biology and life sciences forum

MDPI

Proceeding Paper

Protein Characterization, Functional Annotation, Active Site Analysis of Novel Uncharacterized Conserved Protein of *Bacteroides xylanisolvens*: An In Silico Approach †

Mohammed Emon

Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh; 18bge034@bsmrstu.edu.bd

† Presented at the 2nd International Electronic Conference on Microbiology, 1–15 December 2023; Available online: https://ecm2023.sciforum.net.

Abstract: *Bacteroides xylanysolvens* a gram-negative, anaerobic rods as well xylan-degrading bacterium isolated from human feces samples. An in silico technique can help us better comprehend the uncharacterized protein of *Bacteroides xylanysolvens* (accession ID: CUN90054) by examining its functional annotations, identification, and structural characterization. In silico pathway is used to investigate including the protein's physiochemical properties, functional annotation, structural prediction, active site analysis, and sub-cellular localization. The annotated conserved protein is connected to metal-independent alpha-mannosidase (MIAM) including both bacterial and fungal glycoside hydrolases (GH125), according to the current study. This finding may be of considerable relevance to future bacterial genetics research.

Keywords: *Bacteroides xylanysolvens*; in silico characterization; functional annotation; active site analysis; bacterial research

Citation: Emon, M. Protein Characterization, Functional Annotation, Active Site Analysis of Novel Uncharacterized Conserved Protein of *Bacteroides xylanisolvens*: An In Silico Approach. *Biol. Life Sci. Forum* **2023**, *31*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: date

1. Introduction

Bacteroides xylanisolvens is a member of the Bacteroides genus, gram negative, and non-fragilis species which is the most prevalent species of bacteria in the human gut microbiome also contributes to the breakdown of complex carbohydrates, such as xylan, which is a component of plant cell walls, as well as also upregulates the potential for pectin utilization [1–4]. The 2008 description of *Bacteroides xylanisolvens* was centered on the xylanolytic capability of some *Bacteroides* strains that were separated from samples of human feces [4,5]. They can create short fatty acid chains (propionate, succinate) that may have positive health benefits and have immune-modulatory qualities [3]. Due to a renewed interest in the phage treatment and an increasing understanding of the significance of the gut microbiome in human health, there has been a significant upsurge in interest in the phage communities that live in the human gut, also known as the gut phageome [6,7]. Human health and illness may be influenced by the phageome found in the human gut [6]. They can also carry a carbapenem resistance mechanism resembling that of the *B. fragilis* cfiA system (MBL, activation of the IS element, a genetic element encoding the gene for the GNAT toxin, and a comparable frequency of "silent" as well as resistant cases) [4].

2. Materials and Methods

2.1. Protein Sequence Retrieval

The selected protein's amino acid sequence (AAS) was obtained with the accession number CUN90054 (version: CUN90054.1) in the FASTA format from the NCBI website.

2.2. Protein Physicochemical Property Analysis

The physicochemical characteristics was performed using the ExPASy server's Prot Param tool [8] to measure the amino acid sequence composition, instability index, aliphatic index, GRAVY, and extinction coefficients. Furthermore, used the upgrading of the theoretical isoelectric point (pI) for the CUN90054 protein.

2.3. Functional Annotation Anticipation

We utilized the conserved domain (CD) prediction in the protein of CUN90054. We predicted protein motif determination using the NCBI CD search [9] on the NCBI website and the ScanProsite tool [10] of the ExPASy server, correspondingly.

2.4. Prediction of Secondary Structure

The prediction of the secondary structure of the selected protein was anticipated using numerous server such as PSIPRED [11], SOPMA [12], and SABLE [13] widely used remarkable computational tools.

2.5. Three-Dimensional Structure Prediction and Validation

The tertiary (3D) structure of protein predicted by the HHpred [14] and SWISS-MODEL [15] bioinformatics tools and the structural quality assessment studies were performed using the PROCHECK Sever [16] of the SAVES v.6.0 platform.

2.6. Active Site Determination

The CASTp v.3.0 serve was used to predict active sites of the modeled protein [17]. Protein regions may be located, highlighted, their areas calculated, and their dimensions estimated using the CASTp web server [17]. Additionally, it offers accurate and thorough designations measurements of protein topography [17].

2.7. Identification of the Subcellular Location

By using many tools CELLO v.2.5[18], PSLPred [19], PSORTb (v.3.0.2) [20], BUSC [21], HMMTOP v.2.0 [22] and CCTOP [23] protein subcellular localization was identified.

