Type: Research Article

Title of the manuscript: Molecular characterization and functional annotation of a hypothetical protein (TDB29877.1) from probiotic bacteria *Lactobacillus acidophilus*: an *in-silico* approach

Md Ataul Goni Rabbani*

Poultry Production Research Division, Bangladesh Livestock Research Institute (BLRI), Savar, Dhaka, Bangladesh

*Corresponding author

E-mail: magrabbani@blri.gov.bd (MAGR)

Short Title: *In-silico* characterization of hypothetical protein *L. acidophilus*

Abstract

Lactobacillus acidophilus bacteria are widely used as probiotic and to produce various healthy fermented food products. The PNW3 strain of the bacteria has numerous proteins in its genome and some are considered as hypothetical proteins. The aim of the present study was to predict the structures and biological functions of the hypothetical protein (accession number: TDB29877.1) from *L. acidophilus* through an *in-silico* approach applying various bioinformatics tools. The sequence similarity was searched on the available biological databases using BLASTp program to find out the homologues proteins. Besides, determination of various properties like physicochemical characteristics, subcellular localization, phylogenetic analysis, functional annotation, protein-protein interaction, determination of secondary and tertiary structures, active site detection and further quality assessment analysis were done using appropriate computational methods of bioinformatics. *In-silico* analysis revealed that the hypothetical protein has contained TerB-N and TerB-C domains with the presence of YjbR-like superfamily. The Protein-protein interaction network analysis revealed that the protein highly interacted with various known and unknown proteins responsible for iron ion binding, DNA and RNA metabolisms and numerous repair mechanisms that maintain cellular integrity. It was also found that the protein has predominantly alpha-helices in its secondary structure and the three dimensional model has been found to be novel as it possessed expected quality while gone through various quality assessment tools. Thus, the present result indicated that the selected hypothetical protein which is cytoplasmic in nature with Belta-grasp fold, plays important role in responding during stress condition or phage defense mechanism.

Key-words:

Lactobacillus acidophilus, stress-response, probiotic, functional annotation, energy minimization, defense mechanism

Introduction

Lactobacillus acidophilus bacteria are widely used as probiotic for commercial use in the livestock sector especially in poultry feed as additive for the improvement of production performance [1–6]. In addition to produce various healthy food products like yogurt, cheese, and other dairy products [7,8]. Anyway, probiotics are the direct fed micro-organisms which when administered in adequate amounts confer a health benefit on the host and by improving its intestinal microbial balance [9–11]. These beneficial properties of *L. acidophilus* bacteria has aroused interest of the researchers across the globe to investigate the function of different proteins involved in defending mechanism.

Advancement in computational biology, development of various bioinformatics tools and analysis servers make it easier to predict functions of the protein, identify sequence similarity, conduct phylogenetic analysis, evaluate active site residue similarity, protein–protein interaction, conserved domains, motif phosphorylation regions and so on [12–17]. In most of the completely sequenced genomes, almost 60% genes have known functions. The number of genes having unknown functions called hypothetical proteins [18] which can be classified as either uncharacterized protein families or domain of unknown functions [19]. Besides, with the advancement in sequencing technologies, the number of sequences deposited in the public biological databases like Swiss-Prot or GenBank have been increasing day by day [20,21] in comparison to experimentally determined structures deposited in the Protein Data Bank (PDB) [22]. This is resulting a gap between the number of known sequences and confirmed functions. The *in-silico* approaches can minimize the gap predicting structures and biological functions of the proteins [23]. The development of various bioinformatics tools come as a boon in this regard [24] which may also help in designing potential drug against pathogenic organisms and confer efficient pharmacological targets [25,26].

The annotation report from NCBI-Genome (<http://www.ncbi.nlm.nih.gov/genome/>) stated that the *Lactobacillus* bacteria have over 50 species. The PNW3 strain of *L. acidophilus* is gram-positive and has a total of 1776 proteins of which 255 proteins are uncharacterized and termed as hypothetical proteins. As the structures and biological functions of these hypothetical proteins are yet to known, molecular characterization and functional annotation of such proteins can lead the relevant researchers to a new dimension of knowledge about the structures, pathways, functions and potential uses in different areas of science. Tremendous development of *in-silico* analysis, numerous bioinformatics tools make it easier to analyze functional annotation of those hypothetical proteins. Thus, the present study was designed to reveal molecular characterization and functional annotation of a hypothetical protein (accession number: TDB29877.1) of the important probiotic bacteria *L. acidophilus* for better understanding of the protein applying various bioinformatics tools.

Materials and methods

Retrieval of hypothetical protein sequence

The sequence information of the hypothetical protein from *L. acidophilus* organism (TDB29877.1) was retrieved from the NCBI-Protein database [8]. Then, the sequence was stored as a FASTA format sequence for further use.

Physicochemical properties analysis

The physicochemical properties of the uncharacterized protein were obtained using the PortParam tool of the ExPaSy server [27]. Various parameters like the molecular weight, amino acid composition, atomic composition, estimated half-life, theoretical pI, extinction coefficient, instability index, aliphatic index, grand average of hydropathicity (GRAVY) etc were analyzed using the tool.

Homology identification

Similarity search for finding homologous proteins from related organisms that might have structural similarities with the selected hypothetical protein was carried out using BLASTp program of NCBI against non-redundant protein sequences and UniProt databases [28–30].

Multiple sequence alignment and phylogeny analysis

Multiple sequence alignment (MSA) was done by MUSCLE using MEGA software [31] between the selected hypothetical protein and other proteins obtained from BLASTp program of NCBI. MSA was also cross-checked by Clustal Omega program of EMBL-EBI [32]. Then, Phylogeny.fr tool [33] was used for the phylogeny analysis of the selected protein sequences.

