Research Article

Molecular Structure and Functional Elucidation of IF-3 Protein of *Chloroflexus aurantiacus* Involved in Protein Biosynthesis: An In-Silico Approach

Abu Saim Mohammad Saikat*

Department of Biochemistry and Molecular Biology, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh

*Correspondance

Abu Saim Mohammad Saikat

Department of Biochemistry and Molecular Biology, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh

Email: asmsaikat.bmb@gmail.com; Phone: +8801795840353

Abstract

Chloroflexus aurantiacus is a thermophilic bacterium that produces a multitude of proteins within its genome. Bioinformatics strategies can facilitate comprehending this organism through functional and structural interpretation assessments. This study aimed to allocate the structure and function through an in-silico approach required for bacterial protein biosynthesis. This in-silico viewpoint provides copious properties, including the physicochemical properties, subcellular location, three-dimensional structure, protein-protein interactions, and functional elucidation of the protein (WP_012256288.1). The STRING program is utilized for the explication of protein-protein interactions. The in-silico investigation documented the protein's hydrophilic nature with predominantly alpha (α) helices in its secondary structure. The tertiary-structure model of the protein has been shown to exhibit reasonably high consistency based on various quality assessment methods. The functional interpretation suggested that the protein can act as a translation initiation factor, a protein required for translation and protein biosynthesis. Protein-protein interactions also demonstrated high credence that the protein interconnected with 30S ribosomal subunit involved in protein synthesis. This study is bioinformatically examined that the protein (WP_012256288.1) is affiliated in protein biosynthesis as a translation initiation factor IF-3 of *C. aurantiacus*.

Keywords

Chloroflexus aurantiacus, translation initiation factor IF-3, Homology modeling, Protein-protein interaction, Functional annotation, Protein synthesis

Introduction

In specific, *Chloroflexus aurantiacus* is a Gram-negative organism possessing exceptional characteristics, such as anoxygenic, filamentous, thermophilic, phototrophic, and gliding properties [1] [2][3]. Keeping out other phototrophic anoxygenic, *Chloroflexus aurantiacus* sprout effectively in environments with a moderate temperature of 50–60 °C [4][5]. They can mostly acclimatize in various environmental circumstances, including wetlands, river water, hot springs, and sediments containing elevated-sulfide conditions [6][7]. Surprisingly, the species of bacteria have specific similar characteristics, particularly chimeric photosystem, with purple-photosynthetic bacteria (PPB) and green-sulfur bacteria (GSB) [8][9]. Placed in a certain way, the Chloroflexi are the first expanding bacteria that can generate their nutrients using photosynthesis [8]. Regarding the remarkable photosynthetic and thermophilic properties, the bacterium compelled investigators to examine multiple proteins involved with heat tolerance, formulating industrially crucial enzymes including propionyl-CoA synthase [10][11], maltotetraose producing amylase [12], malonyl-CoA reductase, and so on in recent years [13]. Additionally, highlighting the genome's special features has attracted much attention by studying genome repositories.

Due to advances in computational biology, various platforms and methods have been built for predicting protein structure, recognizing sequence similarities performing phylogenic research, analyzing active site residue correlation, protein-ligand interaction, protein-protein interaction, gene expression screening, motif phosphorylation areas recognition, conserved domains determination [14][15][16][17]. A study using bioinformatics methods of the proteins allows one to evaluate their three-dimensional structural conformation, classify new domains, examine specific pathways to obtain a perspective of our evolutionary tree, identify additional clusters, and attach the proteins' role [18]. This accomplished knowledge can also impart effective pharmacological strategies and assistance in prospective drug design against many diseases [19][20][21].

The protein translation initiation factor IF-3 (WP_012256288.1) is deeply-associated with protein biosynthesis in *C. aurantiacus*. The translation is the final phase of gene expression, which involves translating DNA into RNA and using the RNA to create amino acid chains. Translation includes four distinct stages. These phases include a pre-translational step, initial elongation, termination, and ribosome retrieval. Throughout each step, ribosomes interact with allied translation elements to relay signals essential for protein formulation. It is also crucial to know that the ribosome's conformational mechanisms, translation stimuli, and ribosomal complexes perform a crucial function in directing the translation system's directionality. A key obstacle for the scientists is to grasp how the poorly combined movements of the translational elements contribute to right and rapid protein synthesis [22]. IF-3 is one of the crucial elements required to stimulate the start of protein synthesis in prokaryotes. IF-3 attaches to the 30S ribosomal subunit (RS) and switches the balance between the 70S ribosomes and their available subunits in a manner that enhances the supply of free subunits, thereby maximizing the abundance of novel proteins ready to be constructed [23][24] [25].

Additionally, this assessment enables the recognization of novel biotechnological targets through an adaptive mechanism that involves functional annotation, contemporary gene annotation, and three-dimensional protein modeling.

Methodology

1. Protein selection and sequence retrieval

The amino acid (aa) sequence of the translation initiation factor IF-3 protein present in *Chloroflexus aurantiacus* was retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) in FASTA format.

2. Physicochemical characterization

The characterization of the physicochemical parameters of the protein (WP_012256288.1) was evaluated by the ProtParam assessment tool of the ExPASy server (https://web.expasy.org/protparam/) and the SMS v.2.0 server (https://www.bioinformatics.org/sms2/).

3. Subcellular location identification

The subcellular location of the protein was documented by utilizing the CELLO v.2.5 [26][27], PSORTb v.3.0.2 [28], the SOSUI assessment tool [29], PSLpred server [30], HMMTOP v.2.0 [31], and TMHMM server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

4. Functional annotation prediction

The NCBI platform's CD Search tool [32] was utilized to predict the conserved domain in the protein WP_012256288.1. Protein motif determination was performed using the GenomeNet (Motif) server [33], Pfam tool [34], ScanProsite tool (https://prosite.expasy.org/scanprosite/) of the ExPasy program, and the SuperFamily program [35] assigned the evolutionary relationships of the protein WP_012256288.1.

