Journal of Proteomics & Bioinformatics
ISSN: 0974-276X
Appaiah and Vasu, J Proteomics Bioinform 2016, 9:11
DOI: 10.4172/jpb.1000417

Research Article Open Access

In silico Designing of Protein Rich in Large Neutral Amino Acids Using Bovine α s1 Casein for Treatment of Phenylketonuria

Prakruthi Appaiah and Prasanna Vasu*

Department of Food Safety and Analytical Quality Control Laboratory, CSIR-Central Food Technological Research Institute, Mysuru-570020, Karnataka, India

Abstract

Phenylketonuria (PKU) is a genetically inherited disease where the body fails to convert Phenylalanine (Phe) to Tyrosine (Tyr) due to the defective Phenylalanine Hydroxylase (PAH) enzyme, resulting in the elevated blood Phe level causing neurological damage. Of all therapies, Large Neutral Amino Acid (LNAA) supplementation has become a promising approach to the dietary treatment of PKU, where the LNAA compete with Phe for the same L-Type Large Neutral Amino Acid Transporter (LAT1, SLC7A5) across the blood-brain barrier, decreasing brain Phe level. In this study we have designed an easily digestible protein enriched with LNAA (except Phe), using bovine αs1 casein as template by homology modeling. Our challenge was to maximize the LNAAs (except Phe) in the protein model by finding a suitable scaffold for hosting LNAAs, thereby turning the usual concept of homology modeling. Bioinformatics tools like SWISS-MODEL, EXPASY, PROFUNC, I-TASSER, RaptorX, and SAVeS Server were used for the structure prediction, and validation of the designed protein. Out of 63 different models designed, Protein Model-54 was selected based on sequence similarity to template (61.4%), compact 3D structure containing only α-helices and coils without β-sheets (QMEAN score of 0.567), and good *in silico* digestibility. The molecular weight of protein was 12,094.6 Da. Ramachandran plot revealed that the designed protein contained 89.9% amino acid residues in the favored region with ERRAT score of 88.04. Based on these evaluations, the Protein Model-54 was found to be the best, stable and reliable model which may be of high nutritional significance for PKU patients..

Keywords: Phenylketonuria; Large neutral amino acids; SWISS-MODEL; I-TASSER; SAVeS server

Introduction

PKU is a genetically inherited disease. Though it occurs one in 18,300 persons in Indian populations [1], development of suitable food protein formulations for PKU is still the main focus of many food researches in India. Food proteins contain different amount of Phe and consumption of correct amount is required for the proper protein synthesis, normal growth and development of the body and brain in growing children [2]. In PKU patients, the body fails to eliminate the excess Phe due to the defective enzyme Phenylalanine Hydroxylase (PAH) resulting in the elevated blood Phe level causing neurological damage [3]. Many therapies like gene therapy, enzyme substitution therapy, and protein hydrolysate diet have so far been implemented. But, patients fail to follow the diet as it is unpalatable, or the therapies are too expensive. However, supplementation with large neutral amino acids (LNAA, i.e., histidine, isoleucine, leucine, methionine, tyrosine, threonine, tryptophan, and valine) other than Phe has become a promising approach in the dietary treatment of PKU as it increases the melatonin and neurotransmitter synthesis resulting in improved brain function [4,5]. The balanced ratio of histidine: isoleucine: leucine: methionine: tyrosine: threonine: tryptophan: valine is 6: 9: 20: 8: 6: 9: 3: 13, respectively, as nutritionally specified for infants by WHO/FAO/UNU [6].

Normally, all the LNAAs compete for transport across the blood-brain barrier via the same L-type large neutral amino acid transporter (LAT1, SLC7A5). LAT1 transporter has a high affinity for Phe, resulting in an increased transport velocity of Phe even at slightly higher concentration in the blood [7]. Thus, supplementation of LNAA has become a promising approach for the dietary treatment of PKU. However, protein enriched with LNAA and devoid of Phe is not found in nature. Our search for a protein containing low levels of Phe resulted in the identification of αs1 casein protein, which is used as a template for protein homology modeling. In milk, αs1 casein is the major fraction comprising of 40% of the casein fraction [8]. According to Sanchez and co-workers [9], long-term consumption of milk casein hydrolysate provides cardiovascular benefits and reduces hypertension.

Protein designing has been extensively used to increase the essential amino acid contents in proteins for the enhancement of its nutritional quality [10]. Protein designing can be performed by homology modeling, which predicts the 3D structure of the target protein based on its alignment to the template whose structure has already been experimentally determined [11]. The four main steps in homology modeling include fold assignment and template selection, template-target alignment, model building and model evaluation. Template selection, alignment and model building is done until a satisfactory model is obtained [12]. The model should have above 30% identity to the template so that it can be modeled with accuracy [13,14]. In the model building, the challenge is to maximize the LNAA content (and to remove Phe) in the protein using a particular secondary structure. Thus, we turned the usual concept of homology modeling and tried to find a suitable scaffold (especially α-helix) to host the enriched LNAAs for increased functionality (stability and digestibility). The proper folding of the designed protein into 3D structure is important for its biological activity. The designing was based on the binary patterning using Polar (P) and Non-polar (N) amino acids [15]. A high percentage of β-sheets

***Corresponding author:** Prasanna Vasu, Senior Scientist, Department of Food Safety and Analytical Quality Control Laboratory, CSIR-Central Food Technological Research Institute, Mysore-570020, India, Tel: +91-0821-2514972; Fax: +91-0821-2412064; E-mail: vprasanna@cftri.res.in

Received October 06, 2016; **Accepted** November 11, 2016; **Published** November 16, 2016

Citation: Appaiah P, Vasu P (2016) *In silico* Designing of Protein Rich in Large Neutral Amino Acids Using Bovine αs1 Casein for Treatment of Phenylketonuria. J Proteomics Bioinform 9: 287-297. doi: 10.4172/jpb.1000417

may cause low access to gastrointestinal digestive enzymes resulting in low protein value and low protein bioavailability [16]. Understanding the structure of whole protein is important to know the digestive behavior, nutritive value, utilization and availability in animals [17].

Here we have shown how different bioinformatics tools can be utilized and applied to design a protein enriched with LNAA, and comprised mainly of α-helices secondary structures for easy digestion in human gastrointestinal tract. Bioinformatics online software like SWISS-MODEL, EXPASY tool, Allergenonline.com, PROFUNC, I-TASSER, RaptorX, SuperPose, ProSA-web and SAVeS Server were used for the structure prediction, evaluation and validation of the predicted structure. We envisage that this model protein can be used as a dietary supplement or treatment to increase the level of LNAAs (except Phe) in the blood, and thereby decreasing the Phe concentration in the brain, owing to their competition for the same LAT1 transporter. This designed protein will be of high nutritional significance for PKU patients.

Materials and Methods

Enrichment of αs1 casein with LNAA

The protein sequence of bovine αs1 casein (Accession: AAA30428.1) was obtained from NCBI in FASTA format (www.ncbi.nlm.nih.gov/protein/162794?report=fasta#) [18]. The first amino acid Met was retained as an initial amino acid after deleting the amino acid sequence KLLILTCLVAVALA from the N-terminal end of the protein sequence [19]. The FASTA sequence of αs1 casein and template are as follows (the underlined portions are deleted from the template):

>gi|162794|gb|AAA30429.1| alpha-S1-casein [*Bos taurus*]

M<u>KLLILTCLVAVALA</u>RPKHPIKHQGLPQEVLNENLLRFFVALFPEVFGKEKVNELSKDIGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLL<u>RLKKYKVPQLEIVPNSAEERLHSMKEGIDAQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSEKTTISLW</u>

The actual template is as follows:

MRPKHPIKHQGLPQEVLNENLLRFFVALFPEVFGKEKVNELSKDIGSESTEDQAMEDIKQMEASISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLL

The LNAAs were incorporated in excess levels or in accordance with their relative nutritional levels specified for infants by WHO/FAO/UNU [6]. The selection criteria for the best target model was (i) enrichment of LNAA (except Phe) with their balanced ratio of histidine: isoleucine: leucine: methionine: tyrosine: threonine: tryptophan: valine is 6: 9: 20: 8: 6: 9: 3: 13, respectively, in accordance with nutritionally specified by WHO/FAO/UNU [6], (ii) at least more than 50% of homology to the template protein, and (iii) three-dimensional structure prediction with only α-helix and devoid of β-sheets. The order of arrangement of Polar (P) and Non-polar (N) residues for α-helices was PNPPNNPPNPPNNP [15,20]. It allows the hydrophobic amino acids to burry inside and hydrophilic amino acids to expose to the solvent [21,22]. The order of arrangement for β-sheets was with alternating Polar (P) and Non-polar (N) residues, i.e., PNPNPNP [23,24].

