA web service analysis of human amylin protein model.

indicates the overall model quality and measures the deviation of the total energy of human amylin protein (Figure 6).

The QMEAN score of the model was 0.20 and the Z-score was -1.88 , which was very close to the value of 0 and this shows the fine quality of the model 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 (Figure 6).

4. Discussion

The basic principle of homology modeling is the selection of

template and sequence alignment between the target and the template[19]. This was met by performing a BLAST P search against known protein structures deposited in PDB. The studies of Rost[20] and Yang and Honig[21] demonstrated that 3D structures will be similar if the sequence identity between target and template proteins is higher than 25%. Generally, a target which shares a sequence similarity of 30% or more to an experimentally solved protein structure (template) can only be employed for homology modeling.

Good stereochemical property was shown for the target protein as observed in terms of overall G-factor value of -0.53 indicating that geometry of the model corresponds to the probability conformation

with 66.7% residues in the core region of Ramachandran plot showing high accuracy of model predicted[13,22].

The predicted value of Z-score -6.57 was in a range characteristic of native proteins indicating very less erroneous structures[23].

The quality of estimated model is based on the QMEAN scoring function which was normalized with respect to the number of interactions[24]. 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 (-0.71), interactions between all atoms (-1.26), solvation (-0.88), torsion (-1.99)[25,26]. The Z-score measures the total energy deviation of the human amylin protein structure with respect to an energy distribution derived from random conformations[27].

The generated models in this paper can be very useful to understand the functional characteristics of the human amylin that is responsible in causing the 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 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 -0.53 indicating that geometry of model corresponds to the probability conformation with 66.7% residue in the core region of Ramachandran plot showing high accuracy of model prediction. The Z score of -6.57 predicted by ProSA represents the good quality of the model. 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.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

The authors greatly acknowledge the Krishna University, Machilipatnam and NRI Medical College, Guntur authorities for providing the facilities for the current research.

References

- [1] Schmitz O, Brock B, Rungby J. Amylin agonists: a novel approach in the treatment of diabetes. *Diabetes* 2004; **53**(Suppl 3): S233-8.
- [2] Hayden MR, Tyagi SC. "A" is for amylin and amyloid in type 2 diabetes mellitus. *JOP* 2001; **2**(4): 124-39.
- [3] Bhattacharya S, Latha JN, Kumresan R, Singh S. Cloning and expression of human islet amyloid polypeptide in cultured cells. *Biochem Biophys Res Commun* 2007; **356**(3): 622-8.
- [4] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**(17): 3389-402.
- [5] Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 2011; **9**(2): 173-5.
- [6] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* 2006; **22**: 195-201.
- [7] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997; **18**: 2714-23.
- [8] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 2003; **31**: 3381-5.
- [9] UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 2012; **40**: D71-5.
- [10] Rajasekaran S, Nick H, Pardalos P, Sahni S, Shaw G. Efficient algorithms for local alignment search. *J Combinatorial Optimiz* 2001; **5**: 117-24.
- [11] Dastmalchi S, Hamzeh-Mivehroud M, Sokouti B. *Methods and algorithms for molecular docking-based drug design and discovery advances in medical technologies and clinical practice*. Hershey: IGI Global; 2016, p. 339-60.
- [12] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993; **26**: 283-91.
- [13] 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 Theor Comput Sci* 2014; **1**: 112.
- [14] Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res* 2009; **37**: 4-10.
- [15] Novotny WF, Girard TJ, Miletich JP, Broze GJJ. Platelets secrete a coagulation inhibitor functionally and antigenically similar to the lipoprotein associated coagulation inhibitor. *Blood* 1988; **72**: 2020-5.
- [16] 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**: 407-10.
- [17] Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. *TIBS* 2000; **25**(3):147-50.
- [18] Sahu R, Shukla N. *In silico* analysis of different plant protein and their essential compound with sulfonylurea binding protein of β -cells of homo sapiens for curing diabetes mellitus type II disease. *Eur Chem Bull* 2014; **3**: 568-76.
- [19] Chhabra G, Dixit A. Structure modeling and antidiabetic activity of a seed protein of momordicacharantia in non-obese diabetic (NOD) mice. *Bioinformation* 2013; **9**: 766-70.
- [20] Rost B. Twilight zone of protein sequence alignments. *Protein Eng* 1999; **12**(2): 85-94.
- [21] Yang AS, Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J Mol Biol* 2000; **301**: 691-711.
- [22] Prajapat R, Marwal A, Bajpai V, Gaur RK. Genomics and proteomics characterization of alphasatellite in weed associated with begomovirus. *Int J Plant Pathol* 2011; **2**: 1-14.
- [23] Mohammed M, Mustufa A, Chandra S, Wajid S. Homology modeling and molecular docking analysis of human RAC-alpha serine/threonine protein kinase. *Int J Pharm Bio Sci* 2014; **5**: 1033-42.
- [24] Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 2011; **27**: 343-50.
- [25] Wiederstein M, Sippl MJ. Protein sequence randomization: efficient estimation of protein stability using knowledge based potentials. *J Mol Biol* 2005; **345**: 1199-212.
- [26] 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; doi: 10.1155/2014/752656.
- [27] Rekić I, Chaabene Z, Douglas Grubb C, 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.