Subcellular localization analysis

The subcellular localization of the selected hypothetical protein was predicted by using CELLO server [34]. PSORTb [35] and SOSUI tool [36] were also used for the verification of the subcellular localization. In addition, TMHMM, HMMTOP and CCTOP tools [37–39] were also used to cross-check the results.

Functional annotation analysis

For the purpose of functional annotation analysis of the selected hypothetical protein, search carried out at Conserved Domain Database (CDD) of NCBI for conserved domain(s) [40,41]. Motif search were carried out using Motif server [42] and ScanProsite tool of ExPASy server [43]. Pfam [44] and Superfamily [45] database searches were also done to assign the protein's evolutionary relationships. InterProScan [46] was employed for the functional analysis of the protein. For protein folding pattern recognition, PFP-FunD SeqE tool [47] was used. Detection of coiled-coil conformation within the protein was performed using COILS server [48].

Protein-protein interaction analysis

Protein residues interact with each other for their accurate functions. STRING database [49], known for the prediction of protein-protein interactions, search was performed to identify the

possible functional interaction networks of the selected hypothetical protein of *L. acidophilus* bacteria (TDB29877.1).

Secondary structure prediction

The retrieved protein sequence was used for the prediction of the secondary structure of the hypothetical protein. SOMPA server [50,51] was used in this regard. The secondary structure prediction was also further cross-checked and validated by SABLE [52] and PSIPRED servers [53].

Three-dimensional (3D) prediction

The three-dimensional (3D) structure prediction of the hypothetical protein was performed using HHpred server [54,55] of the Max Planck Institute for Development Biology, Tubingen based on pairwise comparison profile of Hidden Markov Models (HMMs). Visualization of the 3D structure obtained from the HHpred server was then done by PyMOL program [56]. SWISS-MODEL interactive workspace [57] was also further used to verify the prediction of 3D structure of hypothetical protein by automated comparative modeling. Later, the 3D structure was further refined through YASARA energy minimization server [58] and YASARA view software.

Quality assessment of the model

Several assessment tools were used for the quality assessment of the predicted 3D structure of the hypothetical protein. PROCHECK, VERIFY3D and ERRAT tools of SAVES server [59–61] were used for the quality assessment of the build model.

Active site detection

Determination of active site of the hypothetical protein was done using Computer Atlas of Surface Topography of Protein (CASTp) server [62,63] which provides an online resource for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins.

Results

Retrieval of hypothetical protein sequence

The selected a hypothetical protein (TDB29877.1) from *L. acidophilus* is a gram positive bacteria. It contains 606 amino acid residues. Additional information collected from the NCBI database regarding this hypothetical protein is given in Table 1.

Table 1. Primary information of the selected hypothetical protein

Protein individualities	Hypothetical protein information
Locus	TDB29877
Definition	hypothetical protein E1P27_08605 [<i>Lactobacillus acidophilus</i>]
Accession number	TDB29877
Version	TDB29877.1
Organism	<i>Lactobacillus acidophilus</i>
Source strain	PNW3
Host	<i>Sus scrofa domesticus</i>
Country and collection time	South Africa: Pretoria; Jun-2012

Physicochemical properties analysis

The PortParam tool of the ExPaSy server was used to retrieve the physicochemical properties of the uncharacterized protein. The most abundant amino acid residue observed was lysine (9.7%), followed by leusine (9.6%), aspartic acid (8.6%), glutamine (7.8%) and isoleucine (7.3%). The lowest number of amino acids were cysteine (0.3%), tryptophan (1.2) and histidine (1.5%). Other physicochemical properties of the protein are given in Table 2.

Table 2. Physicochemical properties of the hypothetical protein

Properties	Value
Molecular Formula	C ₃₂₇₄ H ₅₀₂₄ N ₈₃₈ O ₉₆₁ S ₁₂
Molecular weight	71885.66
Theoretical pI	5.59
Total number of negatively charged residues (Asp+Glu)	99
Total number of positively charged residues (Arg+Lys)	88
Instability index	38.21
Aliphatic index	84.80
Grand average of hydropathicity (GRAVY)	-0.636

Homology identification

BLASTp program was used against non-redundant protein sequences and UniProt databases to find out the homologous proteins with having structural similarities to the selected hypothetical protein. The result of BLASTp program were given in Tables 3 and 4.

Table 3. Similar proteins obtained from non-redundant protein sequences (nr) database

Accession No	Organism	Protein name	Score	Ident. %	E-value
WP_003546145.1	<i>Lactobacillus acidophilus</i>	TerB N-terminal domain-containing protein	1238	100	0
WP_170089063.1	<i>Lactobacillus amylovorus</i>	TerB N-terminal domain-containing protein	801	64.76	0
WP_052542817.1	<i>Lactobacillus sp. OTU4228</i>	TerB N-terminal domain-containing protein	796	64.27	0
WP_202017233.1	<i>Lactobacillus kitasatonis</i>	TerB N-terminal domain-containing protein	790	64.05	0
WP_007125014.1	<i>Lactobacillus ultunensis</i>	TerB N-terminal domain-	774	63.39	0

		containing protein			
WP_098044344.1	<i>unclassified Lactobacillus</i>	MULTISPECIES: TerB N-terminal domain-containing protein	759	61.83	0
WP_060462377.1	<i>Lactobacillus crispatus</i>	TerB N-terminal domain-containing protein	759	59.87	0
WP_204781356.1	<i>Lactobacillus gallinarum</i>	TerB N-terminal domain-containing protein	757	61.79	0