Construction of protein model by SWISS-MODEL

SWISS-MODEL, ExPASy ProtParam, and Allergenonline.com programs were used to screen the target model and to predict the best target model. Homology modeling was performed to design the protein model. SWISS-MODEL server (http://swissmodel.expasy.org/interactive) was used to predict the 3-D structure of the designed protein based on homology or comparative modeling. The user provides a sequence to be modeled, and SWISS-MODEL automatically calculates a model containing all non-hydrogen atoms, and known templates were chosen by the server. The homology between the template and target was predicted, as well as sequence similarity and alignment of the selected template [25]. A total of 63 different models were designed, constructed and predicted using this software. The overall quality of the 3D structure was predicted by the SWISS-MODEL server. The QMEAN (Quality Model Energy Analysis) Score values assess the structure modeled by the server and the reliable scores should be between 0 and 1. SWISS-MODEL perform comparative protein structure modeling by satisfying the spatial restraints, and performing additional tasks, including *de novo* modeling of loops in protein structures, multiple alignment of protein sequences and structures, comparison of protein model structures, and so on. Each protein has different Z-score values in PDB, and therefore it is not known what range of Z-score is required for correct protein models [26].

Allergenonline.com and ClustalW2

Allergenonline.com is an online server that provides data on the sequence similarity with natural proteins, especially allergens. Allergenonline server (http://www.allergenonline.com/) was used to check if the designed protein is having any sequence similarity to known allergens which may be a potential risk of allergenic cross-reactivity. It also showed sequence similarity to the template. It gives the list of allergens to which the designed protein matches, and provides the percentage of similarity with the allergen by sequence search database. If the full length shows similarity more than 50%, then the sequence is a probable allergen, and should not be considered further. Additionally, if the sequence similarity of more than 35% for the 80 amino acid segment, then the sequence is considered as a probable allergen [27].

Pairwise sequence alignment between template and target proteins was performed using ClustalW2 server (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and EMBOSS Stretcher Pairwise Sequence Alignment (http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/) [28]. These servers were used to identify the regions of sequence similarity between the template and target models. Higher the sequence similarity better the protein model. It was noticed recently that the Clustal server is retired, so we did additional sequence homology with Pairwise sequence alignment using EMBOSS stretcher.

Physicochemical properties of the target protein models by ExPASy (Expert Protein Analysis System) tool

ExPASy ProtParam tool is an analysis tool used to predict various physicochemical parameters like molecular weight, theoretical pI, amino acid composition, atomic composition, extinction co-efficient, half-life, instability index, aliphatic index and grand average of hydropathicity index (GRAVY) (http://web.expasy.org/protparam/). The ExPASy ProtParam tool is used to analyze protein sequences, structures and physicochemical properties of protein models [29].

ExPASy Peptide Cutter

The main aim of this work is to design a protein which is easily digestible in the human gastrointestinal tract, and thus *in silico* digestibility was performed using the ExPASy Peptide Cutter tool (http://web.expasy.org/peptide_cutter/). Peptide cutter is used to predict the cleavage sites of proteases [29].

Secondary structure and tertiary structure prediction

Secondary structures include α-helix, β-sheets, coils and turns. The spatial arrangement of secondary structures results in the tertiary structure. X-ray crystallography and Nuclear Magnetic Resonance (NMR) are advanced techniques used to determine the structure of protein. Recently, computational methods like homology modeling, threading, and *ab initio* were developed. Homology modeling is found to be the most accurate method to predict secondary and tertiary structures. Several homology modeling software, which predict all possible three-dimensional arrangement of protein is now available.

Protein structure predicted by Raptor X

Raptor-X server was used to confirm the structure predicted by SWISS-MODEL. Raptor-X server was used to predict the 3-D structure of the protein sequence without close homologs in the Protein Data Bank (PDB) (http://raptorx.uchicago.edu/). The secondary and tertiary structures, contacts, solvent accessibility, disordered regions and binding sites were predicted by the server [30]. To indicate the quality of the protein 3D model, Raptor-X assigns the following confidence scores: p-value, GDT (global distance test) and uGDT (un-normalized GDT), and RMSD for the relative global quality, absolute global quality and absolute local quality of each residue in the protein model.

3D structure predictions by I-TASSER

I-TASSER (Iterative Threading ASSEmbly Refinement) was considered to predict the secondary structure as well as the three-dimensional structure of the protein sequence (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). Structure template was detected from the PDB by fold recognition or threading [31,32]. The server provided details on ligand binding site, gene ontology and enzyme commission by structurally matching the template and the target in protein function databases. I-TASSER is the most accurate method for the prediction of 3D structure and function of model protein, which compares with the sequence similarity of known protein (template) in the Protein Data Bank (PDB) by fold recognition [32].

Functionality of protein predicted by PROFUNC SERVER

PROFUNC is used for the prediction of the three-dimensional structure of a protein model (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/.) It can also predict the secondary structures like α-helix, β-sheet, coils and loops. This software predicts the protein function using the tertiary structure of the protein. It uses both sequence and structure-based methods for the prediction of conserved domains in protein. Sequence method search includes BLAST search against the UniPort knowledgebase, FASTA search against PDB, InterPro Scan, super-family search, and residue conservation mapped onto structure and genome location analysis. Structure-based methods involve fold-matching using MSD Fold and DALI, SSM (secondary structure matching) fold match, Helix-Turn-Helix motif search (HTH), nest analysis, surface cleft analysis, template methods-enzyme active site, ligand binding site, DNA binding site and reverse template search [33]. The server also provides secondary structure and Ramachandran plot for the protein.

SuperPose Version 1.0 server

SuperPose server (http://wishart.biology.ualberta.ca/SuperPose/) is used to superimpose two or more structures to generate the sequence and structure alignments of target proteins. It also predicts the difference distance plots, PDB coordinates, images and RMSD statistics [34].

Molecular graphics programs for protein model visualization

The structures obtained from tertiary predicting software were visualized using Discovery, UCSF Chimera, and Pymol tools to generate better quality images.

Structure analysis and validation

Validation and evaluation of the predicted 3D structure were performed by SAVeS (Structural Analysis and Verification) Server to check for the errors (http://services.mbi.ucla.edu/SAVES/). SAVeS possesses tools such as PROCHECK, WHAT CHECK, ERRAT, VERIFY 3D and PROVE [35-37]. Ramachandran Plot was also generated by the server. The overall quality of the structure was obtained through this server. PROCHECK analyzes the stereochemistry of the protein structure by checking residue-by-residue geometry and overall structure geometry [35]. ERRAT determines the statistics of non-bonded interactions between different atom types. Ramachandran plot visualizes the dihedral angle φ (phi) against ψ (psi) for amino acids in the protein structure. The PDB file of the protein was uploaded, and all the programs mentioned above were selected to run the programs. ProSA web (Protein Structure Analysis) is used to validate the errors in the generated 3D structure of protein that furnish details on the energy plot, z-score and theoretical model (https://prosa.services.came.sbg.ac.at/prosa.php) [38].