5. Protein-Protein interaction

The STRING v.11.0 program [36] was used for determining the possible protein-protein (pr-pr) interactions.

6. Secondary Structural assessment

The SOPMA tool [37] utilized the secondary structural elements' prediction following the default parameters (window width of 17, number of states of 4, and the similarity threshold of 8) of the protein translation initiation factor IF-3 present in *Chloroflexus aurantiacus*. The SPIPRED v.4.0 [38] and the DISOPRED v.3.0 [39] tools were utilized to predict the secondary structure and the disordered areas, respectively.

7. Three-dimensional structure prediction and validation

The three-dimensional (tertiary) structure was predicted with Modeller [40] following the HHpred tool [41][42]. The most suitable template (Hit: 5LMN_X; PDB ID: 5LMN) was selected for designing the tertiary structure among the number of hits of 130 with the probability, E-value, Aligned cols, and target lengths of 100, 2.5×10^{-37}, 168, and 171, respectively. The PROCHECK [43] of the SAVES v.6.0 program (https://saves.mbi.ucla.edu/) was performed to predict the Ramachandran Plot and validate the predicted tertiary structure.

8. Active site determination

The CASTp v.3.0 server [44] performed to predict the active sites of the modeled protein.

Results and Discussion

1. Sequence retrieval

The amino acid (aa) sequence of the protein (WP_012256288.1) of *Chloroflexus aurantiacus* was gathered from the NCBI database. The protein contains 275 amino acids. Further information on the protein (WP_012256288.1) is mentioned in **Table 1.**

Table 1 Protein retrieval

Protein individualities	**Protein Information**
Locus	WP_012256288
Amino acid	275 aa
Definition	Translation initiation factor IF-3 [*Chloroflexus aurantiacus*]
Accession	WP_012256288
Version	WP_012256288.1
Source	*Chloroflexus aurantiacus*
Keywords	RefSeq
Organism	*Chloroflexus aurantiacus*
FASTA sequence	>WP_012256288.1 translation initiation factor IF-3 [Chloroflexus aurantiacus] MPRLSPVARRRSRAIRDRFRINNRIRAREVRLIDENGTQVGI VPLREALAMAEERGFDLVEVAPNAVPPVCRLLDYGKFRYE QSKKEREARRNQKQSELKQIRLMPKTDDHDVAVKANQARR FLLAGDKVKFNLRFRGREMAHPEIGRQMLDQIAEQLSDIAVI EQKPLMEGRVLSMLLAPTAKVLKAAQQAQKAAAQRTTTA ESAKPATSAASTPATAEPADEEEEELIDDGDVVEEDEDDDD TFVADYDDEDDDFEDDDDDDEDDERNRRRRR

2. Physicochemical properties

Through studying the characteristics of each of the amino acids in the protein, it can be understood how the physicochemical features of the protein are defined. The ProtParam program of the ExPASy server was utilized to define the physicochemical properties of the protein (WP_012256288.1). The protein is consist of 275 amino acids where Arg (34) was the most abundant amino acid followed by Ala (33), Asp (33), Glu (29), Leu (20), Val (17), Lys (14), Gln (14), Pro (13), Ile (11), Thr (10), Gly (9), Ser (9), Asn (8), Phe (8), Met (7), Tyr (3), His (2), and Cys (1). There was no amino acid residue tryptophan (Trp) in the protein. Protein half-life is characterized as the period it requires for the radiolabeled focus protein density to be decreased by 50 percent compared to the amount at the onset of the chase [45]. The protein (WP_012256288.1) *Chloroflexus aurantiacus* has an estimated half-life of about 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo), and >10 hours (*Escherichia coli*, in vivo). The calculated isoelectric point (pI), molecular weight, the total number of atoms were 4.88 (4.62*), 31444.01 Dalton, and 4384, respectively (**Table 2**). Besides, the molecular formula of the protein was $C_{1336}H_{2179}N_{417}O_{444}S_8$. The pI value introduced the protein is negatively charged where the total number of negatively charged residues (Asp + Glu) was 62, and the total number of positively charged residues (Arg + Lys) was 48. Other parameters, including the instability index (II), describe the proteins' stability, whereas the aliphatic index (73.89) determines its balance over a broad temperature scale. The GRAVY index determines the proteins' solubility [46]. The negative value of GRAVY (-0.931) indicated the hydrophilic nature of the protein.

Table 2 Physicochemical parameters

Parameters	**Value**
Molecular weight	31444.01
Theoretical pI	4.88, 4.62*
Total number of negatively charged residues (Asp + Glu)	62
Total number of positively charged residues (Arg + Lys)	48
Formula	$C_{1336}H_{2179}N_{417}O_{444}S_8$
Total number of atoms	4384
The estimated half-life	a) 30 hours (mammalian reticulocytes, in vitro). b) >20 hours (yeast, in vivo) c) >10 hours (*Escherichia coli*, in vivo)
Instability index (II)	60.38
Aliphatic index	73.89
Grand average of hydropathicity (GRAVY)	-0.931

*pI calculated by the SMS v.2.0

3. Subcellular location determination

The CELLO (v.2.5), PSORTb (v.3.0.2), $SOSUI_{Gram}N$, and PSLpred tools utilized for subcellular location assessment of the protein (WP_012256288.1). The tools predicted the subcellular location of the protein as a cytoplasmic protein. The HMMTOP (v.2.0) and the TMHMM (v.2.0) programs

predicted that there were no transmembrane-helices in the protein (WP_012256288.1) and emphasized the cytoplasmic location of the protein present in *Chloroflexus aurantiacus* (**Table 3**).