161 **Table 4. Similar proteins obtained from UniProt database**

Accession No	Organism	Protein Name	Score	Ident. %	E-value
A0A4V3BIJ5_9LACO	<i>Lactobacillus crispatus</i>	TerB_N domain-containing protein	1,163	62.40	1.00E-152
A0A0U5K922_LACDE	<i>Lactobacillus delbrueckii</i> subsp. <i>Bulgaricus</i>	TerB-N/TerB-C domain	739	32.30	1.60E-85
A0A0R2D8P4_9LACO	<i>Lactobacillus taiwanensis</i> DSM 21401	TerB_N domain-containing protein	663	34.20	1.30E-76
A0A1G6BEX5_9FIRM	<i>Ruminococcaceae</i> bacterium FB2012	TerB-C domain-containing protein	629	29.60	8.70E-70
A0A1H9EUD7_9FIRM	<i>Butyrivibrio</i> sp. TB	Predicted DNA-binding protein, MmcQ/YjbR family	540	27.00	3.70E-56
A0A315Y7B5_RUMFL	<i>Ruminococcus flavefaciens</i>	TerB-like protein	523	29.40	1.60E-55
A0A3D5MZF8_9FIRM	<i>Erysipelotrichaceae</i> bacterium	TerB_N domain-containing protein (Fragment)	492	27.10	8.20E-51
A0A562SBM9_9SPIO	<i>Treponema putidum</i>	TerB-like protein	475	27.60	2.30E-48
A0A2M8Z4U1_9FIRM	<i>Clostridium</i>	TerB-like protein	433	28.90	2.10E-42

	<i>celerecrescens 18A</i>				

Multiple sequence alignment and phylogeny analysis

Sequences obtained from BLASTp program and the query sequence (TDB29877.1) were aligned by MUSCLE using MEGA software is shown in Fig 1. Multiple sequence alignment was also cross-checked by Clustal Omega program of EMBL-EBI (S1 File). For the confirmation of homology assessment between the proteins, down to the complex and subunit level, phylogenetic analysis was also carried out using Phylogeny.fr server. One click method was applied to construct the phylogenetic tree on the basis of BLASTp result and multiple sequence alignment which given the similar concept about the query protein (Fig 2).

Fig 1. Multiple sequence alignment of different homologous proteins aligned by MUSCLE

Fig 2. Phylogenic tree with bootstrap confidence values of different proteins from *Lactobacillus* organism

Subcellular localization analysis

CELLO server was used to identify the subcellular localization of the selected uncharacterized protein. It was found that it's a cytoplasmic protein. The result obtained from the other servers (PSORTb, SOSUI, TMHMM, HMMTOP and CCTOP) were also revealed the similar indication (Table 5).

Table 5. Subcellular localization of the hypothetical protein

Subcellular localization analysis	Result
CELLO 3.0	Cytoplasmic
PSORTb	Cytoplasmic membrane
SOSUI	Soluble protein
TMHMM 2.0	No transmembrane helices present

HMMTOP	No transmembrane helices present
CCTOP	Not transmembrane protein

Functional annotation analysis

The conserved domain search (CD-search) revealed that (shown in Fig 3) the selected hypothetical protein had two domains, TerB-N terminal domain (accession no: pfam 13208) and TerB-C domain (accession no: cl21414). The TerB-N domain is found N-terminal to TerB, and TerB-C containing proteins. TerB-C occurs in the C terminus of TerB in TerB-N containing proteins. Pfam server predicted the TerB N-terminal domain at 141-360 amino acid residues with an e-value 8.8e-47 and TerB-C domain at 467-599 amino acid residues with an e-value 3.2e-21. Motif and InterProScan servers also forecasted the same domains with at similar alignment position. However, ScanProsite tool of ExPASy server did not find any hit while searching for motif. Superfamily server revealed presence of YjbR-like superfamily. PFP-FunD SeqE tool predicted the fold type of the selected hypothetical protein as Beta-grasp. The x-axis of output graph from COILS server represented the position of the amino acid number in the protein (starting at the N-terminus) and the y-axis showed the coiled coil whereas 'window' refers to the width of the amino acid window that is scanned at one time.

Fig 3. Functional annotation of the hypothetical protein: Fig 3(a) NCBI CD-search result; Fig 3(b) Search result in Pfam server; Fig 3(c) Result of COILS server: coil shows the heptads corresponding to the residue window 14 (green), 21 (blue) and 28 (red)

Protein-protein interaction analysis

STRING database was used to analyze the protein-protein interaction. The result obtained from the STRING server revealed that the query protein interacted with other functionally known and unknown or uncharacterized proteins (Fig 4). The selected hypothetical protein of

L. acidophilus organism (TDB29877.1) showed a high confidence interaction with LBA0469 and LBA0470 protein (same score 0.979) followed by LBA0471 (score 0.845), LBA0466 (score 0.550), LBA0110 (score 0.478), amtB (score 0.464), PspC (score 0.458) and LBA1740 (score 0.418). Of them, there are one COG1201 Lhr-like helicases, one ammonium transport protein, one surface protein PspC and one putative membrane protein. One protein is from Cytochrome P450 71C1 and annotation of three proteins are not available yet.

Fig 4. STRING network analysis of the hypothetical protein, indicates as LBA0468

Secondary structure prediction

Prediction about the secondary structure of the hypothetical protein which includes α -helices, β -sheets, extended strands, turn and coils were obtained from the SOMPA, SABLE and PSIPRED servers (Fig 5). The result of predicted secondary structure of the hypothetical protein from SOMPA server showed that alpha-helices were most predominant (50%) followed by random coil (36.8%), extended strand (10.73%) and beta-turn (2.48%). Similar type of outputs were obtained while validating the secondary structure using SABLE and PSIPRED tools.