Results and Discussion

Enrichment of αs1 casein with LNAA

LNAA supplementation results in the decreased accumulation of Phe in the brain, owing to their competition for the same LAT1 transporter at the blood-brain barrier [5]. Thus, the aim of this study was to design a protein model with the enriched amount of LNAA using homology modeling. A total of 63 different protein models were designed using different template sequence including bovine αs1 casein to enrich LNAA contents (except Phe). Furthermore, a high percentage of β-sheets may cause hindrance to gastrointestinal digestive enzymes that result in low protein value and less protein availability [16]. Therefore, the designed protein should have α-helical structures and coils, without any β-sheets. This was achieved by binary patterning based on the arrangement of polar (P) and nonpolar (N) amino acids for α-helices, which was in the order sequence; PNPPNNPPNPPNNP [15,20].

The FASTA sequence of the best designed protein model (Protein Model-54) enriched with LNAA based on the binary patterning is as follows:

MDLVHLIKHVMRLVQEVLKDMLHRYWTVPQKLVN-VYTELVTELIKILEIVNLITMYSWITYMTLVSILTMWHIVQHLV-EDLMHILMDVLTNYLTYLEQLL.

Construction of protein model by SWISS-MODEL

Designing a protein that would adopt a suitable secondary and tertiary structure, despite high content of LNAA was a challenging task. Earlier attempts of homology modeling using different templates were unsuccessful, since the models had high instability index, presence of β-sheets, flaccid or loose 3D structure, or had similarity to known allergens. Here, bovine αs1 casein (AAA30429.1) was chosen as the template for homology modeling, as it contained a low level of Phe [19]. Model building was performed to enrich LNAA levels other than Phe, in the target using binary patterning for α-helix formations, so that the designed protein contains only this secondary structure in the final model. This is required because the modeled protein is used as a dietary protein, and has to be digested easily in the human gut.

For expression of the designer protein, proper folding of the protein is necessary to avoid aggregation, which in turn forms inclusion bodies [39]. Folding also stabilizes the protein by resisting towards endogenous protease activity. The protein was properly folded into

Citation: Appaiah P, Vasu P (2016) *In silico* Designing of Protein Rich in Large Neutral Amino Acids Using Bovine αs1 Casein for Treatment of Phenylketonuria. J Proteomics Bioinform 9: 287-297. doi: 10.4172/jpb.1000417

a compact structure with four α-helices and devoid of any β-sheets, as depicted in the figure (Figure 1A (a)). The structure was showing sequence similarity to early antigen protein R, since bovine αs1 casein protein structure was not available in PDB. Bovine αs1 casein has a high number of prolines and devoid of disulfide bridges, which gives it a relatively little tertiary structure.

The QMEAN score is a comprehensive scoring function of a linear combination of six structures such as C_beta interaction energy, all-atom pairwise energy, solvation energy, torsion angle energy, secondary structure agreement and solvent accessibility [40]. The QMEAN score of the designed protein (Protein Model-54) is summarized in Table 1. The QMEAN Score obtained was 0.567 (Table 1), which indicated that the model predicted by SWISS-MODEL was highly acceptable because the reliable score should be between 0 and 1.

Sequence similarity analysis by allergenonline.com and ClustalW2

The designed protein sequence was not showing any similarity to the known allergens. The sequence similarity with bovine αs1 casein was found to be 61.4% (Figure 2A), which is expected, since the template used was bovine αs1 casein. High sequence similarity (more than 50%) is good for homology modeling. If the similarity to allergen is more than 50% (with full length) or more than 35% similarity (with 80 amino acid segment), the sequence is considered as a probable allergen and was not considered further in this work [27].

The sequence alignment between bovine αs1 casein and designer protein was performed using ClustalW2 server (Figure 2B). It allows faster alignment of large sets of data and increase alignment accuracy [41]. However, we recently noticed that the ClustalW2 server was retired, and thus the EMBOSS stretcher Pairwise sequence alignment tool was also used to identify the regions of sequence identity between the template and target model (Figure 2C). Both the alignment tools showed different alignment data, because the default parameters used for the alignments were different. In ClastalW2, the matrix used was Gonnet, while that in EMBOSS Stretcher was Eblosum62. The gap open penalty and gap extension penalty in ClastalW2 were 10.0 and 0.2, while those for EMBOSS Stretcher were 12.0 and 2.0. Thus the difference in alignment data is because of these differences in the default alignment parameters of the two server. However the target protein sequence, which contained only 100 amino acids have high sequence homology (61.4%) with the template protein, bovine αs1 casein.

Physicochemical analysis of designed protein

ExPASy ProtParam tool: ExPASy ProtParam tool was used to predict the physicochemical properties of a protein model. The ExPASy ProtParam results are summarized in Table 2. The sequence contained histidine: isoleucine: leucine: methionine: tyrosine: threonine: tryptophan: valine in the ratio 6: 9: 20: 8: 6: 9: 3: 13 respectively, as nutritionally specified for infants by WHO/FAO/UNU [6]. The molecular weight of designed protein (Protein Model-54) containing 100 amino acids was predicted to be 12094.6 Da and theoretical pI was 5.77 which indicate that the protein is negatively charged and that it can be precipitated in acidic medium. Protein pH was calculated based on the pK_a values of individual amino acids, and it depends on the side chains. The extinction co-efficient, which is the quantity of light that could be absorbed by the protein at 280 nm, of Protein Model-54 is 25440 M^{-1} cm^{-1}. This is based on the amount of tyrosine, tryptophan and cysteine residues, which are 6%, 3%, and 0% in the modeled protein. Half-life is the prediction of the time taken by the protein to reduce to half of its amount in the cell after its synthesis in the cell. The server considered human, yeast and *Escherichia coli* cells and it was 30

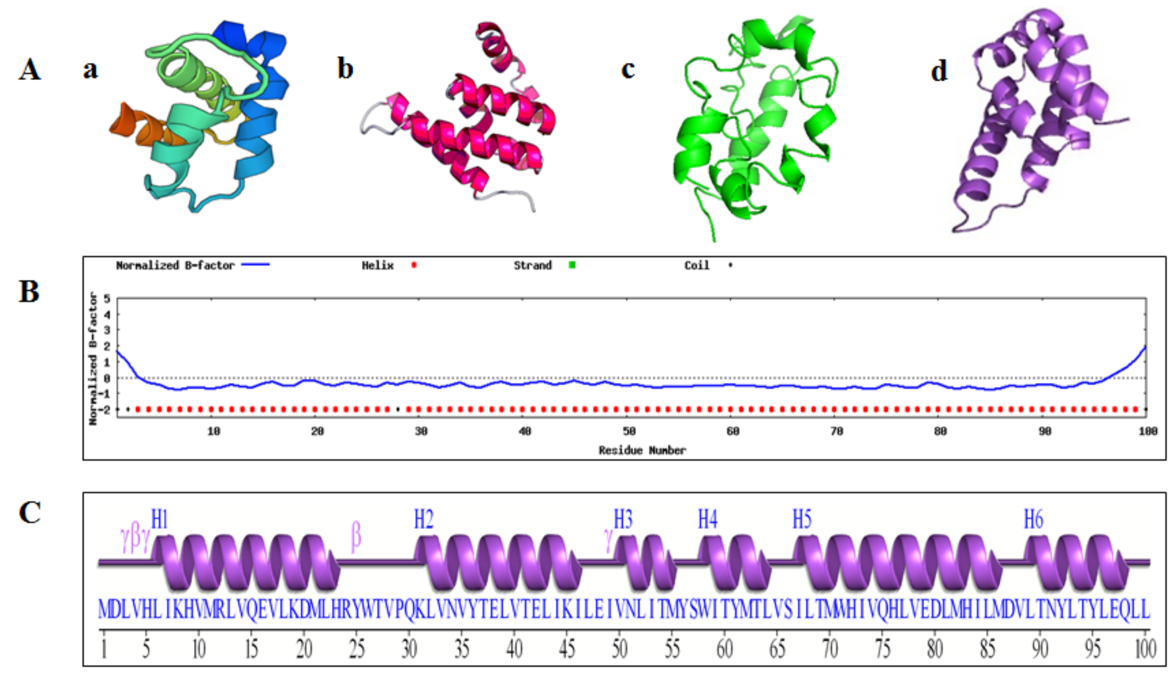

Figure 1: A) Tertiary structure (3D) of the Protein Model-54, generated by: a) Swiss-Model, b) RaptorX, c) I-TASSER, and d) PROFUNC. B) The normalized B-factor profile from I-TASSER. C) Secondary structure prediction by PROFUNC, where α-helices are labeled with 'H', β-sheets are labeled with β and γ.