Table 3 Subcellular localization assessment

Analysis	**Result**
CELLO (v.2.5)	Cytoplasmic
PSORTb (v.3.0.2)	Cytoplasmic
$SOSUI_{Gram}N$	Cytoplasmic
PSLpred	Cytoplasmic
HMMTOP (v.2.0)	No transmembrane helices present
TMHMM (v.2.0)	No transmembrane helices present

4. Functional annotation of WP_012256288.1

The CDD tool of NCBI characterizes the domain that is found in the identical protein sequences. CD-Search employs RPS-BLAST to assess a test sequence across position-specific rating datasets that have been assembled from conserved domain (CD) alignments contained in the CD protein-cluster. The CD-search tool predicted a conserved-domain as a translation initiation factor IF-3 (infC, accession No. PRK00028) of the protein WP_012256288.1. The IF-3 is one of the crucial elements for the onset of protein synthesis. It attaches to a 30S ribosomal subgroup, shifting the balance between 70S ribosomes and their 50S and 30S subgroups towards free subunits and thereby increasing the suitability of 30S subunits where protein synthesis activation starts. Besides, the ScanProsite program predicted a motif (position: 72 – 85; accession No. PS00938) as IF-3 (gene: infC), which is one of the primary elements required for protein biosynthesis in bacterial [47]. Also, the Pfam program described two different motifs at the positions of 98 – 181 (Pfam ID: IF3_C; IF-3, C-terminal domain; e-value of 2.4×10^{-34}), and 21 – 90 (Pfam ID: IF3_N; IF-3, N-terminal domain; e-value of 4.0×10^{-33}). The CDD tool also validated the domains IF3_C and IF3_N at 98 – 181 and 21 – 90. The IF3_C (CDD No. pfam00707) is the only member of the superfamily cl29551, whereas the IF3_N (CDD No. pfam05198) is the only member of the superfamily cl04980 as of the Conserved Protein Domain Family search feature by the CDD program. The SuperFamily tool predicted the protein WP_012256288.1 (**Figure 1**) as deeply-associated with the infC Superfamily (e-value of 2.09×10^{-98}). The x-axis of the diagram displays the location in the amino acid (aa) count protein (beginning at the N-terminus), and the Y-axis indicates the coiled-coil, while the 'Window' corresponds to the amino acids' 'window' which is examined simultaneously (**Figure 1**).

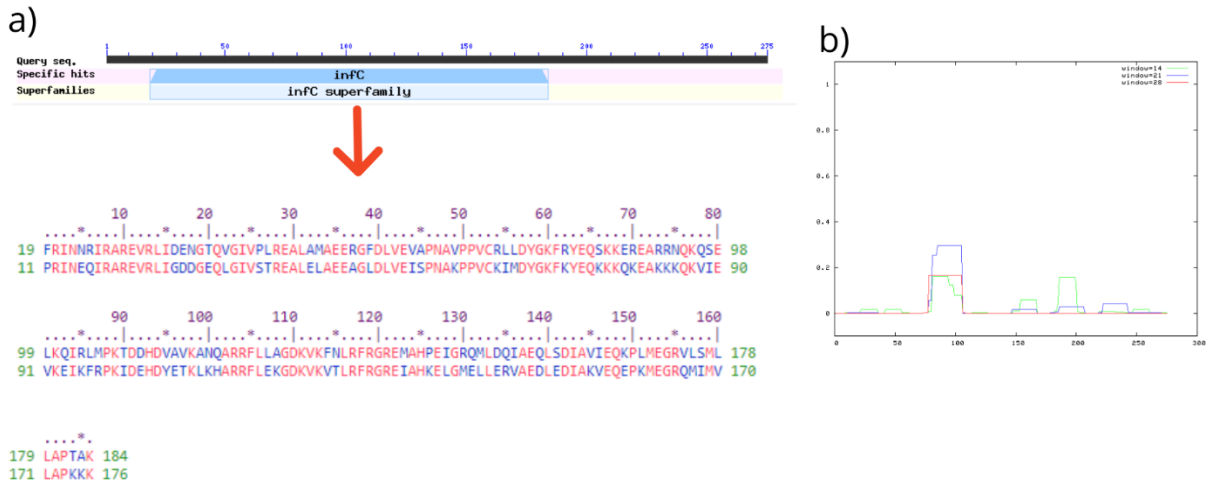

Figure 1 a) Functional annotation of the hypothetical protein. Moreover, **b)** Coil reveals the heptads of existing windows 14 (green color), 21(blue color), and 28 (red color).

5. Protein-protein interaction

The primary focus of protein-protein interactions is acknowledging how cellular systems operate. Such connections allow the filtering, evaluating, and validating of functional genomics data and offering an insightful platform for annotating functional, structural, and evolutionary features of proteins. The platform can furnish predictions for prospective experiments and map the interactions between different species [48]. The STRING v.11.0 program was performed to determine the protein-protein (pr-pr) interaction. The STRING program determined the functional fellows with scores as of rpsM (0.990), rpsE (0.988), rpsK (0.988), rpsS (0.987), rpsI(0.983), rpIT(0.980), rpsC (0.980), rpsJ (0.964), rpsR (0.955), and rpsB (0.951). The rpsM, rpsE, rpsK, rpsS, rpsI, rpIT, rpsC, rpsJ, rpsR, and rpsB are the 30S ribosomal protein S13, 30S ribosomal protein S5, 30S ribosomal protein S11, 30S ribosomal protein S19, ribosomal protein S9 belongs to the universal ribosomal protein uS9 family, 50S ribosomal protein L20, 30S ribosomal protein S3, 30S ribosomal protein S10, 30S ribosomal protein S18, and ribosomal protein S belongs to the universal ribosomal protein uS2 family, respectively (**Figure 2**).

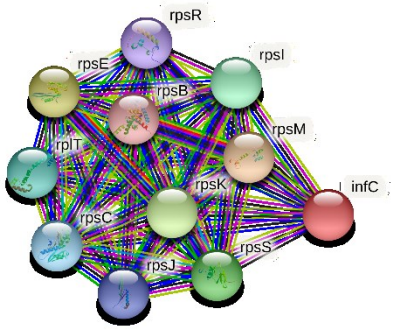

Figure 2 String network of the protein determines the protein-protein interactions

6. Secondary structure inquiry

Protein structure and function are strongly connected. The secondary structural components, e.g., helix, coil, sheet, and turn, have an excellent relationship with protein s function, structure, and engagement [49][50]. The SOPMA program predicted the secondary-structural element of the protein (WP_012256288.1) as of the Alpha helix (Hh), Extended strand (Ee), Beta turn (Tt), Random coil (Cc) were 121 (44.00%), 45 (16.36%), 23 (8.36%), and 86 (31.27%), respectively (**Table 4**). The SPIPRED v.4.0 and DISOPRED v.3.0 tools predicted the sequence plot, secondary structure, and transmembrane topology (**Figure 3**).

