Gene 626 (2017) 258–263

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Structural annotation of Beta-1,4-*N*-acetyl galactosaminyltransferase 1 (B4GALNT1) causing Hereditary Spastic Paraplegia 26

Rubina Dad[a], Uzma Malik[a], Aneela Javed[b], Berge A. Minassian[c], Muhammad Jawad Hassan[b,*]

[a] Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[b] Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[c] Division of Neurology, Department of Paediatrics, The Hospital For Sick Children, University of Toronto, Toronto, ON, Canada

ARTICLE INFO

Keywords:
Structural insights
B4GALNT1
Bioinformatics tools
HSP26
3-D structure of B4galnt1

ABSTRACT

Beta-1,4-*N*-acetyl galactosaminyltransferase 1, B4GALNT1, is a GM2/GD2 synthase, involved in the expression of glycosphingolipids (GSLs) containing sialic acid. Mutations in the gene *B4GALNT1* cause Hereditary Spastic Paraplegia 26 (HSP26). In present study we have made attempt to predict the potential structural of the human B4GALNT1 protein. The results illustrated that the amino acid sequences of B4GALNT1 are not 100% conserved among selected twenty species. One signal peptide and one transmembrane domain predicted in human wild type B4GALNT1 protein with aliphatic index of 92.76 and theoretical (iso-electric point) pI of 8.93. It was a kind of unstable protein with Grand average of hydropathicity (GRAVY) of −0.127. Various post-translational modifications were also predicted to exist in B4GALNT1 and predicted to interact with different proteins including ST8SIA5, SLC33A1, GLB1 and others. In the final round, reported missense mutations have shown the further decrease in stability of the protein. This *in-silico* analysis of B4GALNT1 protein will provide the basis for the further studies on structural variations and biological pathways involving B4GALNT1 in the HSP26.

1. Introduction

In the Central Nervous System of mammals, ganglioside structure and expression levels are highly conserved. The four predominant gangliosides are GM1, GD1a, GD1b and GT1b. Beta-1-4-*N*-acetyl galactosaminyltransferase 1 (GalNAc-T) B4GALNT1 is a GM2/GD2 synthase. It has role in the expression of GM2 and GD2 gangliosides which are glycosphingolipids (GSLs) containing sialic acid. They are located in the external leaflet of the plasma membrane and are abundant in the central nervous system. GalNAc-T is involved in the synthesis of asialo, a, b, c series gangliosides and their derivatives. It catalyzes the transfer of GalNAc into Lactosylceramide (LacCer), monosialoganglioside (GM3), disialo-ganglioside (GD2) or trisialo-ganglioside (GT3) (Nagata et al., 1994; Lloyd and Furukawa, 1998; Schnaar, 2010). Mutations of *B4GALNT1* lead to loss of GM3 synthase activity affecting a crucial branch point in the GSL biosynthetic pathway and signaling pathways in the nervous system (Simpson et al., 2004; Plomp and Willison, 2009).

Hereditary spastic paraplegia syndromes (HSPs) are grouped as genetically heterogeneous disorders. They also have a complex associated pathologies and overlapping phenotypes (Finsterer et al., 2012). The clinical and genetic diagnosis of HSPs is not an easy task because of the complex genetic background and overlapping phenotypes. The key feature of the syndrome is spasticity of lower limb due to dysfunctional pyramidal tract (Finsterer et al., 2012; Boukhris et al., 2013).

The genes known to underlie HSP encode molecules with a diverse range of biological functions in the cell; therefore, there are many defective pathways in the pathogenesis of the syndrome. One class of HSP is HSP 26 known to be caused by mutations in the *B4GALNT1* gene (Boukhris et al., 2013; Harlalka et al., 2013; Wakil et al., 2014).

Little literature is presented regarding the structure insights of the B4GALNT1 wild type and mutated protein (Yamaguchi et al., 2016). In the present article, we have identified the some structural and molecular characteristic features of B4GALTN1 wild type protein using bioinformatics tools and also predicted the impact of known missense mutations on the stability of the predicted 3-D structure of the wild type protein.