Parameters	Score
C beta interaction	-68.07 (Z-score: 0.14)
All-atom pairwise energy	-3518.87 (Z-score: 0.10)
Solvation energy	4.26 (Z-score: -2.03)
Torsion angle energy	-4.85 (Z-score: -2.63)
Secondary structure agreement	81.0% (Z-score: -0.22)
Solvent accessibility agreement	69.0% (Z-score: -0.22)
Total QMEAN score	0.567 (Z-score: -1.58)

Table 1: QMEAN score of designed Protein Model-54.

Amino acid	No.	Percentage	Amino acid	No.	Percentage
Ala (A)	0	0.0%	Lys (K)	4	4.0%
Arg (R)	2	2.0%	Met (M)	8	8.0%
Asn (N)	3	3.0%	Phe (F)	0	0.0%
Asp (D)	4	4.0%	Pro (P)	1	1.0%
Cys (C)	0	0.0%	Ser (S)	2	2.0%
Gln (Q)	4	4.0%	Thr (T)	9	9.0%
Glu (E)	6	6.0%	Trp (W)	3	3.0%
Gly (G)	0	0.0%	Tyr (Y)	6	6.0%
His (H)	6	6.0%	Val (V)	13	13.0%
Ile (I)	9	9.0%	Pyl (O)	0	0.0%
Leu (L)	20	20.0%	Sec (U)	0	0.0%
Aliphatic index		150.80	-	-	-
GRAVY		0.630	-	-	-
Instability index		40.50	-	-	-
Molecular weight (Da)		12,094.6	-	-	-
Theoretical pI		5.77	-	-	-
Extinction coefficient (M^{-1} cm^{-1})		25,440	-	-	-

Table 2: Physicochemical properties of designed protein model (Protein model-54).

h, >20 h and >10 h, respectively. As our model protein contained Met as the N-terminal residue, the half-life would be 30 h in mammalian reticulocytes (*in vitro*), >20 h in yeast (*in vivo*) and >10 h in *E. coli* (*in vivo*). N-terminal Arg, Lys, Leu, Phe, Tyr and Trp residues decrease the half-life of the protein in *E. coli* [42].

Stability of the protein was provided by instability index. The designed protein showed instability index of 40.50% whereas αs1 casein was 56.03%. A protein whose instability index less than 40 was predicted as stable and has the half-life of >16 h; those having above 40 was unstable and has a half-life of fewer than 5 h [43]. The instability index is related to the half-life of protein. The sequence contained no regions that are enriched in Pro, Glu, Ser and Thr (PEST) residues as it may result in destabilization of the protein in eukaryotes systems [43].

The aliphatic index is the relative volume of the protein occupied by the aliphatic amino acids like lysine, valine, isoleucine and leucine. The aliphatic index of the Protein Model-54 was 150.80. A high aliphatic index suggests that the protein have a wide range of temperature stability [44]. Higher the aliphatic index greater the thermostability of globular protein [45]. Hydropathicity is the relative hydrophobicity or hydrophilicity of amino acids residues present in the protein sequence. Grand average of hydropathicity (GRAVY) was found to be 0.630 (Table 2). It is calculated by adding hydropathy values of each amino acid residues and dividing by the number of residues in the sequence [46]. The positive value indicated greater hydrophobicity and the protein is sparingly soluble in water.

ExPASy peptide cutter: For the proper metabolism, digestion and absorption of protein in humans, it has to be cleaved or hydrolysed with digestive enzymes like trypsin, chymotrypsin and pepsin to yield peptides and free amino acids. The server shows the probability of the sequence cleaved by many enzymes. Table 3 summarizes the results obtained by the peptide cutter tool. The total number of cleavages for chymotrypsin (high and low specificity), pepsin and trypsin were found to be 9, 42, 36 and 6, respectively (Table 3), which indicated that the digestion rate of the protein is good.

Secondary structure and tertiary structure prediction

Protein 3D structure predicted by Raptor X: The server generated the 3D structure of the given protein sequence along with the PDB format that opened through the Discovery software. The structure had four α-helices that are connected by the coils, and is devoid of β-sheets (Figure 1A.b). The structure contained 95% helix, 0% strands, and 5% coil. The p-value is 1.65e-02, uGDT (GDT) is 45(45), uSeqId (SeqId) is 16(16) and Score is 92. The p-value indicates the quality of the model. Smaller the p-value better is the quality of the model. The score is the alignment score lying between 0 and domain sequence length, where 0 being the worst. This shows that the designed protein is good and acceptable [47].

Protein 3D structure predicted by I-TASSER: I-TASSER is one of the methods of predicting the 3D structure and function of a model protein. The result obtained from the I-TASSER server showed the presence of only α-helical structures with coils, with no β-sheets (Figure 1A (c)). The quality of predicted protein structure was estimated by confidence score (C-score). The server gives 5 best structures in the PDB format, and based on the best or highest C-score value, one of them will be selected. C-score value usually comes in the range of -5 to 2, which is calculated based on the threading alignment. C-score of higher value signifies that the model is predicted with high confidence. The server generated 5 models viz. Model 1, Model 2, Model 3, Model 4 and Model 5 whose C-score values were -3.29, -2.62, -4.52, -4.56 and -3.32 respectively (Table 4). The best model was selected based on C-score value (-3.29) with the estimated accuracy of 0.35 ± 0.12 (TM-score) and RMSD 11.3 ± 4.5Å (Table 4). TM-score (metric to measure the similarity between two proteins) and RMSD are used to measure the structure similarity between two structures which in turn measures the accuracy of structure modeling using the native structure as a reference.

The cluster density is the number of structure decoys (low-temperature replicas) at a unit of space, and a higher cluster density means the structure occurs more often in the simulation trajectory and therefore considered a better quality model. The number of decoys in Table 5 represents the number of structural decoys that are used in generating each model [32]. B-factor is the thermal mobility of amino acids or atoms present in the proteins. The normalized B-factor profile of all the residues was less than 0 (Figure 1B), so they are considered stable, as described by Yang and coworkers [48].