Table 4 Secondary structural elements

Structural elements	**Values (%)**
Alpha helix (Hh)	121 (44.00)
3_{10} helix (Gg)	0
Pi helix (Ii)	0
Extended strand (Ee)	45 (16.36)
Beta bridge (Bb)	0 (0.00)
Bend region (Ss)	0 (0.00)
Beta turn (Tt)	23 (8.36)
Random coil (Cc)	86 (31.27)
Ambiguous states	0
Other states	0

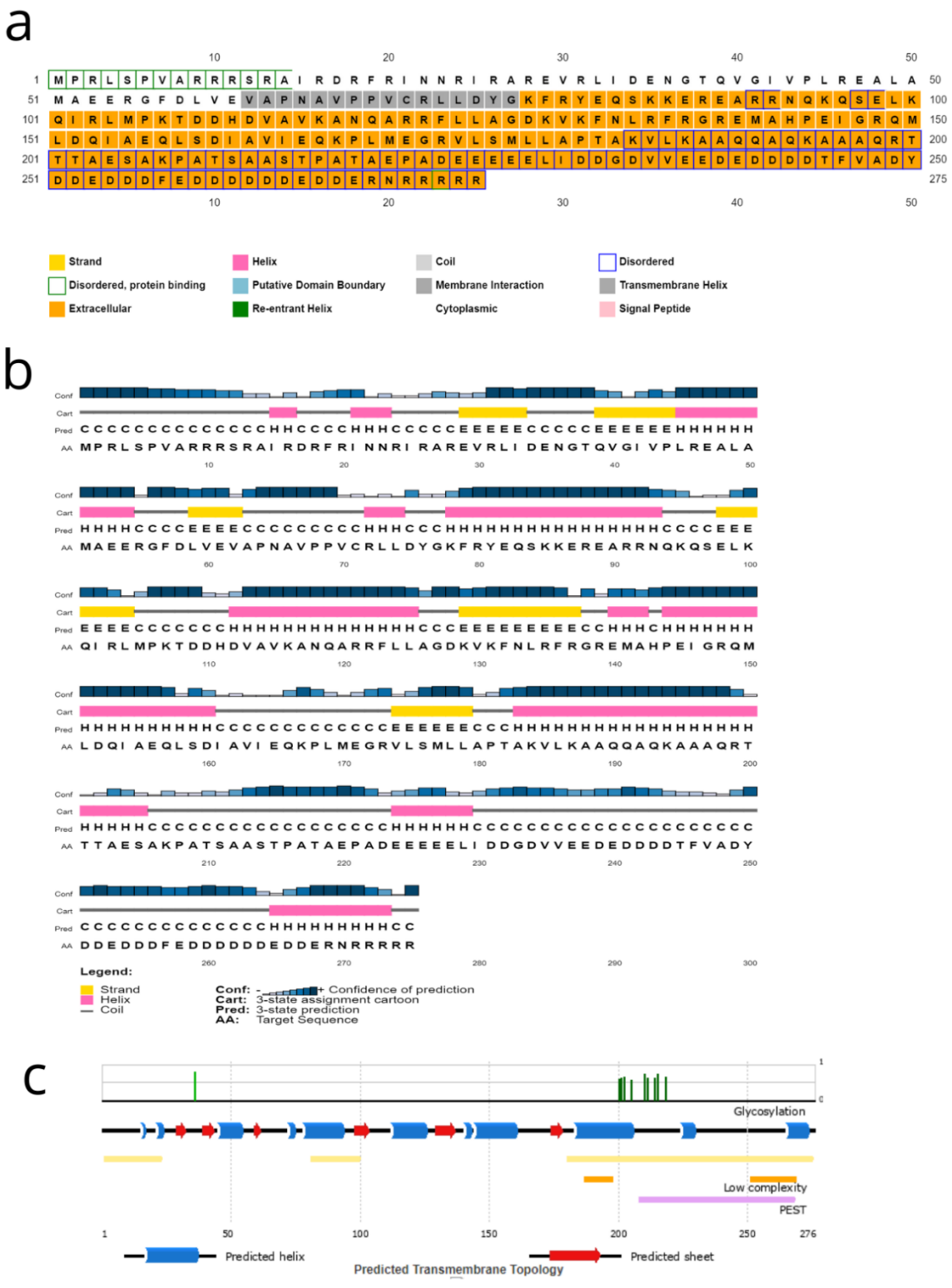

Figure 3 The secondary structural assessment. **a**) Sequence Plot, **b**) The predicted secondary

structure, **c**) Predicted Transmembrane Topology (Position dependent feature predictions are mapped onto the sequence schematic phenomena. The line-height of the Phosphorylation and Glycosylation features reflects the confidence of the residue prediction).