3. Results and Discussions

3.1. Protein Sequence Retrieval

The essential building blocks of biological structure as well as function are determined by protein sequences. The precise NCBI protein sequence database is a protein sequence platform collected from various reliable sources, such as RefSeq, Gene-Bank, SwissProt. PIR etc. The *Bacteroides xylanisolvens* protein obtained from the NCBI protein database with the accession number CUN90054.1 (version: CUN90054.1) is present in the locus CUN90054.1 containing 487 amino acid sequence.

3.2. Physicochemical Properties

In order to calculate and analysis physicochemical parameters of the amino acid sequence of CUN90054.1 (Table 1), which is present in *B. xylanisolvens*, was retrieved in FASTA format and used as a query sequence. A protein whose instability index is smaller than 40 is predicted as stable. To focus on it selected this specific protein the protein's instability index predicted score by ProtParam 32.45 (below 40.00), which means it is stable protein [8]. The protein has a pH of 7.18, which corresponds to its theoretical isoelectric point (pI) and shows that the molecular weight (MW) 55348.50 Da, aliphatic index 78.52, instability index (II) 32.45, and GRAVY value −0.361 (hydrophilic) [8].

Table 1. Several physicochemical properties of the selected conservative protein.

Physicochemical Properties	Value
Number of amino acids	487
Molecular wight	55348.50
Theoretical pI	7.18
Total no. of negatively charged residues (Asp + Glu)	58
Total no. of positively charged residues (Arg + Lys)	58
Total no. of atoms	7741
Instability index (II)	32.45
Aliphatic index	78.52
Grand Average of hydropathicity (GRAVY)	−0.361

3.3. Functional Annotation Prediction

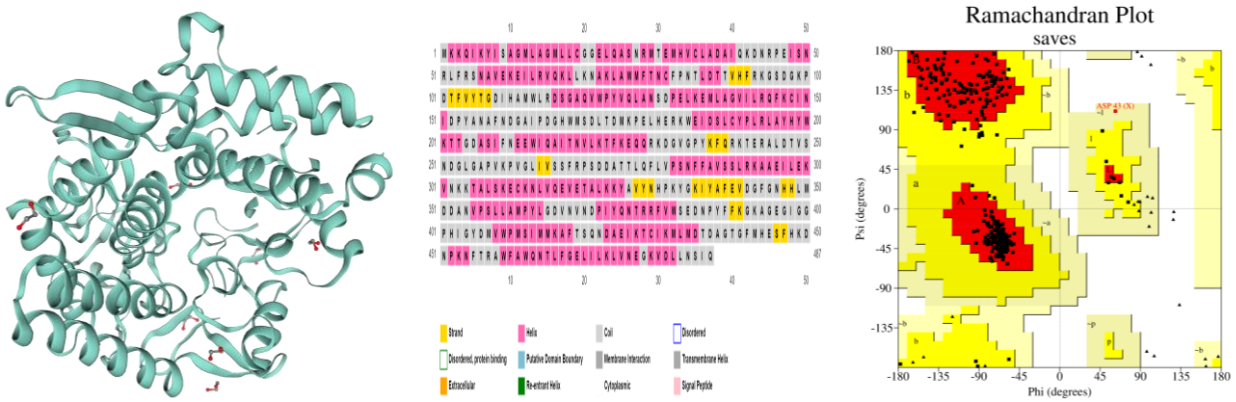
The NCBI Conserved Domain (CD) search tool anticipated two significant domains, including Glyco hydro 125 (GH125) and COG3538. The GH125 (domain architectural ID: 10536255) domain's accession is pfam06825 and interval at 71–472 amino acid residue [9,24]. This domain belongs to family of glycoside hydrolases (GH125) known as Metal-independent alpha-mannosidase (MIAM), which includes enzymes from bacteria and fungi [9,24]. They perform as metal-independent alpha-mannosidases (MIAM) with a preference for terminal mannose residues that are alpha-1,6-linked and non-reducing, and this family belongs to superfamily of six hairpin glycosidase structurally [9]. Another domain COG3538 domain (accession is COG3538 and interval at 49–474 amino acid residue) along with the Glycoside hydrolase family 125 proteins include metal-independent alpha-mannosidase from *C. perfringens* and meiotically up-regulated gene 157 (Mug157) protein from *Schizosaccharomyces pombe* [9,24].

3.4. Secondary Structure Prediction

The ability to forecast the secondary structure of proteins is greatly enhanced by the protein consensus projection's secondary structure assertiveness, which is derived from many alignments. The secondary-structural elements of the protein (CUN90054) was prediction done by the SOPMA server where the alpha helix (Hh), extended strand (Ee), beta-turn (Tt), random coil (Cc) was 221 is 45.38%, 61 is 12.53%, 29 is 5.95%), and 176 is 36.14%, respectively [12].