Protein functionality predicted by PROFUNC server: PROFUNC can able to predict the secondary structure, 3D structure and function of protein. PROFUNC server is also used to study the biochemical function of a protein such as conserved regions, active sites, functions of templates, etc. from its 3D structure. The server provided the 3D structure of the designed protein that contained five α-helical structures, packed closely with coils, and is devoid of any β-strands (Figure 1A (d)). The result also showed that the protein sequence contained only α-helices and coils but not β-sheets (Figure 1C). It revealed that the protein contained 77.0% of α-helix, 23% of other structure (coil) and 0.0% β-sheets (Table 5). The overall average G-factor of the designed protein was 0.01. Dihedral angles and main chain covalent forces were taken into consideration under G factors [33]. The functional

Citation: Appaiah P, Vasu P (2016) *In silico* Designing of Protein Rich in Large Neutral Amino Acids Using Bovine αs1 Casein for Treatment of Phenylketonuria. J Proteomics Bioinform 9: 287-297. doi: 10.4172/jpb.1000417

A

```
The best scores are:                                      opt z-sc E(1956) %_id  %_sim  alen
gi|159793197|gid|367|alpha S1 casein [Bos taurus] [Bos  ( 205)   93 148.6  0.0035 0.396 0.614  101
gi|30794348|gid|367|casein alpha s1 [Bos taurus] [Bos t ( 214)   93 148.4  0.0036 0.396 0.614  101
gi|162794|gid|367|alpha-S1-casein [Bos taurus]          ( 214)   80 128.7   0.045 0.386 0.604  101

>>>query, 100 aa vs fasta/version16.fasta library
```

```
>>gi|30794348|gid|367|casein alpha s1 [Bos taurus] [Bos taurus]            (214 aa)
 initn:  80 init1:  47 opt:  93  Z-score: 148.4  bits: 33.4 E(): 0.0036
Smith-Waterman score: 93; 39.6% identity (61.4% similar) in 101 aa overlap (5-100:19-114)

                             10        20        30          40         50        60
query                 MDLVHLIKHVMRLVQEVLKDMLHRYWTVPQKLVNVY-TELVTELIKIL--EIVNLITMYSWITYMT
                          : ::: . : ::::.. : :....:  . .:.  : :.:: : .  : ..  .: . :  :
gi|307 MKLLILTCLVAVALARPKHPIKH-QGLPQEVLNENLLRFFVAP--FPEVFGKEKVNELSKDIGSESTEDQAMED-IKQME
               10        20         30        40          50        60        70

            70        80         90        100
query  LVSILTMWHIVQHLVEDLMHILM-DVLTN-YLTYLEQLL
        :: .  .:: . ::.  ::   :: .. :: ::::::
gi|307 AESISSSEEIVPNSVEQ-KHIQKEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQQKEPMIGVNQ
          80        90        100       110       120       130       140       150
```

B

```
CLUSTAL 2.1 multiple sequence alignment


gi|162794|gb|AAA30429.1|      MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVALFPEVFGKE 50
EMBOSS_001                    MDLVHLIKHVMRLVQEVLKDMLHR---YWTVPQKLVNVYTELVTELIKIL 47
                              *.*: *   *   : .  : : *:     .: ::*:..:. *..*::

gi|162794|gb|AAA30429.1|      KVNELSKDIGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKED 100
EMBOSS_001                    EIVNLITMYS----------WITYMTLVSILTMWHIVQHLVEDLMHILMD 87
                              :: :* .  .          *. *   ** :  .** : **:       *

gi|162794|gb|AAA30429.1|      VPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIDAQQKEPM 150
EMBOSS_001                    VLTN-YLTYLEQLL------------------------------------ 100
                              * :: ** ******

gi|162794|gb|AAA30429.1|      IGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNP 200
EMBOSS_001                    --------------------------------------------------


gi|162794|gb|AAA30429.1|      IGSENSEKTTISLW 214
EMBOSS_001                    --------------
```

C

```
EMBOSS_001          1 MKLLILTCLVAVALARPKHPIKH-QGLPQEVLNENLLRFFVALFPEV-FG     48
                      |.|:                |.||| ..|.||||.:.|.|::......| ..
EMBOSS_001          1 MDLV--------------HLIKHVMRLVQEVLKDMLHRYWTVPQKLVNVY     36

EMBOSS_001         49 KEKVNELSKDIGSESTEDQAMED-IKQMEAESISSSEEIVPNSVEQKHIQ     97
                      .|.|.||.|.:  |......|.. |..|...||.:...||.:.||.....
EMBOSS_001         37 TELVTELIKIL--EIVNLITMYSWITYMTLVSILTMWHIVQHLVEDLMHI     84

EMBOSS_001         98 KEDVPSERYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIDAQQK    147
                      ..||.:. ||.||||||
EMBOSS_001         85 LMDVLTN-YLTYLEQLL---------------------------------    100

EMBOSS_001        148 EPMIGVNQELAYFYPELFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDI    197

EMBOSS_001        101 --------------------------------------------------    100

EMBOSS_001        198 PNPIGSENSEKTTISLW    214

EMBOSS_001        101 -----------------    100
```

Figure 2: Protein sequence showing homology of target to template: A) Protein Model-54 showing homology to template by allergenonline.com, B) CLUSTAL W2.1 multiple sequence alignment of template and target models, and C) Pairwise sequence alignment between template and target models using EMBOSS Stretcher

Name of enzyme	No. of cleavages	Positions of cleavage sites (amino acid residue numbers)
Chymotrypsin-high specificity (C-term to [FYW], not before P)	9	25, 26, 36, 56, 58, 61, 71, 92, 95
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	42	1, 3, 5, 6, 9, 11, 13, 18, 21, 22, 23, 25, 26, 32, 36, 39, 43, 47, 52, 56, 58, 61, 62, 64, 68, 70, 71, 72, 76, 77, 81, 82, 83, 85, 86, 89, 92, 93, 95, 96, 99, 100
Pepsin (pH 1.3)	36	2, 3, 5, 6, 13, 17, 18, 22, 31, 32, 38, 39, 42, 43, 46, 51, 52, 63, 64, 67, 68, 76, 77, 80, 81, 84, 88, 89, 92, 93, 95, 96, 98, 99, 99, 100
Pepsin (pH>2)	50	2, 3, 5, 6, 13, 17, 18, 22, 31, 32, 35, 36, 38, 39, 42, 43, 46, 51, 52, 55, 56, 57, 58, 60, 61, 63, 64, 67, 68, 70, 71, 76, 77, 80, 81, 84, 88, 89, 91, 92, 92, 93, 94, 95, 95, 96, 98, 99, 99, 100
Trypsin	6	8, 12, 19, 24, 31, 45

Table 3: Cleavage of amino residues by trypsin, chymotrypsin and protease enzymes generated by ExPASy peptide cutter.

Name	C-score	Exp. TM-Score	Exp. RMSD	No. of decoys	Cluster density
Model 1	-3.29	0.35 ± 0.12	11.3 ± 4.5	2502	0.0618
Model 2	-2.62	-	-	1440	0.1208
Model 3	-4.52	-	-	574	0.0181
Model 4	-4.56	-	-	571	0.0173
Model 5	-3.32	-	-	1347	0.0597

Table 4: C-score of 5 models from I-TASSER.

prediction of the designer protein was listed in Table 5. The low values suggest that the designed protein have no biochemical or biological function (Table 5).

RMSD by SuperPose server

SuperPose tool was used to generate sequence alignments, structure alignments, PDB (Protein Data Bank) coordinates, RMSD statistics and super imposed molecular images of template and target models [34]. The sequence alignment of template and target models of SuperPose tool (Figure 3A) showed almost similar with that of pairwise sequence alignment (Figure 2C). The similarity of target model for the whole length template (bovine αs1 casein) was low (26.7%), and thus the truncated template was used, which showed a similarity of 42.3%. The overall RMSD values of alpha carbons, back bone and heavy atoms of superposition truncated template and target models was 2.27 Å (Figure 3B), which was better than the full length sequence (RMSD=4.58 Å). The ribbon shaped superposition images of template (red color) and target (yellow color) models were also generated from the SuperPose server (Figure 3C). Based on these results, it can be confirmed that both the truncated template and target protein models have similar structures.

Model visualization programs for protein models

The structure produced from the 3D predicted servers were visualized through Pymol tool, Discovery tool and UCSF Chimera tool (Figures 4a-4c). These software were used for the better visualization of the 3D structures [49,50]. It is very clear from the visualized structures that Protein Model-54 contains only α-helix and coils, and is devoid of any β-sheets.