7. Tertiary-structure prediction and validation

Homology modeling (HM) is an essential method for estimating protein architecture when solely amino acid sequence information is accessible. Protein activities can be derived from the composition of the chain. Using homology modeling (HM) or comparative modeling (CM), scientists would quickly evaluate two closely related sequences' similarities and roles. Sequence similarity to a defined structure is typically representative of translational and structural similarities to that structure. In the face of these constraints, sequence similarity below 30% will never provide suitable efficiency in structure prediction [51][52]. The HHpred is a powerful platform used for distant homology identification and structure estimation, implemented initially as hidden Markov models (HMMs), pioneered by the earliest pairwise comparative analysis of homologous protein profiles. It enables a broad range of repositories, including PDB, Pfam, SCOP, COG, SMART, and CDD. It admits a solitary query array or multiple lineups as input, and it delivers the findings to a PSI-BLAST-like user-friendly interface. Search features are including local or global integration and the detection of secondary systems. HHpred can generate a pair of query prototypes, multiple model alignments with several frameworks from the lookup findings, and 3D structural models from these configurations computed with the Modeller program [53]. The three-dimensional structure of the protein (WP_012256288.1) predicted by Modeller following the HHpred tool (**Figure 4**) with 100% identity and the probability of the template (HHped Hit: 5LMN_X; PDB ID: 5LMN). The 5LMN is the crystal structure (Method: electron microscopy; Resolution: 3.55 Å; Reconstruction Method: Single Particle) of the bacterial 30S-IF1-IF3-mRNA translation pre-initiation complex (state-1A) of *Thermus thermophilus* HB8, where the expression system was *Escherichia coli* BL21(DE3) [54]. The SAVES server's PROCHECK program was utilized for structural quality assessment of the modeled protein, where the arrangement of the ψ angle and the φ angle are shown (**Table 5, Figure 4**). Residues in most favored regions engulfed 92.0%, which validated the protein's modeled tertiary structure (WP_012256288.1). Also, residues in additional allowed regions generously allowed regions, disallowed regions, No. of non-glycine and non-proline residues, No. end-residues (excl. Gly and Pro) were 10 (6.7%), 1 (0.7%), 1 (0.7%), 150, and 2, respectively. The No. of glycine residues and the No. of proline residues were similar (8 residues) found in the protein 3D structure.

Table 5 Ramachandran Plot statistics of the modeled protein

Ramachandran Plot statistics	**Value (%)**
Residues in most favored regions [A,B,L]	138 (92.0)
Residues in additional allowed regions [a,b,l,p]	10 (6.7)
Residues in generously allowed regions [~a,~b,~l,~p]	1 (0.7)
Residues in disallowed regions	1 (0.7)
Number of non-glycine and non-proline residues	150
Number of end-residues (excl. Gly and Pro)	2
Number of glycine residues (shown as triangles)	8

Number of proline residues	8
Total number of residues	168

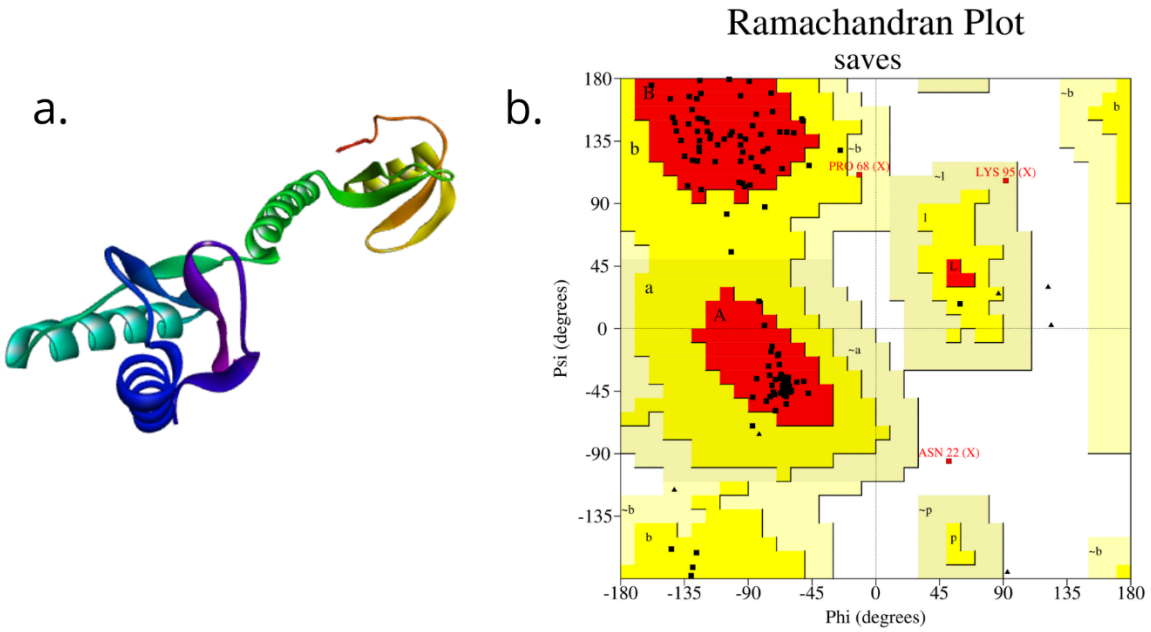

Figure 4 Tertiary structure prediction. **a**) Predicted tertiary structure by Modeller following HHpred, and **b**) The Ramachandran Plot statistics of the modeled three-dimensional (tertiary) structure validated by the PROCHECK program.

8. Active site determination

The CASTp v.3.0 program predicted 21 different active sites of the modeled protein (**Figure 4**). CASTp is a database server that can locate areas on proteins, delineate their outline, find the areas' dimensions, and calculate the regions' area. This involves pockets on protein surfaces and vacuums concealed within proteins. The calculation consists of a pocket and volume spectrum or vacuum, both mathematically determined by a solvent-accessible surface (surface of Richards) and molecular surface model (surface of Connolly). CASTp could be utilized for the investigation of surface properties and protein operational zones. CASTp provides a pictorial, user interface versatile, dynamic view and user-submitted constructs on-the-fly measurement [44]. The maximum active-sites of the modeled protein were identified between the area of 85.302 and the volume of 50.667 (**Figure 5**).