Abbreviations: B4GALNT1, Beta-1,4-*N*-acetyl galactosaminyltransferase 1; HSP26, Hereditary Spastic Paraplegia 26; GSLs, glycosphingolipids; GD2, disialo-ganglioside; GT3, trisialo-ganglioside; GRAVY, The Grand average of hydropathicity; TM, transmembrane; ST8SIA5, (st8 alpha-*n*-acetyl-neuraminide alpha-2,8-sialyltransferase 5); SLC33A1, (solute carrier family 33 (acetyl-coa transporter), member 1); GLB1, (galactosidase, beta-1)

* Corresponding author.
E-mail addresses: rdad.phdabs04asab@asab.nust.edu.pk (R. Dad), uzmamalik_86@yahoo.com (U. Malik), javedaneela@gmail.com (A. Javed), berge.minassian@sickkids.ca (B.A. Minassian), mjh@asab.nust.edu.pk (M.J. Hassan).

http://dx.doi.org/10.1016/j.gene.2017.05.041
Received 31 January 2017; Received in revised form 18 May 2017; Accepted 19 May 2017
Available online 20 May 2017

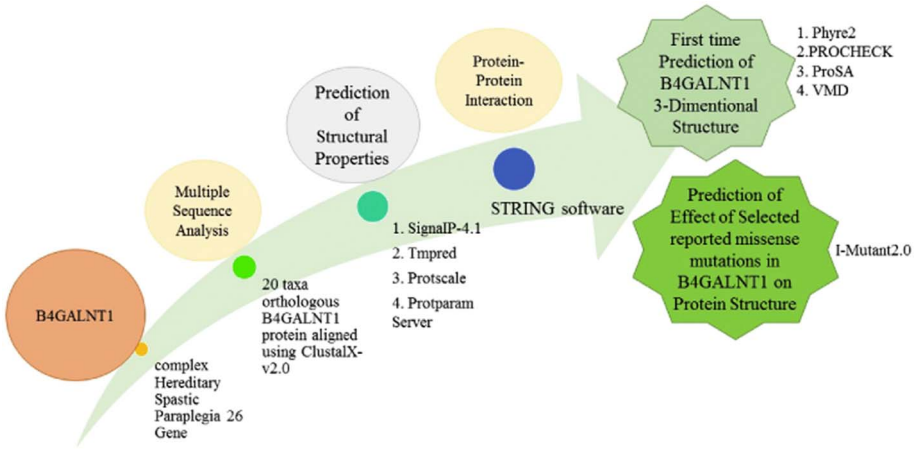

Fig. 1. Illustration of the methodology used for Prediction of Structural Insights of B4GALNT1.

2. Materials and methods

2.1. Multiple sequence alignment (MSA) of B4GALNT1 in different species

The FASTA sequence of B4GALNT1 wild type protein was retrieved from UniProt (http://www.uniprot.org/) using its primary accession number Q00973. BLASTp (Altschul et al., 1997) was utilized for searching homologous sequences with known function. Multiple sequence alignment (MSA) was achieved by aligning the human B4GALNT1 protein sequence with 20 orthologous protein sequences derived from various taxa using ClustalX-v2.0 (Larkin et al., 2007) with gap opening and gap extension multiple alignment penalties set to 10.00 and 0.20 respectively. The names of organisms and their UniProt ID have been tabulated in Supplementary Table 1 (S_Table1).

2.2. Computational analysis of B4GALNT1

For the prediction of human wild type B4GALNT1 protein signal peptide, SignalP-4.1 Server of Center for Biological Sequence Analysis (CBS) platform (http://www.cbs.dtu.dk/services/SignalP/) was used (Petersen et al., 2011). For the prediction of transmembrane region and associated orientations of B4GALNT1 wild type protein, TMpred (http://embnet.vital-it.ch/software/TMPRED_form.html) program was used (Hofmann and Stoffel, 1993). Expasy platform with its server Protscale (http://web.expasy.org/ protscale was used to predict the accessibility, bulkiness, flexibility, hydrophilicity, mutability polarity, and refractivity of B4GALNT1 wild type protein. In order to predict post-translational modifications like glycosylation sites, mannosylation sites, kinase specific phosphorylation sites and other phosphorylation sites, for B4GALNT1 wild type protein, server (http://www.cbs.dtu.dk/services/) was used. For validation of some physiological parameters of B4GALNT1, Protparam Server from expasy platform (http://web.expasy.org/protparam/) was used. For analysis of protein-protein interaction of B4GALNT1 with other proteins, Search Tool for Retrieval of Interacting Genes and proteins (STRING), was used, which is a database for interaction analysis of predicted and known proteins (Szklarczyk et al., 2011).