Structure analysis and validation

The validation and evaluation of the protein structure were performed by SAVEs server, a well-known protein structure validating program. Additionally, protein structure validation was evaluated by ProSA-web server. The structural images (Jmol C^{α} trace) of Protein Model-54 was generated by ProSA-web, where the amino acid residues are colored from blue to red with increasing residue energy (Figure 5A). Red color indicates the high residue energy [38,51].

Two important protein evaluation programs were utilized to check the stereochemistry of the model. ERRAT showed that the overall quality of the Protein Model-54 was 88.04 (Figure 5B). Normally, a score of 50 is acceptable [52]. The overall quality of bovine αs1 casein was found to be 80.68, which is slightly lesser when compared to modeled protein. VERIFY 3D was also used to validate the refined structure. The results revealed that 60% of the residues had an averaged 3D-1D score >0.2 for modeled protein, and 24.73% of the residues had an averaged 3D-1D score >0.2 for bovine αs1 casein template. Modeled 3D structure of nitrogenase iron protein of nitrogen fixing Actinomycete *Arthrobacter* species was evaluated by ERRAT, VERIFY 3D, PROVE and PROCHECK, and had similar scores [53]. ERRAT and Verify 3D confirm the quality of predicted 3D structure [44], and the predicted structure of Protein Model-54 was found to be good, stable, reliable and consistent.

The Ramachandran Plot by PROCHECK was depicted in Figure 5C. Ramachandran plot generated by PROCHECK revealed that 89.9% of the amino acid residues were in the most favored region, 7.2% in additional allowed region, 2.1% in the generously allowed region and 1.0% in the disallowed region (Table 6). The bottom left box indicated the presence of right-handed α-helix. The red region indicated the favored region where no steric clashes occur (Figure 5C). These results showed that the majority of amino acids fall in phi–psi distribution, which is in consistent with right-handed α-helix, as described elsewhere [44]. Ramachandran plot revealed that the designed Protein Model-54 was good, stable, and reliable.

Z-score of the protein model was -4.22 indicated as a large dark dot which was within the range of scores usually found for the native protein of equivalent sizes (Figure 5D). Z-score was determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to the length of protein chains in PDB. In the energy plot, the positive value shows the problematic or erroneous part of the model structure (Figure 5E). The plot of single residue energy (Figure 5E) indicated that the quality of protein is good, and is devoid of error, since the values are negative. ProSA-web tool, ERRAT and PROCHECK were used for Structural evaluation and validation [51,52,53]. This validation result of target protein indicates that the predicted Protein Model-54 is best, reliable, stable, consistent and highly acceptable.

The amino acid sequence of the Protein Model-54 was further modified by removing the initial N–terminal Met, and by adding hexa Histidine tag at the C-terminal end. The protein model was checked with the software mentioned above, and found no difference in the predicted model. The protein sequence will be reverse-translated to DNA sequence, and the synthetic DNA obtained will be cloned and over-expressed in *Pichia pastoris* or *E. coli.* The expressed protein will be used for characterization studies of the protein for its toxicity and *in vitro*- and *in vivo*-digestibility. The dietary effect of this designer protein will be studied in PKU induced rats. The expected outcome of this work may be of high nutritional significance for PKU patients.

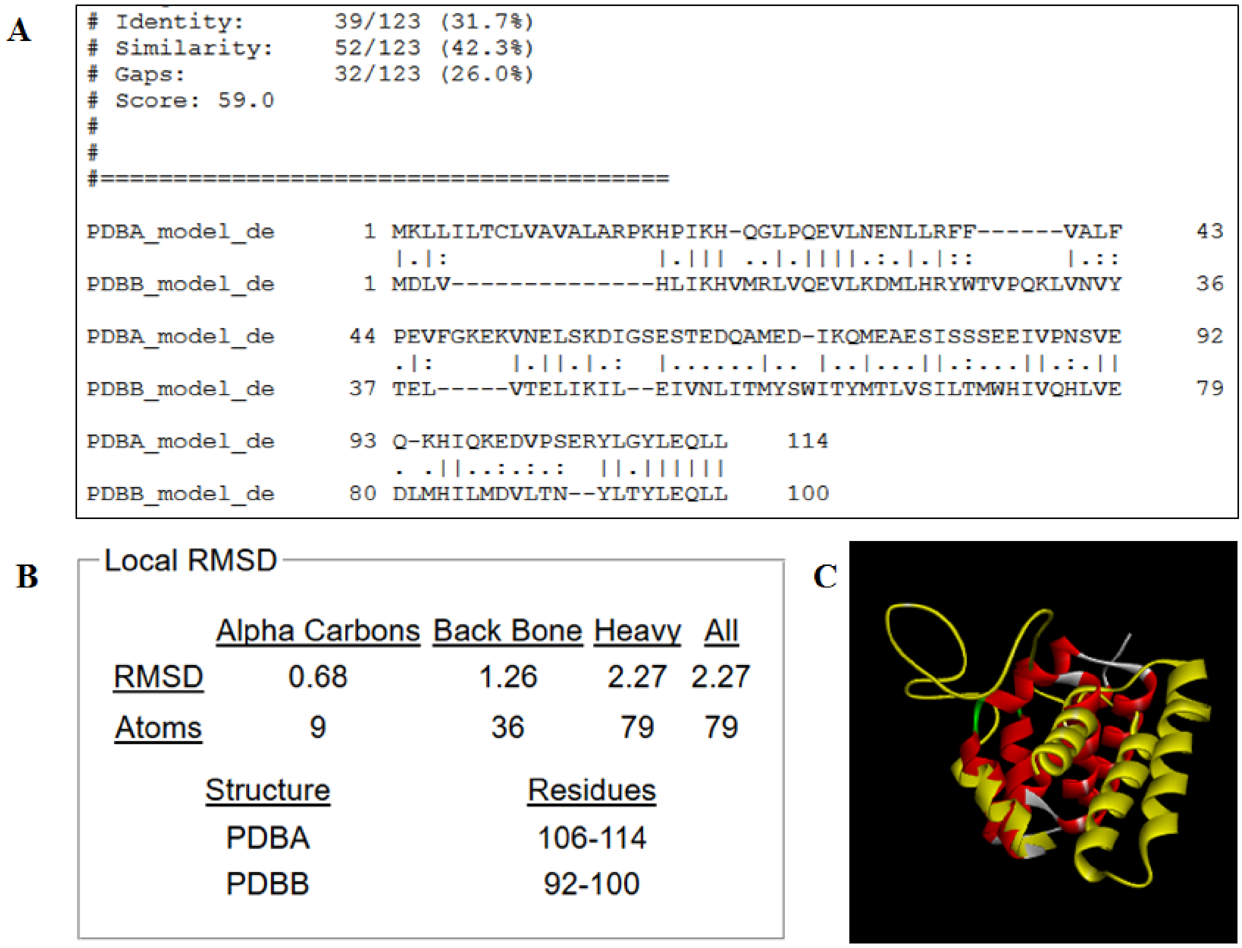

Figure 3: SuperPose Version 1.0. A) Sequence alignment between truncated template and target. B) RMSD. C) Superposition of truncated template and target.