Fig 5. Secondary structure of hypothetical protein predicted by-Fig 5(a) SOMPA server (The window width, similarity threshold and number of states were 17, 8 and 4 respectively); Fig 5(b) SABLE server; Fig 5(c) PSIPRED server

Three-dimensional (3D) prediction

Prediction of the three-dimensional (3D) structure of hypothetical protein was done by using HHpred server. This server predicted 3D structure of the protein (Fig 6) having 99.24% identity with the highest scoring template (PDB ID: 3H9X_A). 3H9X_A is a crystal structure of the PSPTO_3016 protein from *Pseudomonas syringae* organism with four chains (Chain A, B, C and D). Further validation of the 3D structure prediction by SWISS-MODEL interactive workspace revealed that the oligo-state of the protein is a monomer [64]. The

crystallographic resolution of the template used to the model protein was 2.51Å by adopting the X-ray diffraction method [65]. Global quality estimate, local quality estimate, comparison of protein size residue and model template alignment were also explored from this server (Fig 7). Later, the 3D structure was further modified by YASARA energy minimization server. The energy calculated before energy minimization was -10988.9 kJ/mol whereas it was reduced to -55991.2 kJ/mol after energy minimization. The initial score was -2.99 while the final score was -0.19 after energy minimization.

Fig 6. Predicted 3D structure of the hypothetical protein

Fig 7. The assessment of 3D structure using SWISS-MODEL interactive workspace: Fig 7(a) Global quality estimate; Fig 7(b) local quality estimate; Fig 7(c) comparison of the protein size residue

Quality assessment of the model

Validation of 3D structure of the hypothetical protein was done through several quality assessment steps. Assessment of the 3D model was done by PROCHECK tool through Ramachandran plot analysis (Fig 8), where the distribution of ϕ and ψ angle in the model within the limits were shown. This result also showed that residues in the most favored regions covered 92.4% (Table 6). Then, the structure again verified by VERIFY 3D and ERRAT tools and found 90.65% of the residues had average 3D-1D score ≥ 0.2 and overall quality factor was 72.72 respectively.

Fig 8. Ramachandran plot for the 3D model of the hypothetical protein validated by PROCHECK program

Table 6. Ramachandran plot statistics of the hypothetical protein

Residues in the most favored resins [A,B,L]	85	92.4%
Residues in the additional allowed resins [a,b,l,p]	4	4.3%

Residues in the generously allowed regions [$\sim$ a, $\sim$ b, $\sim$ l, $\sim$ p]	2	2.2%
Residues in the disallowed regions	1	1.1%
	----	-----
Number of non-glycine and non-proline residues	92	100%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown in triangles)	7	
Number of proline residues	6	

Total number of residues	107	

Active site detection

Computer Atlas of Surface Topography of Protein (CASTp) server was used to determination the active sites with the amino acid residues of the hypothetical protein (Fig 9). The result from CASTp calculation revealed a total of 18 active pockets of the hypothetical protein. The best active site found in the areas (SA) with 126.75 and a volume (SA) of 78.13 amino acids.

Fig 9. Active site detection of the hypothetical protein using CASTp: Fig 9(a) The red sphere indicates the active site of the protein; Fig 9(b) Sequence of active amino acid residues of the largest pocket

Discussion

Since the *L. acidophilus* bacteria are well known for its beneficial properties and are being used as probiotics, a hypothetical protein from this organism have targeted to examine it's involvement in the defensive mechanism against pathogenic organisms. Thus, the amino acid sequence of the targeted hypothetical protein was retrieved for further investigation. The selected hypothetical protein had a total of 606 amino acids. The computed instability index (38.21) classifying the protein as stable one because instability index value below 40

indicates a protein as stable and above 40 indicates as unstable [66]. The selected protein's aliphatic index (94.34) indicated its stability over a wide range of temperature and the negative GRAVY value (-0.636) indicated the hydrophilicity nature of the hypothetical protein [67].

Homology analysis revealed that the query protein has structural similarities with other TerB-N and TerB-C domain containing proteins from various *Lactobacillus* species. Multiple sequence alignment using MUSCLE and Clustal Omega program produced alignments between the selected sequences from BLASTp using seeded guide trees and HMM profile-profile techniques. The phylogenetic tree displayed the highest degree of similarity between the studied hypothetical protein and its related proteins for homology modeling obtained by BLASTp of NCBI. The bootstrapping confidence levels of the analysis stated the closed similarity between the query protein (TDB29877.1) and TerB N-terminal domain-containing protein (WP_003546145.1) of *L. acidophilus* organism. Other homologous proteins from various *Lactobacillus* species formed separate clades having varied structural similarity with the studied hypothetical protein.

Basically, subcellular localization of protein indicates where the protein resides in a cell; it may be in outer membrane, inner membrane, periplasm, extracellular or in cytoplasm [68]. The functional properties, interaction and genome annotation are highly influenced by its subcellular localization. Results from CELLO and PSORTb servers indicated the hypothetical protein as a cytoplasmic one. SOSUI server was also depicted the selected protein as a soluble protein. Absent of transmembrane helices predicted by TMHMM and HMMTOP also emphasized the result of being cytoplasmic protein. In addition, CCTOP server also summarized that the query protein was not a transmembrane protein, thus it's a cytoplasmic protein. Such subcellular identification analysis indicated that the hypothetical protein might be involved in recovering disease state through discovering some novel drugs

[69]. As the membrane protein can be used as a potential vaccine target and the cytoplasmic proteins may act as promising drug targets [70], the selected hypothetical protein may be a good source for producing various beneficial pharmaceutical products or healthy food items. The response against chemical stress and anti-viral defense systems of bacteria are constituted by the Ter gene products. The TerB_N has a predominantly alpha-helical structure and contains an absolutely conserved glutamate. The presence of a conserved acidic residue suggested that it might chelate metal like TerB. These proteins occur in a two-gene operon containing an AAA+ ATPase and SF-II DNA helicase suggesting a role in stress-response or phage defense [71]. TerB-C domain also displays multiple conserved acidic residues. The presence of conserved acidic residues in both TerB-N and TerB-C suggested that they, like the TerB domain, might also chelate metals. These two domains might also occur together in the same protein independently of TerB [71]. Motif, Pfam and InterProScan servers also confirmed the presence of TerB-N and TerB-C domains in the selected hypothetical protein. YjbR-like superfamily of the hypothetical protein is expected to contain the DNA binding domain comprising the 'double wing' motif [72]. Moreover, protein fold plays a significant role in their function and hence the fold prediction has also been applied in order to further validate the predicted function. PFP-FunD SeqE tool revealed the fold type of the hypothetical protein as 'Belta-grasp' which indicated that the protein might play role in hydrolase activity [73].