2.3. Prediction of secondary and tertiary structure of B4GALNT1

Lastly, protein data bank (PDB) (Berman et al., 2000) was searched using BLASTp for the existence of any suitable entry (having > 37% similarity) as a template for modeling the structure of B4GALNT1. As the results did not return any suitable template, Phyre2.0 (Kelley et al., 2015) was used for the automated prediction of B4GALNT1 structure. The 3-D model generated by Phyre2 was subjected to energy minimization using UCSF Chimera (Pettersen et al., 2004), a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. Once minimized, the model was analyzed for its stereochemical quality by PROCHECK (Laskowski et al., 1993) and ProSA (Wiederstein and Sippl, 2007), both of which are well-established structure analysis tools for modeled proteins. PROCHECK provides the distribution of residues in different regions based on resolution of N = 2.0 Å and R-factor b = 20.0. ProSA returns a z-score that calculates the deviation of the total energy of the structure in comparison to an energy distribution derived from random conformations to exhibit the model quality. The negative z-score of model depicts its good quality. VMD program (Humphrey et al., 1996) was then used for the visualization of protein structure. The point mutations selected in this article (one unpublished mutation from our group and four already published missense mutations from (Boukhris et al., 2013; Harlalka et al., 2013) given in Table 2), were then inserted in the 3-D structure of wild type B4GALNT1 protein using Chimera. The changes in protein stability upon single point mutations were predicted by I-Mutant2.0 (Capriotti et al., 2005). Finally, the individual mutated structures of B4GALNT1 were superimposed with the wild type B4GALNT1 using VMD to visualize the structural changes inflicted by the mutations, methodology illustrated in the Fig. 1.

3. Results

The sequence of the human B4GALNT1 protein is given below.

MWLGRRALCALVLLLACASLGLLYASTRDAPGLRLPLAPWAPPQSPRRPELPDLAPEPRYAHIPVRIKEQVVGLLAWNNCSCESSGGGLPLPFQKQVRAIDLTKAFDPAELRAASATREQEFQAFLSRSQSPADQLLIAPANSPLQYPLQGVEVQPLRSILVPGLSLQAASGQEVYQVNLTASLGTWDVAGEVTGVTLTGEGQADLTLVSPGLDQLNRQLQLVTYSSRSYQTNTADTVRFSTEGHEAAFTIRIRHPPNPRLYPPGSLPQGAQYNISALVTIATKTFLRYDRLRALITSIRRFYPTVTVVIADDSDKPERVSGPYVEHYLMPFGKGWFAGRNLAVSQVTTKYVLWVDDDFVFTARTRLERLVDVLERTPLDLVGGAVREISGFATTYRQLLSVEPGAPGLGNCLRQRRGFHHELVGFPGCVVTDGVVNFFLARTDKVREVGFDPRLSRVAHLEFFLDGLGSLRVGSCSDVVVDHASKLKLPWTSRDAGAETYARYRYPGSLDESQMAKHRLLFFKHRLQCMTSQ.

The UniProt ID of human B4GALNT1 is Q00973 which was used to retrieve its sequence for performing systematic sequence analysis using various tools. The multiple sequence alignment of B4GALNT1 from 20 different taxa including human revealed 203 perfect, 121 high and 36 weakly conserved residue groups. The conserved residues and their specific positions are depicted in the multiple sequence alignment presented in Supplementary Fig. 1. Moreover, the point mutations selected in the study have also been highlighted in the figure.