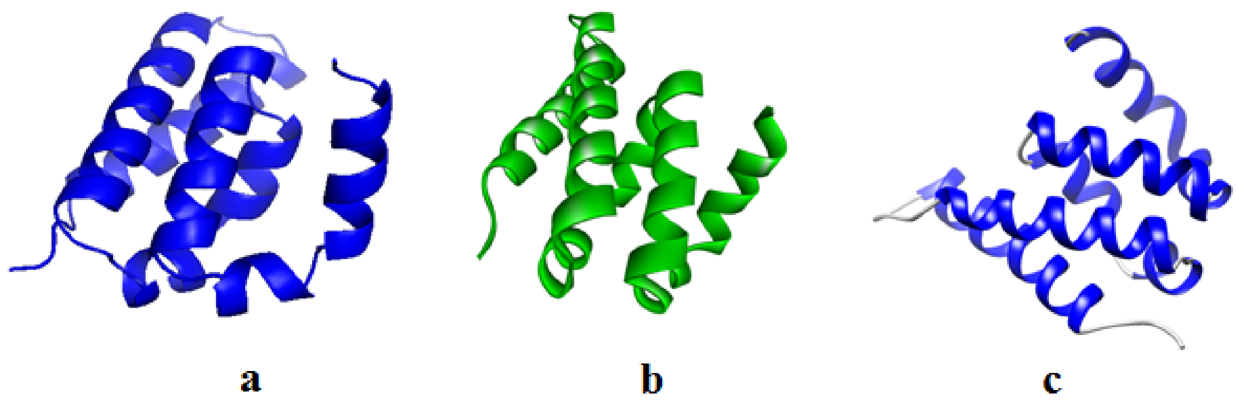

Figure 4: Molecular visualization of 3D structures of target protein from molecular graphics programs: a) PyMOL, b) Discovery studio, and c) UCSF Chimera tool.

Conclusion

Here we have shown that how different bioinformatics tools can be used and applied to design a small protein enriched with LNAA (except Phe). Out of 63 different models designed, Protein Model-54 was selected based on sequence similarity to template (61.4%), compact 3D structure containing only α-helices and coils without β-sheets (QMEAN score of 0.567). The protein sequence of the Protein Model-54 contained histidine: isoleucine: leucine: methionine: tyrosine: threonine: tryptophan: valine in the ratio 6: 9: 20: 8: 6: 9: 3: 13 respectively. Bioinformatics software tools like, SWISS-MODEL, EXPASY tool; PROFUNC, I-TASSER, RaptorX, ProSA-web and

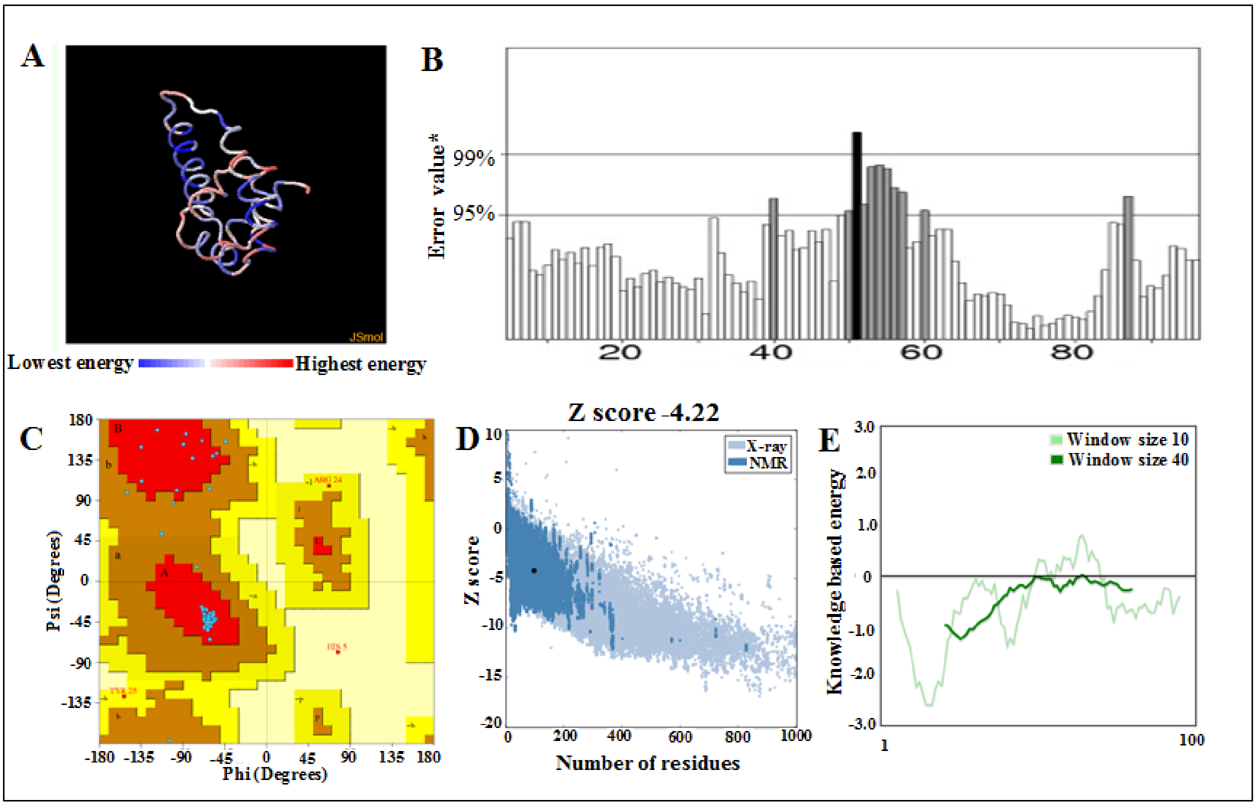

Figure 5: Validation of protein model-54: A) Jmol Cα trace by ProSA web, B) Overall quality by ERRAT, C) Ramachandran plot by PROCHECK, D) Z-score by ProSA web, and E) Energy plot by ProSA web.

Profunc	Protein model-54
Secondary structure summary	
Strand	0 (0.0%)
Alpha helix	77 (77.0%)
3-10 helix	0 (0.0%)
Other	23 (23.0%)
Total residues	100
Cellular component	
Cell	5.73
Cell part	5.73
Intracellular	5.73
Intracellular part	5.73
Biological process	
Cellular process	8.85
Metabolic process	8.2
Primary metabolic process	7.01
Cellular metabolic process	6.78
Biochemical function	
Binding	7.40
Catalytic activity	5.38
Nucleic acid binding	3.87
Metal ion binding	3.30

Table 5: Functional prediction of protein model-54 by Profunc server.

SAVeS server were used for the structure prediction, and validation of the designed protein. Protein Model-54 was designed using bovine αs1 casein as template. The data generated by the secondary structure prediction servers were matching with one another showing that the sequence contained only α-helices and coils, and lacks β-sheets. The

Plot statistics	% of Residues
Residues in most favored region	89.9
Residues in additional allowed region	7.2
Residues in generously allowed/outer region	2.1
Residues in disallowed region	1.0
Number of non-glycine and non proline-residues	97
Number of end residues (excl. Gly and Pro)	2
Number of glycine residues (shown as triangles)	0
Number of proline residues	1
Total number of residues	100

Table 6: Ramachandran plot from SAVeS server (PROCHECK).

sequence was not showing any similarity to known allergens, and the protein was completely hydrolyzed with chymotrypsin, pepsin, and trypsin, indicating good *in silico* digestibility of the protein. The structure was stable as the N-terminal amino acid was Met and the sequence contained no regions that are enriched in Pro, Glu, Ser and Thr residues. The Ramachandran plot showed that 89.9% of the amino acid residues were in the most favored region, 7.2% in additional allowed region, and 2.1% in the generously allowed region, with ERRAT score of 88.04. Ramachandran plot, ERRAT and VERIFY 3D results showed that the designed protein structure was the best, stable, reliable and consistent model. The results clearly indicate that the available bioinformatics tools can be used for homology modeling to design proteins for specific purposes, especially for human health and nutrition.

Acknowledgement

Authors are thankful to Prof. Ram Rajasekharan, Director, CSIR-CFTRI, Mysore, India, for providing infrastructural facilities, and Indian Council of Medical Research (ICMR), Delhi, India, for giving the senior research fellowship award to PA.

Authors' Contributions

PV designed and supervised the work. PA implemented and carried out the experiments. PV & PA analyzed the data. Both the authors wrote, read and approved the final manuscript.