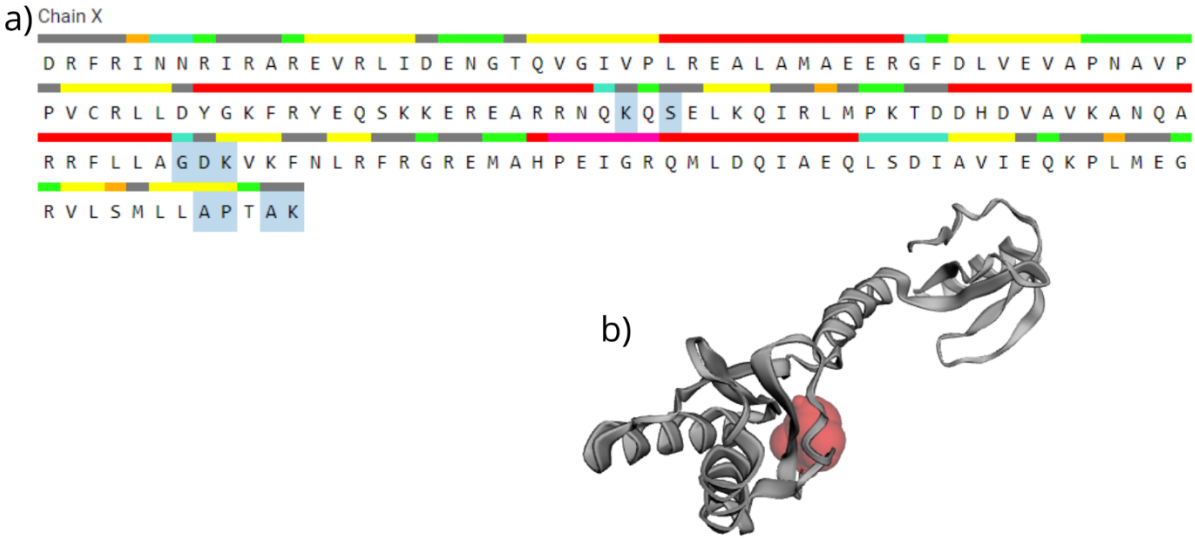

Figure 5 Active site determination. **a**) The amino acid residues in the active site (blue color), and **b**) Active sites of the protein (WP_012256288.1). Here, the 'red-sphere' indicates the active protein sites.

Conclusions

Comprehending how proteins act is essential for explaining how they operate, and this protein contains IF-3, a crucial factor in protein synthesis considered to initiate protein synthesis. IF-3 connects to the 30S ribosomal subunit and alters the balance between the 70S ribosomes and their 50S and 30S subunits, thereby strengthening the abundance of 30S subunits, enhancing the affordability of amino acids for the initiation of protein biosynthesis. This investigation reveals the fundamental characteristics, including cytoplasmic nature and functional annotation of the protein in association with tertiary structure. Thus, the study findings show the efficiency and scale of further studies on the IF-3 protein of bioinformatics methods used in this investigation.

Abbreviations

NCBI: National Center for Biotechnology Information; PDB: Protein Data Bank; SMS: The Sequence Manipulation Suite; CDD: Conserved Domain Database; RPS-BLAST: Reverse Position-Specific BLAST; BLAST: Basic Local Alignment Search Tool; GRAVY: Grand Average of Hydropathicity; CELLO: Subcellular Localization predictor; PSLpred: Prediction of subcellular localization of bacterial proteins; HMMTOP: Prediction of transmembrane helices and topology of proteins; SPIPRED: PSI-blast based secondary structure Prediction; SOPMA: Self Optimized Prediction Method; CASTp: Computed Atlas of Surface Topography of proteins.

ORCID

Abu Saim Mohammad Saikat: https://orcid.org/0000-0002-6850-1245

Conflicts of Interest

There is no conflict of interest.

Acknowledgment

None.

References

[1] B.K. Pierson, R.W. Castenholz, A phototrophic gliding filamentous bacterium of hot springs, Chloroflexus aurantiacus, gen. and sp. nov., Arch. Microbiol. 100 (1974) 5–24. https://doi.org/10.1007/BF00446302.

[2] J.F. Holden, Extremophiles: Hot Environments, Encycl. Microbiol. (2009) 127–146. https://doi.org/10.1016/B978-012373944-5.00281-9.

[3] Y. Xin, Y.K. Lu, R. Fromme, P. Fromme, R.E. Blankenship, Purification, characterization and crystallization of menaquinol:fumarate oxidoreductase from the green filamentous photosynthetic bacterium Chloroflexus aurantiacus, Biochim. Biophys. Acta - Bioenerg. 1787 (2009) 86–96. https://doi.org/10.1016/j.bbabio.2008.11.010.

[4] R.G. Feickt, R.C. Fuller, Topography of the photosynthetic apparatus of Chloroflexus aurantiacus, Biochemistry. 23 (1984) 3693–3700. https://doi.org/https://doi.org/10.1021/bi00311a019.

[5] L. He, Y. Wang, L. You, Y. Khin, J.K.H. Tang, Y.J. Tang, Glycine cleavage powers photoheterotrophic growth of Chloroflexus aurantiacus in the absence of H2, Front. Microbiol. 6 (2015) 1–7. https://doi.org/10.3389/fmicb.2015.01467.

[6] J. Oelze, R.C. Fuller, Temperature dependence of growth and membrane-bound activities of Chloroflexus aurantiacus energy metabolism, J. Bacteriol. 155 (1983) 90–96. https://doi.org/10.1128/jb.155.1.90-96.1983.

[7] C.G. Klatt, Z. Liu, M. Ludwig, M. Kühl, S.I. Jensen, D.A. Bryant, D.M. Ward, Temporal metatranscriptomic patterning in phototrophic Chloroflexi inhabiting a microbial mat in a geothermal spring, ISME J. 7 (2013) 1775–1789. https://doi.org/10.1038/ismej.2013.52.

[8] K.H. Tang, K. Barry, O. Chertkov, E. Dalin, C.S. Han, L.J. Hauser, B.M. Honchak, L.E. Karbach, M.L. Land, A. Lapidus, F.W. Larimer, N. Mikhailova, S. Pitluck, B.K. Pierson, R.E. Blankenship, Complete genome sequence of the filamentous anoxygenic phototrophic bacterium Chloroflexus aurantiacus, BMC Genomics. 12 (2011) 1–21. https://doi.org/10.1186/1471-2164-12-334.

[9] O.A. Zadvornyy, E.S. Boyd, M.C. Posewitz, N.A. Zorin, J.W. Peters, Biochemical and structural characterization of enolase from Chloroflexus aurantiacus: evidence for a

thermophilic origin, Front. Bioeng. Biotechnol. 3 (2015) 74. https://doi.org/10.3389/fbioe.2015.00074.