The function of a target protein and drug availability of molecules can be predicted by analyzing protein-protein interaction [49]. Protein-protein interaction network analysis showed that the query protein (TDB29877.1, shown as LBA0468 in Fig 4) highly interacted with proteins LBA0469, LBA0470 and LBA0471 within the network; they are also neighborhood in the genome. These interactions give an indication about the selected

314 hypothetical protein that it might be involved in iron ion binding, DNA and RNA
315 metabolisms and numerous repair mechanisms that maintain cellular integrity [29,74].
316 It was obtained from SOMPA server prediction that the selected secondary structure of
317 hypothetical protein was an alpha-helices dominating protein. The window width, similarity
318 threshold and number of states were 17, 8 and 4 respectively. Confidence of prediction from
319 PSIPRED server also stated alpha-helices dominating output. In addition, SABLE server
320 forecasted the secondary structure of the protein having a good confidence of prediction.
321 The HHpred server forecasted a 3H9X_A protein template with highest score (106.27). 3H9X
322 belongs to the protein Pspto_3016 of *Pseudomonas syringae*. Pspto_3016 is a 117-residue
323 member of the protein domain family PF04237, which is to date a functionally
324 uncharacterized family of proteins [75]. The GMQE (Global Model Quality Estimation) and
325 QMEAN value of the selected model from the SWISS-MODEL interactive workspace
326 analysis were 0.06 and -1.69 respectively [57]. GMQE is a quality estimation which
327 combines properties from the target-template alignment and the template structure and the
328 resulting GMQE score is expressed as a number between 0 and 1. The QMEAN Z-score
329 provides an estimate of the ‘degree of nativeness’ of the structural features observed in the
330 model on a global scale [76,77] and scores of -4.0 or below are an indication of models with
331 low quality. The GMQE and QMEAN score of the selected model indicated it as a
332 comparatively reliable and better quality model. The comparison plot and model quality
333 scores of individual models are related to scores obtained for experimental structures of
334 similar size. The x-axis shows protein length (number of residues). The y-axis is the
335 normalized QMEAN score. Every dot represents one experimental protein structure. Black
336 dots are experimental structures with a normalized QMEAN score within 1 standard
337 deviation of the mean ($|Z\text{-score}|$ between 0 and 1), experimental structures with a $|Z\text{-score}|$
338 between 1 and 2 are grey. Experimental structure that are even further from the mean are

light grey [76,77]. The actual model is represented as a red star meant that our model was within the grey region. In addition, reduced energy and improved score of the predicted model applying YASARA energy minimization tools indicated the model structure as more stable one [58].

Ramachandran plot analysis showed that 92.4% of the residues belonged to the most favored regions. Residues in additional allowed regions and generously allowed regions were 4.3% and 2.2% respectively, which indicated reliability of the model quality. It is generally accepted that more than 90% of the residues in the most favored regions is likely to be a reliable 3D model [78]. The environmental profile or the amino acid environment for non-bonded atomic interactions of the model is good as VERIFY 3D analysis revealed that 90.65% of the residues had average 3D-1D score ≥ 0.2 . Overall quality factor obtained through ERRAT was 72.72 which indicated a high quality model. Higher scores indicate higher quality and the generally accepted range is >50 for a high quality model [79].

A probe radius of 1.4Å was used for computing solvent accessible surface area while calculating the active sites of the hypothetical protein using CASTp server. It also measured the exact volumes and areas, as well as sizes of the mouth openings of the active pockets. These metrics were calculated analytically, using both the solvent accessible surface model called Lee and Richards' surface model [80] and the molecular surface model called Connolly's surface model [81].

Conclusion

The current study was designed to forecast the structures and biological functions of a hypothetical protein (TDB29877.1) from *L. acidophilus* bacteria through an *in-silico* approach. All the above findings applying various bioinformatics tools suggested that the selected hypothetical protein from probiotic type bacteria plays role in responding during stress condition or phage defense mechanism. It was also found that the hypothetical protein

of interest is cytoplasmic in nature containing ‘Belta-grasp’ fold. These findings may encourage researchers who are interested to work with such beneficial probiotic bacteria to produce various feed additives or healthy food products. Therefore, the outcome of this study in determining structures and functions of the uncharacterized protein indicate reliability of computational approach using bioinformatics tools, thereby assisting experimental validation research on a protein.

Acknowledgement

The author is grateful to Mr Mohammad Uzzal Hossain of Bioinformatics Division at National Institute of Biotechnology, Bangladesh for providing extraordinary mental support, guidelines and courage to work on such topic.