References

1. Rama Devi AR, Naushad SM (2004) Newborn screening in India. Indian J Pediatr 71: 157-160.
2. Reeds PJ (2000) Dispensable and Indispensable amino acids for humans. J Nutr 130: 1835S-1840S.
3. Huttenlocher PR (2000) The neuropathology of phenylketonuria: human and animal studies. Eur J Pediatr 159: S102-S106.
4. Yano S, Moseley K, Azen C (2013) Large neutral amino acid supplementation increases melatonin synthesis in phenylketonuria: a new biomarker. J Pediatr 162: 999-1003.
5. Pietz J, Kreis R, Rupp A, Mayatepek E, Rating D, et al. (1999) Large neutral amino acids block phenylalanine transport into brain tissue in patients with phenylketonuria. J Clin Invest 103: 1169-1178.
6. WHO/FAO/UNU (2007) Protein and amino acid requirements in human nutrition, Report of the Joint WHO/FAO/UNU Expert Consultation. World Health Organization Technical Report Series No. 935, Geneva, Switzerland.
7. Pardridge WM (1998) Blood-brain barrier carrier-mediated transport and brain metabolism of amino acids. Neuro Res 23: 635-644.
8. Farrell HMJ, Flores JR, Bleck GT, Brown EM, Butler JE, et al. (2004) Nomenclature of the proteins of cow's milk-sixth revision. J Dairy Sci 87: 1641-1674.
9. Sanchez D, Kassan M, Contreras MM, Carron R, Recio I, et al. (2011) Long-term intake of a milk casein hydrolysate attenuates the development of hypertension and involves cardiovascular benefits. Pharmacol Res 63: 398-404.
10. Beauregard M, Hefford MA (2006) Enhancement of essential amino acid contents in crops by genetic engineering and protein design. Plant Biotechnol 4: 561-574.
11. Petrey D, Honig B (2005) Protein structure prediction: inroads to biology. Mol Cell 20: 811-819.
12. Marti-Renom MA, Stuart CA, Fiser A, Sanchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Ann Rev Biophy Biomol Struct 29: 291-325.
13. Brindha S, Sailo S, Chhakchhuak L, Kalita P, Gurusubramanian G, et al. (2011) Protein 3D structure determination using homology modeling and structure analysis. In Colloquium 11: 125-133.
14. Sanchez R, Sali A (1997) Advances in comparative protein-structure modeling. Curr Opi Struct Biol 7: 206-214.
15. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein Design by Binary Patterning of Polar and Nonpolar Amino Acids. Science 262: 1680-1685.
16. Yu P (2005) Protein secondary structures (a-helix and b-sheet) at a cellular level and protein fractions in relation to rumen degradation behaviours of protein: a new approach. Br J Nutr 94: 655-665.
17. Yu P, Christensen DA, Christensen CR, Drew MD, Rossnagel BG, et al. (2004) Use of synchrotron FTIR microspectroscopy to identify chemical differences in barley endosperm tissue in relation to rumen degradation characteristics. Can J Anim Sci 84: 523-527.
18. Gorodetskii SI, Zakhar'ev VM, Kyarshulite DR, Kapelinskaya TV, Skryabin KG (1986) Cloning and nucleotide sequence of cDNA for bovine alpha-S1-casein. Biokhimiia 51: 1402-1409.
19. Goda SK, Sharman AF, Yates M, Mann N, Carr N, et al. (2000) Recombinant expression analysis of natural and synthetic bovine alpha-casein in *Escherichia coli*. App Microbiol Biotechnol 54: 671-676.
20. Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) *De novo* proteins from designed combinatorial libraries. Protein Sci 13: 1711-1723.
21. Nelson DL, Cox MM (2008) Lehninger, Principles of Biochemistry (5th edn.) New York.
22. Xiong H, Buckwalter BL, Shieh HM, Hecht MH (1995) Periodicity of polar and non-polar amino acids is the major determinant of secondary structure in self assembling oligomeric peptides. Proc Natl Acad Sci USA 92: 6349-6353.
23. Wang W, Hecht MH (2002) Rationally designed mutations convert *de novo* amyloid-like fibrils into soluble monomeric β-sheet proteins. Proc Natl Acad Sci USA 99: 2760-2765.
24. Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003) Solution structure of a *de novo* protein from a designed combinatorial library. Proc Natl Acad Sci USA 100: 13270-13273.
25. Dongardive J, Abraham S (2013) Predicting 3D structure of proteins from genomic sequences: A genetic algorithm approach. International Conference on Advances in Computing, Communications and Informatics (ICACCI) pp: 1207-1212.
26. Zhang L, Skolnick J (1998) What should the z-score of native protein structures be? Protein Sci 7: 1201-1207.
27. Codex Alimentarius Commission (2003) Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, Italy 30 June-5 July, 2003. Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants and Appendix IV, Annex on the assessment of possible allergenicity pp: 47-60.
28. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends in Genetics 16: 276-277.
29. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook pp: 571-607.
30. Kallberg M, Wang H, Wang S, Peng J, Wang Z, et al. (2012) Template-based protein structure modeling using RaptorX web server. Nat Protoc 7: 1511-1522.
31. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40-48.
32. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 5: 725-738.
33. Laskoswki RA, Watson JD, Thorton JM (2005) ProFunc: a server for predicting protein function and 3D structure. Nucleic Acids Res 33: W89-W93.
34. Maiti, R, Gary HVD, Haiyan Z, David SW (2004) SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 32: 590-594.
35. Laskoswki RA, MacArthur MW, Moss DS, Thorton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystal 26: 283-291.
36. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 2: 1511-1519.
37. Dhurigai N, Daniel RR, Auxilia LR (2014) Structure determination of Leghemoglobin using Homology Modeling. Int J Curr Microbiol App Sci 3: 177-187.
38. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in the three-dimensional structures of proteins. Nucleic Acid Res 35: W407-W410.
39. Rosen R, Biran D, Gur E, Becher D, Hecker M, et al. (2002) Protein aggregation in *Escherichia coli*: Role of proteases. FEMS Microbiology Letter 207: 9-12.
40. Benkert P, Tosatto SCE, Schomburg D (2008) "QMEAN: A comprehensive scoring function for model quality assessment. Proteins, Structure, Function, and Bioinformatics 71: 261-277.
41. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clusal W and Clustal X version 2.0. Bioformatics 23: 2947-2948.
42. Tobias JW, Shrader TE, Rocap G, Varshavsky A (1991) The N-end rule in bacteria. Science 254: 1374-1377.
43. Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. Science 234: 364-368.
44. Gupta S, Kathait A, Sharma V (2015) Computational sequence analysis and structure prediction of jack bean urease. Intl J Adv Res 3: 185-191.
45. Ikai A (1980) Thermostability and Aliphatic Index of Globular Proteins. J Biochem 88: 1895-1898.
46. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-132.

47. Ma J, Wang S, Xu ZFJ (2013) Protein threading using context-specific alignment potential. Bioinformatics 29: 1257-1265.

48. Yang J, Yan R, Roy A, Xu D, Poisson J, et al. (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12: 7-8.

49. Liang MP, Banatao DR, Klien TE, Brutlag DL, Altman RB (2003) Web FEATURE: an interactive web tool for identifying and visualizing functional sites on macromolecular structures. Nucleic Acid Res 31: 3324-3327.

50. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics 7: 339.

51. Tran NT, Jakovlic I, Wang WM (2015) *In silico* characterization, homology modelling and structure-based functional annotation of blunt snout bream (*Megalobrama amblycephala*) Hsp70 and Hsc70 proteins. J Anim Sci Technol 57: 74.

52. Wadood A, Huma, Ullah Z, Riaz M, Shams S, et al. (2014) Structural modeling and molecular dynamics simulation studies of camel milk kappa casein protein. Int J Comput Bioinfo In Silico Model 3: 483-490.

53. Sharon FB, Daniel RR (2013) Homology modeling of nitrogenase iron protein of nitrogen fixing Actinomycete *Arthrobacter* sp. Int J Comput Appl 61: 13-19.