[10] B.E. Alber, G. Fuchs, Propionyl-coenzyme a synthase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation, J. Biol. Chem. 277 (2002) 12137–12143. https://doi.org/10.1074/jbc.M110802200.

[11] B.E. Alber, G. Fuchs, Propionyl-coenzyme a synthase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation, J. Biol. Chem. 277 (2002) 12137–12143. https://doi.org/10.1074/jbc.M110802200.

[12] K. Ratanakhanokchai, J. Kaneko, Y. Kamio, K. Izaki, Purification and properties of a maltotetraose- and maltotriose-producing amylase from Chloroflexus aurantiacus, Appl. Environ. Microbiol. 58 (1992) 2490–2494. https://doi.org/10.1128/aem.58.8.2490-2494.1992.

[13] M. Hügler, C. Menendez, H. Schägger, G. Fuchs, Malonyl-Coenzyme A Reductase from Chloroflexus aurantiacus, a Key Enzyme of the 3-Hydroxypropionate Cycle for Autotrophic CO2 Fixation, J. Bacteriol. 184 (2002) 2404–10. https://doi.org/10.1128/JB.184.9.2404-2410.2002.

[14] J. Gong, Y. Chen, F. Pu, P. Sun, F. He, L. Zhang, Y. Li, Z. Ma, H. Wang, Understanding Membrane Protein Drug Targets in Computational Perspective, Curr. Drug Targets. 20 (2018) 551–564. https://doi.org/10.2174/1389450120666181204164721.

[15] F. Canduri, P. Cardoso Perez, R. Caceres, W. de Azevedo, Protein Kinases as Targets for Antiparasitic Chemotherapy Drugs, Curr. Drug Targets. 8 (2007) 389–398. https://doi.org/10.2174/138945007780058979.

[16] A.S.M. Saikat, R. Islam, S. Mahmud, M.A.S. Imran, M.S. Alam, M.H. Masud, M.E. Uddin, Structural and Functional Annotation of Uncharacterized Protein NCGM946K2_146 of Mycobacterium Tuberculosis: An In-Silico Approach, Proceedings. 66 (2020) 13. https://doi.org/10.3390/proceedings2020066013.

[17] J.H. Pereira, F. Canduri, J.S. De Oliveira, N.J.F. Da Silveira, L.A. Basso, M.S. Palma, W.F. De Azevedo, D.S. Santos, Structural bioinformatics study of EPSP synthase from Mycobacterium tuberculosis, Biochem. Biophys. Res. Commun. 312 (2003) 608–614. https://doi.org/10.1016/j.bbrc.2003.10.175.

[18] C.L. Mills, P.J. Beuning, M.J. Ondrechen, Biochemical functional predictions for protein structures of unknown or uncertain function, Comput. Struct. Biotechnol. J. 13 (2015) 182–191. https://doi.org/10.1016/j.csbj.2015.02.003.

[19] W.F. De Azevedo, G.C. Dos Santos, D.M. Dos Santos, J.R. Olivieri, F. Canduri, R.G. Silva, L.A. Basso, G. Renard, I.O. Da Fonseca, M.A. Mendes, M.S. Palma, D.S. Santos, Docking and small angle X-ray scattering studies of purine nucleoside phosphorylase, Biochem. Biophys. Res. Commun. 309 (2003) 923–928. https://doi.org/10.1016/j.bbrc.2003.08.093.

[20] W.F. de Azevedo, Molecular Dynamics Simulations of Protein Targets Identified in, Curr. Med. Chem. 19 (2011) 1353–1366.

[21] W.F. De Azevedo, F. Canduri, V. Fadel, L.G.V.L. Teodoro, V. Hial, R.A.S. Gomes, Molecular model for the binary complex of uropepsin and pepstatin, Biochem. Biophys. Res. Commun. 287 (2001) 277–281. https://doi.org/10.1006/bbrc.2001.5555.

[22] M. V. Rodnina, Translation in prokaryotes, Cold Spring Harb. Perspect. Biol. 10 (2018) 1–22. https://doi.org/10.1101/cshperspect.a032664.

[23] M. Paci, C. Pon, M. Lammi, C. Gualerzi, Structure-function relationship in Escherichia coli translational initiation factors. Characterization of IF-3 by high resolution 1H NMR spectroscopy, J. Biol. Chem. 259 (1984) 9628–9634. https://doi.org/10.1016/s0021-9258(17)42747-5.

[24] D. Brauer, B. Wittmann-Liebold, The primary structure of the initiation factor IF-3 from Escherichia coli, FEBS Lett. 79 (1977) 269–75. https://doi.org/10.1016/0014-5793(77)80801-6.

[25] Y. Hua, D.P. Raleigh, On the global architecture of initiation factor IF3: A comparative study of the linker regions from the Escherichia coli protein and the Bacillus stearothermophilus protein, J. Mol. Biol. 278 (1998) 871–878. https://doi.org/10.1006/jmbi.1998.1736.

[26] C.-S. Yu, C.-J. Lin, J.-K. Hwang, Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n -peptide compositions , Protein Sci. 13 (2004) 1402–1406. https://doi.org/10.1110/ps.03479604.

[27] G. Sanchez, Las instituciones de ciencia y tecnología en los procesos de aprendizaje de la producción agroalimentaria en Argentina, El Sist. Argentino Innovación Inst. Empres. y Redes. El Desafío La Creación y Apropiación Conoc. 651 (2013) 643–651. https://doi.org/10.1002/prot.

[28] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. Cenk Sahinalp, M. Ester, L.J. Foster, F.S.L. Brinkman, PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics. 26 (2010) 1608–1615. https://doi.org/10.1093/bioinformatics/btq249.

[29] T. Hirokawa, S. Boon-Chieng, S. Mitaku, SOSUI: classification and secondary structure prediction system for membrane proteins, Bioinformatics. 14 (1998) 378–379. https://doi.org/10.1093/bioinformatics/14.4.378.