3.5. Three-Dimentional Structure Prediction and Validation of the Selected Protein

The three-dimensional structure of the desired protein was established using the MODELLER tool (HHPred platform) [14] and SWISS-Model computational program [15]. Significantly, the PROCHECK tool of the SAVES methods was utilized for the structural quality assessment of the modeled protein, where the arrangement of the ψ angle and the φ angle is shown [16]. Based on Ramachandran model a good quality model would be expected to have over 90% in the most favoured regions [16]. Our tertiary protein model structure successfully passed there and was validated by the residues in the most preferred region (A,B,L), which were swallowed by 94.3% and residue in additional allowed regions (a,b,l,p) is 5.4% of the protein (CUN90054.1) (Figure 1) [16]

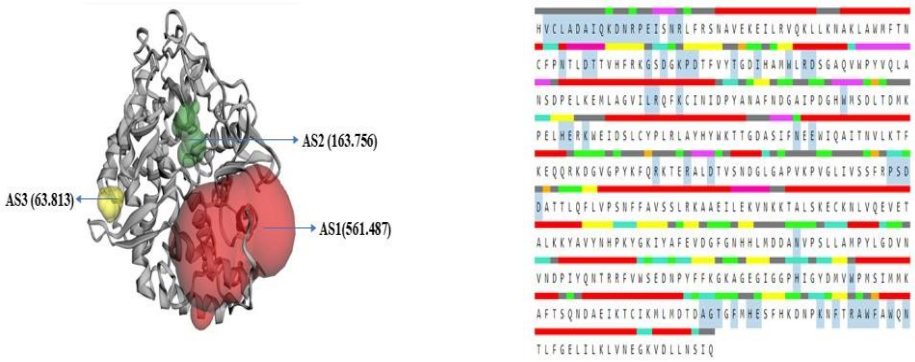


(**a**) (**b**) (**c**)

Figure 1. (**a**) structure of CUN90054 protein predicted by Swiss Model [15] (**b**) The secondary structure of the selected protein anticipated by PSLPred [11] (**c**) Ramachandran plot analysis of conserved protein of *B. xylanisolvens*, protein structure validated by PROCHECK of SAVES V.6.0. server [16].

3.6. Active Site Analysis of Protein

The modeled protein's active sites determinates by using precise and reliable CASTp v.3.0 (Figure 2) [17]. CASTp tool identified 72 surface pockets while probe radius was set to default 1.4 Å, the top act showed in first pocket were found most of area (SA) Å2 561.487 with volume (SA) Å3 1583.046 of *B. xylanisolvens* (CUN90054.1) modeled protein [17].



(**a**) (**b**)

Figure 2. (**a**) Active site of the conserved protein analysis by CASTp server. Here the red sphere indicates active site 1 (AS1) as pocket 1, green is active site 2 as pocket 2 and yellow is active site 3 (AS3) of CUN90054.1 protein. (**b**) Active site sequence indicates blue color in the protein chain [17].

3.7. Subcellular Localization Prediction

The protein's (CUN90054) subcellular localization was evaluated using the several broadly used web tools. CELLO subcellular localization predictor predict as periplasmic with most reliability score 2.881 [18]. These tools including PSLPred and PSORTb (v.3.0.2) predicted subcellular locate of the conservative protein as a cytoplasmic protein and BUSCA web server indicated protein localization at extracellular space [19–21]. HMMTOP and CCTOP servers figure on there is no transmembrane helicase present in the protein CUN90054, all results featured in the Table 2 [22,23].

Table 2. Protein subcellular localization analysis through various computational tools.

Subcellular Servers/Tools	Results
CELLO v.2.5	Periplasmic
PSLPred	Cytoplasmic
PSORTb (v.3.0.2)	Cytoplasmic
BUSCA	Extracellular space
HMMTOP	No transmembrane helices present
CCTOP	No transmembrane helices present