References

1. Gallazzi D, Giardini A, Mangiagalli MG, Marelli S, Ferrazzi V, Orsi C, et al. Effects of *Lactobacillus acidophilus* D2/CSL on laying hen performance. *Ital J Anim Sci.* 2008;7: 27–37. doi:10.4081/ijas.2008.27
2. Salarmoini M, Fooladi MH. Efficacy of *Lactobacillus acidophilus* as probiotic to improve broiler chicks performance. *J Agric Sci Technol.* 2011;13: 165–172.
3. De Cesare A, Sirri F, Manfreda G, Moniaci P, Giardini A, Zampiga M, et al. Effect of dietary supplementation with *Lactobacillus acidophilus* D2/CSL (CECT 4529) on caecum microbioma and productive performance in broiler chickens. *PLoS One.* 2017;12. doi:10.1371/journal.pone.0176309
4. Balevi T, An USU, Coskun B, Kurtoglu V, Etingül IS. Effect of dietary probiotic on performance and humoral immune response. *Br Poult Sci.* 2001;42: 456–461. doi:10.1080/00071660120073133
5. Kurtoglu V, Kurtoglu F, Seker E, Coskun B, Balevi T, Polat ES. Effect of probiotic

supplementation on laying hen diets on yield performance and serum and egg yolk
cholesterol. Food Addit Contam. 2004;21: 817–823.
doi:10.1080/02652030310001639530

6. Ljungh Å, Wadström T. Lactic acid bacteria as probiotics. Curr Issues Intest
Microbiol. 2001;7: 73–90.

7. Ershidat OTM, Mazahreh AS. Probiotics bacteria in fermented dairy products.
Pakistan J Nutr. 2009;8: 1107–1113. doi:10.3923/pjn.2009.1107.1113

8. NCBI. Lactobacillus acidophilus (ID 1099) - Genome - NCBI. [cited 22 May 2021].
Available: <https://www.ncbi.nlm.nih.gov/genome/?term=Lactobacillus+acidophilus>

9. AFRC RF. Probiotics in man and animals. Journal of Applied Bacteriology. John
Wiley & Sons, Ltd; 1989. pp. 365–378. doi:10.1111/j.1365-2672.1989.tb05105.x

10. Miles RD, Bootwalla SM. Direct-fed microbials in animal production. A Rev. 1991;
117–132.

11. FAO. Report of a joint FAO/WHO expert consultation on evaluation of health and
nutritional properties of probiotics in food including powder milk with live lactic acid
bacteria. Amerian Córdoba Park Hotel, Córdoba, Argentina; 2001.

12. Díaz DA, Barreto GE, Santos JG. Structural and functional prediction of the
hypothetical protein PA2481 in Pseudomonas aeruginosa Pao1. Advances in
Intelligent Systems and Computing. Springer Verlag; 2014. pp. 47–55.
doi:10.1007/978-3-319-01568-2_7

13. Canduri F, Fadel V, Basso LA, Palma MS, Santos DS, De Azevedo WF. New catalytic
mechanism for human purine nucleoside phosphorylase. Biochem Biophys Res
Commun. 2005;327: 646–649. doi:10.1016/j.bbrc.2004.12.052

14. Henrique Pereira J, Canduri F, Sim J, de Oliveira oes, Jos Freitas da Silveira N,
Augusto Basso L, et al. Structural bioinformatics study of EPSP synthase from

- 413   Mycobacterium tuberculosis. doi:10.1016/j.bbrc.2003.10.175
- 414   15.   Canduri F, Cardoso Perez P, Caceres R, de Azevedo W. Protein kinases as targets for
415   antiparasitic chemotherapy drugs. Curr Drug Targets. 2007;8: 389–398.
416   doi:10.2174/138945007780058979
- 417   16.   Gong J, Chen Y, Pu F, Sun P, He F, Zhang L, et al. Understanding membrane protein
418   drug targets in computational perspective. Curr Drug Targets. 2019;20: 551–564.
419   doi:10.2174/1389450120666181204164721
- 420   17.   Tan J-X, Lv H, Wang F, Dao F-Y, Chen W, Ding H. A survey for predicting enzyme
421   family classes using machine learning methods. Curr Drug Targets. 2019;20: 540–550.
422   doi:10.2174/1389450119666181002143355
- 423   18.   Naveed M, Tehreem S, Usman M, Chaudhry Z, Abbas G. Structural and functional
424   annotation of hypothetical proteins of human adenovirus: prioritizing the novel drug
425   targets. BMC Res Notes. 2017;10: 706. doi:10.1186/s13104-017-2992-z
- 426   19.   Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, Deacon AM, et al.
427   Exploration of uncharted regions of the protein universe. PLoS Biol. 2009;7:
428   e1000205. doi:10.1371/journal.pbio.1000205
- 429   20.   Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al.
430   The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.
431   Nucleic Acids Research. Nucleic Acids Res; 2003. pp. 365–370.
432   doi:10.1093/nar/gkg095
- 433   21.   Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL.
434   GenBank. Nucleic Acids Res. 2002;30: 17–20. doi:10.1093/nar/30.1.17
- 435   22.   Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The
436   Protein Data Bank. Nucleic Acids Research. Oxford University Press; 2000. pp. 235–
437   242. doi:10.1093/nar/28.1.235

23. Baker D, Sali A. Protein structure prediction and structural genomics. *Science*. Science; 2001. pp. 93–96. doi:10.1126/science.1065659
24. Doerks T, von Mering C, Bork P. Functional clues for hypothetical proteins based on genomic context analysis in prokaryotes. *Nucleic Acids Res*. 2004;32: 6321–6326. doi:10.1093/nar/gkh973
25. F. de Azevedo W. Molecular Dynamics Simulations of Protein Targets Identified in *Mycobacterium tuberculosis*. *Curr Med Chem*. 2011;18: 1353–1366. doi:10.2174/092986711795029519
26. De Azevedo WF, Canduri F, Fadel V, Teodoro LGVL, Hial V, Gomes RAS. Molecular model for the binary complex of uropepsin and pepstatin. *Biochem Biophys Res Commun*. 2001;287: 277–281. doi:10.1006/bbrc.2001.5555
27. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*. Humana Press; 2005. pp. 571–607. doi:10.1385/1-59259-890-0:571
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
29. Bateman A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47: D506–D515. doi:10.1093/nar/gky1049
30. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49: D480–D489. doi:10.1093/nar/gkaa1100
31. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Battistuzzi FU, editor. *Mol Biol Evol*. 2021 [cited 22 May 2021]. doi:10.1093/molbev/msab120
32. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI

- search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47: W636–W641. doi:10.1093/nar/gkz268
33. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 2008;36: W465–W469. doi:10.1093/nar/gkn180
34. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins Struct Funct Genet.* 2006;64: 643–651. doi:10.1002/prot.21018
35. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 2010;26: 1608–1615. doi:10.1093/bioinformatics/btq249
36. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics.* 1998;14: 378–379. doi:10.1093/bioinformatics/14.4.378
37. Möller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics.* 2001;17: 646–653. doi:10.1093/bioinformatics/17.7.646
38. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001;17: 849–850. doi:10.1093/bioinformatics/17.9.849
39. Dobson L, Remenyi I, Tusnady GE. CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* 2015;43: W408–W412. doi:10.1093/nar/gkv451
40. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* 2020;48: D265–D268. doi:10.1093/nar/gkz991

41. Marchler-Bauer A, Bryant SH. CD-Search: Protein domain annotations on the fly. Nucleic Acids Res. 2004;32: W327–W331. doi:10.1093/nar/gkh454
42. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res. 2002;30: 42–46. doi:10.1093/nar/30.1.42
43. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006;34: W362–W365. doi:10.1093/nar/gkl124
44. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Nucleic Acids Res. 2014;42: D222–D230. doi:10.1093/nar/gkt1223
45. Wilson D, Madera M, Vogel C, Chothia C, Gough J. The SUPERFAMILY database in 2007: Families and functions. Nucleic Acids Res. 2007;35: D308–D313. doi:10.1093/nar/gkl910
46. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. Nucleic Acids Res. 2009;37: D211–D215. doi:10.1093/nar/gkn785
47. Shen H Bin, Chou KC. Predicting protein fold pattern with functional domain and sequential evolution information. J Theor Biol. 2009;256: 441–446. doi:10.1016/j.jtbi.2008.10.007
48. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science (80-). 1991;252: 1162–1164. doi:10.1126/science.252.5009.1162
49. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43: D447–D452. doi:10.1093/nar/gku1003

50. Geourjon C, Deléage G. SOMPA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. 1995;11: 681–684. doi:10.1093/bioinformatics/11.6.681
51. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: Network Protein Sequence analysis. *Trends Biochem Sci*. 2000;25: 147–150. doi:10.1016/S0968-0004(99)01540-6
52. Adamczak R, Porollo A, Meller J. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins Struct Funct Genet*. 2005;59: 467–475. doi:10.1002/prot.20441
53. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000;16: 404–405. doi:10.1093/bioinformatics/16.4.404
54. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. *J Mol Biol*. 2018;430: 2237–2243. doi:10.1016/j.jmb.2017.12.007
55. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, et al. Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinforma*. 2020;72: e108. doi:10.1002/cpbi.108
56. PyMOL. The PyMOL Molecular Graphics System. Schrödinger, LLC; Available: <https://pymol.org/>
57. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46: W296–W303. doi:10.1093/nar/gky427
58. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins: Structure, Function and*

- Bioinformatics. Proteins; 2009. pp. 114–122. doi:10.1002/prot.22570
59. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26: 283–291. doi:10.1107/s0021889892009944
60. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277: 396–404. doi:10.1016/S0076-6879(97)77022-8
61. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993;2: 1511–1519. doi:10.1002/pro.5560020916
62. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Res. 2018;46: W363–W367. doi:10.1093/nar/gky473
63. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006;34. doi:10.1093/nar/gkl282
64. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep. 2017;7: 1–15. doi:10.1038/s41598-017-09654-8
65. Hisano T, Tsuge T, Fukui T, Iwata T, Miki K, Doi Y. Crystal structure of the (R)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis. J Biol Chem. 2003;278: 617–624. doi:10.1074/jbc.M205484200
66. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng Des Sel. 1990;4: 155–161.

doi:10.1093/protein/4.2.155

67. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157: 105–132. doi:10.1016/0022-2836(82)90515-0
68. Huang GH, Xu WB. Recent advance in new types of human adenovirus. *Chinese J Virol.* 2013;29: 342–348. Available: <https://europepmc.org/article/med/23905481>
69. Imai K, Asakawa N, Tsuji T, Akazawa F, Ino A, Sonoyama M, et al. SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. *Bioinformation.* 2008;2: 417–421. doi:10.6026/97320630002417
70. Vahisalu T, Kollist H, Wang YF, Nishimura N, Chan WY, Valerio G, et al. SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. *Nature.* 2008;452: 487–491. doi:10.1038/nature06608
71. Anantharaman V, Iyer LM, Aravind L. Ter-dependent stress response systems: Novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. *Mol Biosyst.* 2012;8: 3142–3165. doi:10.1039/c2mb25239b
72. Singarapu KK, Liu G, Xiao R, Bertonati C, Honig B, Montelione GT, et al. NMR structure of protein yjbR from *Escherichia coli* reveals “double-wing” DNA binding motif. *Proteins Struct Funct Genet.* 2007;67: 501–504. doi:10.1002/prot.21297
73. Kuhn P, Tarentino AL, Plummer TH, Van Roey P. Crystal Structure of Peptide-N4-(N-acetyl-β-d-glucosaminyl)asparagine Amidase F at 2.2-Å Resolution. *Biochemistry.* 1994;33: 11699–11706. doi:10.1021/bi00205a005
74. Hajj M, Langendijk-Genevaux P, Batista M, Quentin Y, Laurent S, Abdel Raz-zak Z, et al. Phylogenetic diversity of Lhr proteins and biochemical activities of the *Thermococcales* aLhr2 DNA/RNA helicase. 2021 [cited 23 May 2021]. doi:10.20944/preprints202103.0477.v1
75. Feldmann EA, Seetharaman J, Ramelot TA, Lew S, Zhao L, Hamilton K, et al.