[30] M. Bhasin, A. Garg, G.P.S. Raghava, PSLpred: Prediction of subcellular localization of bacterial proteins, Bioinformatics. 21 (2005) 2522–2524. https://doi.org/10.1093/bioinformatics/bti309.

[31] G.E. Tusnády, I. Simon, The HMMTOP transmembrane topology prediction server, Bioinformatics. 17 (2001) 849–850. https://doi.org/10.1093/bioinformatics/17.9.849.

[32] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 8 (2020) D265–D268.

https://doi.org/10.1093/nar/gkz991.

[33] M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya, Thed KEGG databases at GenomeNet, Nucleic Acids Res. 30 (2002) 42–46. https://doi.org/10.1093/nar/30.1.42.

[34] R.D. Finn, Pfam: the protein families database, Encycl. Genet. Genomics, Proteomics Bioinforma. 42 (2005) D222-D230. https://doi.org/https://doi.org/10.1002/047001153X.g306303.

[35] D. Wilson, M. Madera, C. Vogel, C. Chothia, J. Gough, The SUPERFAMILY database in 2007: Families and functions, Nucleic Acids Res. 35 (2007) 308–313. https://doi.org/10.1093/nar/gkl910.

[36] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C. Von Mering, STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (2019) D607–D613. https://doi.org/10.1093/nar/gky1131.

[37] C. Combet, C. Blanchet, C. Geourjon, G. Deleage, NPS@: network protein sequence analysis, Trends Biochem. Sci. 25 (2000) 147–150. https://doi.org/https://doi.org/10.1016/S0968-0004(99)01540-6https://doi.org/10.1016/S0968-0004(99)01540-6.

[38] D.W.A. Buchan, D.T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on, Nucleic Acids Res. 47 (2019) W402–W407. https://doi.org/10.1093/nar/gkz297.

[39] D.T. Jones, D. Cozzetto, DISOPRED3: Precise disordered region predictions with annotated protein-binding activity, Bioinformatics. 31 (2015) 857–863. https://doi.org/10.1093/bioinformatics/btu744.

[40] B. Webb, A. Sali, Comparative protein structure modeling using MODELLER, Curr. Protoc. Protein Sci. 2016 (2016) 2.9.1-2.9.37. https://doi.org/10.1002/cpps.20.

[41] L. Zimmermann, A. Stephens, S.Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A.N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core, J. Mol. Biol. 430 (2018) 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

[42] F. Gabler, S.Z. Nam, S. Till, M. Mirdita, M. Steinegger, J. Söding, A.N. Lupas, V. Alva, Protein Sequence Analysis Using the MPI Bioinformatics Toolkit, Curr. Protoc. Bioinforma. 72 (2020) 1–30. https://doi.org/10.1002/cpbi.108.

[43] R.A. Laskowski, M.W. MacArthur, J.M. Thornton, PROCHECK: validation of protein-structure coordinates, J. Appl. Crystallogr. 16 (1983) 284–284. https://doi.org/10.1107/s0021889883010444.

[44] W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: Computed atlas of surface topography of proteins, Nucleic Acids Res. 46 (2018) W363–W367. https://doi.org/10.1093/nar/gky473.

[45] P. Zhou, Determining Protein Half-Lives, Signal Transduct. Protoc. 284 (2004) 67–77.

https://doi.org/https://doi.org/10.1385/1-59259-816-1:067.

[46] P. Smialowski, A.J. Martin-Galiano, A. Mikolajka, T. Girschick, T.A. Holak, D. Frishman, Protein solubility: Sequence based prediction and experimental verification, Bioinformatics. 23 (2007) 2536–2542. https://doi.org/10.1093/bioinformatics/btl623.

[47] D. Liveris, J.J. Schwartz, R. Geertman, I. Schwartz, Molecular cloning and sequencing of infC, the gene encoding translation initiation factor IF3, from four enterobacterial species, FEMS Microbiol. Lett. 112 (1993) 211–216. https://doi.org/10.1111/j.1574-6968.1993.tb06450.x.

[48] A.S. Schwartz, J. Yu, K.R. Gardenour, R.L. Finley, T. Ideker, Cost-effective strategies for completing the interactome, Nat. Methods. 6 (2009) 55–61. https://doi.org/10.1038/nmeth.1283.

[49] H.B. Uchôa, G.E. Jorge, N.J. Freitas Da Silveira, J.C. Camera, F. Canduri, W.F. De Azevedo, Parmodel: A web server for automated comparative modeling of proteins, Biochem. Biophys. Res. Commun. 325 (2004) 1481–1486. https://doi.org/10.1016/j.bbrc.2004.10.192.

[50] W. Filgueira de Azevedo, F. Canduri, J. Simões de Oliveira, L. Augusto Basso, M. Sérgio Palma, J. Henrique Pereira, D. Santiago Santos, Molecular model of shikimate kinase from Mycobacterium tuberculosis, Biochem. Biophys. Res. Commun. 295 (2002) 142–148. https://doi.org/10.1016/S0006-291X(02)00632-0.

[51] M.M. Gromiha, R. Nagarajan, S. Selvaraj, Protein structural bioinformatics: An overview, Encycl. Bioinforma. Comput. Biol. ABC Bioinforma. 1–3 (2018) 445–459. https://doi.org/10.1016/B978-0-12-809633-8.20278-1.

[52] Z. Xiang, Advances in Homology Protein Structure Modeling, Curr. Protein Pept. Sci. 7 (2006) 217–227. https://doi.org/https://doi.org/10.2174/138920306777452312.

[53] J. Söding, A. Biegert, A.N. Lupas, The HHpred interactive server for protein homology detection and structure prediction, Nucleic Acids Res. 33 (2005) 244–248. https://doi.org/10.1093/nar/gki408.

[54] T. Hussain, J.L. Llácer, B.T. Wimberly, J.S. Kieft, V. Ramakrishnan, Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation, Cell. 167 (2016) 133-144.e13. https://doi.org/10.1016/j.cell.2016.08.074.