4. Conclusions

This genomics study aims to characterize, functional annotation of conserved protein (CUN90054) of the *B. xylanisolvens* bacteria which is beneficial and marvelous for microbial research. This protein has an ideal AA seq., MW, and GRAVY (negative value) is a hydrophilic nature, that is why it will be easier for the purification of vaccine production. Secondary and 3D qualities of the chosen protein were demonstrated as stable and ideal for further bacterial genomics research by passing several vital validation steps. The selected protein may be used as a target for developing protein-based therapeutics as drugs and vaccines, and immensely help to reduce preponderance of several bacterial infections.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Upon request to corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Despres, J.; Forano, E.; Lepercq, P.; Comtet-Marre, S.; Jubelin, G.; Yeoman, C.J.; Miller, M.E.B.; Fields, C.J.; Terrapon, N.; Le Bourvellec, C.; et al. Unraveling the pectinolytic function of Bacteroides xylanisolvens using a RNA-seq approach and mutagenesis. *BMC Genom.* **2016**, *17*, 147. https://doi.org/10.1186/s12864-016-2472-1.
2. Fernandez-Julia, P.; Black, G.W.; Cheung, W.; Van Sinderen, D.; Munoz-Munoz, J. Fungal β-glucan-facilitated cross-feeding activities between Bacteroides and Bifidobacterium species. *Commun. Biol.* **2023**, *6*, 576. https://doi.org/10.1038/s42003-023-04970-4.
3. Ulsemer, P.; Toutounian, K.; Schmidt, J.; Karsten, U.; Goletz, S. Preliminary safety evaluation of a new Bacteroides xylanisolvens isolate. *Appl. Environ. Microbiol.* **2012**, *78*, 528–535. https://doi.org/10.1128/AEM.06641-11.
4. Sóki, J.; Lang, U.; Schumacher, U.; Nagy, I.; Berényi, Á.; Fehér, T.; Burián, K.; Nagy, E. A novel Bacteroides metallo-β-lactamase (MBL) and its gene (crxA) in Bacteroides xylanisolvens revealed by genomic sequencing and functional analysis. *J. Antimicrob. Chemother.* **2022**, *77*, 1553–1556. https://doi.org/10.1093/jac/dkac088.
5. Chassard, C.; Delmas, E.; Lawson, P.A.; Bernalier-Donadille, A. Bacteroides xylanisolvens sp. nov., a xylan-degrading bacterium isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 1008–1013. https://doi.org/10.1099/ijs.0.65504-0.
6. Guerin, E.; Shkoporov, A.N.; Stockdale, S.R.; Comas, J.C.; Khokhlova, E.V.; Clooney, A.G.; Daly, K.M.; Draper, L.A.; Stephens, N.; Scholz, D.; et al. Isolation and characterisation of ΦcrAss002, a crAss-like phage from the human gut that infects Bacteroides xylanisolvens. *Microbiome* **2021**, *9*, 89. https://doi.org/10.1186/s40168-021-01036-7.
7. Shkoporov, A.N.; Clooney, A.G.; Sutton, T.D.S.; Ryan, F.J.; Daly, K.M.; Nolan, J.A.; McDonnell, S.A.; Khokhlova, E.V.; Draper, L.A.; Forde, A.; et al. The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* **2019**, *26*, 527–541.e5. https://doi.org/10.1016/j.chom.2019.09.009.
8. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. https://doi.org/10.1093/nar/gkg563.
9. Yang, M.; Derbyshire, M.K.; Yamashita, R.A.; Marchler-Bauer, A. NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. *Curr. Protoc. Bioinforma.* **2020**, *69*, e90. https://doi.org/10.1002/cpbi.90.
10. de Castro, E.; Sigrist, C.J.A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **2006**, *34*, W362–W365. https://doi.org/10.1093/nar/gkl124.
11. McGuffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics* **2000**, *16*, 404–405. https://doi.org/10.1093/bioinformatics/16.4.404.

12. Geourjon, C.; Deléage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **1995**, *11*, 681–684. https://doi.org/10.1093/bioinformatics/11.6.681.
13. Adamczak, R.; Porollo, A.; Meller, J. Accurate prediction of solvent accessibility using neural networks–based regression. *Proteins Struct. Funct. Bioinforma.* **2004**, *56*, 753–767. https://doi.org/10.1002/prot.20176.
14. Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **2018**, *430*, 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.
15. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. https://doi.org/10.1093/nar/gky427.
16. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291. https://doi.org/10.1107/S0021889892009944.
17. 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. https://doi.org/10.1093/nar/gky473.
18. Yu, C.-S.; Chen, Y.-C.; Lu, C.-H.; Hwang, J.-K. Prediction of protein subcellular localization. *Proteins* **2006**, *64*, 643–651. https://doi.org/10.1002/prot.21018.
19. Bhasin, M.; Garg, A.; Raghava, G.P.S. PSLpred: Prediction of subcellular localization of bacterial proteins. *Bioinformatics* **2005**, *21*, 2522–2524. https://doi.org/10.1093/bioinformatics/bti309.
20. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; 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. https://doi.org/10.1093/bioinformatics/btq249.
21. Savojardo, C.; Martelli, P.L.; Fariselli, P.; Profiti, G.; Casadio, R. BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* **2018**, *46*, W459–W466. https://doi.org/10.1093/nar/gky320.
22. Tusnády, G.E.; Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **2001**, *17*, 849–850. https://doi.org/10.1093/bioinformatics/17.9.849.
23. Dobson, L.; Reményi, I.; Tusnády, G.E. CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* **2015**, *43*, W408–W412. https://doi.org/10.1093/nar/gkv451.
24. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. https://doi.org/10.1093/nar/gkw1129.