- 588   Solution NMR and X-ray crystal structures of *Pseudomonas syringae* Pspto-3016 from
- 589   protein domain family PF04237 (DUF419) adopt a “double wing” DNA binding motif.
- 590   *J Struct Funct Genomics*. 2012;13: 155–162. doi:10.1007/s10969-012-9140-8
- 591   76.   Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of
- 592   individual protein structure models. *Bioinformatics*. 2011;27: 343–350.
- 593   doi:10.1093/bioinformatics/btq662
- 594   77.   Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T.
- 595   QMEANDisCo—distance constraints applied on model quality estimation.
- 596   *Bioinformatics*. 2020;36: 1765–1771. doi:10.1093/bioinformatics/btz828
- 597   78.   Hooda V, Gundala P babu, Chinthala P. Sequence analysis and homology modeling of
- 598   peroxidase from *Medicago sativa*. *Bioinformation*. 2012;8: 974–979.
- 599   doi:10.6026/97320630008974
- 600   79.   Messaoudi A, Belguith H, Hamida J Ben. Three-dimensional structure of *Arabidopsis*
- 601   *thaliana* lipase predicted by homology modeling method. *Evol Bioinforma*. 2011;2011:
- 602   99–105. doi:10.4137/EBO.S7122
- 603   80.   Lee B, Richards FM. The interpretation of protein structures: Estimation of static
- 604   accessibility. *J Mol Biol*. 1971;55: 379–400. doi:10.1016/0022-2836(71)90324-X
- 605   81.   Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. *Science*.
- 606   *Science*; 1983. pp. 709–713. doi:10.1126/science.6879170

Supporting information

S1 File. Multiple sequence alignment of different homologous proteins aligned by Clustal Omega program of EMBL-EBI

Fig2

Protein Classification

TerB_N and TerB_C domain-containing protein (domain architecture ID 10590934)

TerB_N and TerB_C domain-containing protein

Graphical summary

☐ Zoom to residue level

show extra options »

Pssm-ID: 404157 Cd Length: 203 Bit Score: 160.59 E-value: 8.42e-46

```

      10      20      30      40      50      60      70      80
      .....|.....|.....|.....|.....|.....|.....|.....|
seqsig_MDPNQ_c8cd862d30a647573b0348f5f8ac9673 142 NFKKKPLATDRIHQMQE-LYDYSvlpsvGRYKNFYNQKFMADyeDDYSDYAAFRRFYPTYHDMKIEQLRSYFAWRTQLR 220
Cdd:pfam13208      2 VKVAGYSIPGGLVYVGTsLDGYR----GREEAFIDPSLSVAS--DGSDYFGPFMSYWP SYHSLSPAQRRA YLDWLAQGR 74

      90     100     110     120     130     140     150     160
      .....|.....|.....|.....|.....|.....|.....|.....|
seqsig_MDPNQ_c8cd862d30a647573b0348f5f8ac9673 221 KGnyQKVSTSYAYVYIYELLNNIGVKSADGYQKLIDFKENYVEkfdlsIEPYLNDWLKDYvLPYELDQKLIENfeDEI 300
Cdd:pfam13208      75 KG--PDTDLGYVFLYFYGLERRVGVDNPQEGYVKLIELPTIYAE-----LDRLLADYGDY-SFREYATDLLDLA--RLL 144

      170     180     190     200     210     220
      .....|.....|.....|.....|.....|.....|.....|.....|
seqsig_MDPNQ_c8cd862d30a647573b0348f5f8ac9673 301 AQDHDYIVLHDPESFSAKELAEVFARKTSYWNSTKtiknnsEVFAKVLRCVWQELLLSKYGIAY 365
Cdd:pfam13208      145 ASPDPYLLPPPETSSAYELPLSLRVALGQYAADG-----EPLPAEWALAWALLHPCLKLRTPA 203

```

Pssm-ID: 406129 Cd Length: 143 Bit Score: 70.84 E-value: 1.54e-14

```

      10      20      30      40      50      60      70      80
      .....|.....|.....|.....|.....|.....|.....|.....|
seqsig_MDPNQ_c8cd862d30a647573b0348f5f8ac9673 467 YQIQEKKSRIQINLSNLETIRNNASKTRDSL---LTDEEKALEKEEQKQEEESKIESKVEITDN-DYDLNKDEIFLLL 542
Cdd:pfam15615      5 EEAAEAARK--GISLDLDRIAAIQEETAAVSALLaeiFEEETEELAEILEPPESEEAETALADKvDEGLDEAHSAPLR 82

      90     100     110     120     130     140
      .....|.....|.....|.....|.....|.....|.....|.....|
seqsig_MDPNQ_c8cd862d30a647573b0348f5f8ac9673 543 ALLQDQFW----QEYIKVHHLMDSILVDSINEKLFDEFPDGSVIEFDekDQPRIIEDYQTDLE 600
Cdd:pfam15615      83 LLLARAQWpreevESLARDHGLMPDGATESINENAFDEFDLVIEDD--DPFIEINEDYLEELR 143

```

Fig3(a)

Coils output for UNKNOWN

Fig3(c)

Fig4

Fig5(a)

Legend

Description

1 ————— 1

Amino acid residue numeration

Protein secondary structure

H-alpha and other helices (model 1)

H-alpha and other helices (model 2)

E-beta-strand or bridge

C-coil

Relative solvent accessibility (RSA)

0 1 2 3 4 5 6 7 8 9

0-completely buried (0-9% RSA),
9-fully exposed (90-100% RSA)

Fig5(b)

Fig5(c)

Fig6

Global Quality Estimate

Fig7(a)

Local Quality Estimate

Fig7(b)

Comparison with Non-redundant Set of PDB Structures

Fig7(c)

Fig8

Fig9(a)

Chain X

Fig9(b)

[illegible]

Fig3(b)