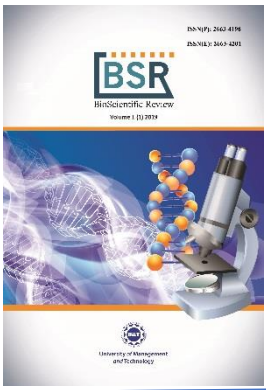

BioScientific Review (BSR)
Volume 2, Issue 2, 2020
ISSN(P): 2663-4198 ISSN(E): 2663-4201
Journal DOI: https://doi.org/10.32350/BSR
Issue DOI: https://doi.org/10.32350/BSR.0202
Homepage: https://journals.umt.edu.pk/index.php/BSR

Journal QR Code:

Article:

A Comprehensive *In Silico* Analysis of Deleterious SNPs of Paraplegin Protein Associated with Hereditary Spastic Paraplegia through Mitochondrial Dysfunction

Author(s):

Ammara Akhtar
Sobia Nazir Choudhry
Rana Muhammad Mateen
Mureed Hussain

Article DOI: https://doi.org/10.32350/BSR.0202.01

Article QR Code:

Ammara Akhtar

To cite this article: Akhtar A, Choudhry SN, Mateen RM, Hussain M. A comprehensive *In Silico* analysis of deleterious SNPs of paraplegin protein associated with hereditary spastic paraplegia through Mitochondrial dysfunction. *BioSci Rev*. 2020;2(2): 01–14.
Crossref

A publication of the
Department of Life Sciences, School of Science
University of Management and Technology, Lahore, Pakistan

A Comprehensive *In Silico* Analysis of Deleterious SNPs of Paraplegin Protein Associated with Hereditary Spastic Paraplegia through Mitochondrial Dysfunction

Ammara Akhtar, Sobia Nazir Choudhry, Rana Muhammad Mateen, Mureed Hussain*

Department of Life Sciences, University of Management and Technology, Lahore, Pakistan

*Corresponding author: mureed.hussain@umt.edu.pk

Abstract

Hereditary spastic paraplegia (HSP) is a heterogenous neurological disorder primarily associated with progressive spasticity. Paraplegin is a mitochondrial protein and mutations in this protein can lead to HSP. In this study, *in silico* analysis was carried out to identify the pathogenic variants of SPG7 (paraplegin protein). To find novel pathogenic mutations, missense and splicing variants were collected from gnomAD database and passed through a detailed and stringent analysis with the help of a variety of bioinformatic tools. The list of mutations was examined and compared in ClinVar. Altogether, 14 missense mutations and 18 splicing mutations were obtained and these mutations were predicted to have the potential of disrupting the normal structural and functional properties of paraplegin protein.

Keyword: *in silico* analysis, hereditary spastic paraplegia (HSP), paraplegin, SNPs, SPG7

1. Introduction

Hereditary spastic paraplegia (HSP) is a genetic neurodegenerative syndrome which is also known as Strumpell-Lorrain disease. It is described on the basis of a major hallmark, that is, axonal degeneration, which leads to abnormalities in the corticospinal region. The corticospinal dysfunction ultimately causes weakness and spasticity in the lower limb area [1, 2].

The heterogeneity associated with the clinical manifestations of HSP reflects that some proteins of HSP contribute in processes or pathways at cellular and molecular level, so the abnormalities or mutations may spur defects in multiple processes. Since HSP genes are also associated with mitochondrial functions, hence mitochondrial abnormalities have been implicated in the development of neurological disorders [3].

Paraplegin (SPG7) shows the autosomal recessive form of HSP. It is a metalloprotease protein with 17 exons and 795 amino acids which have an AAA domain [4]. The protein has a prevalence rate of 2–6/100,000 [5]. It is primarily present with the homologous AFG3L2 protein, which is a protease in nature and is found in mitochondria. The main functions performed by paraplegin include controlling the quality of protein, enzyme processing, maturation of mitochondrial protein and it also plays a role in ribosomal assembly. Mutation in paraplegin can lead to defects in the oxidative phosphorylation process of mitochondria and it has been observed in the muscle biopsies of HSP patients [4, 6].

Over the past many years, massive efforts and high-throughput methods have been formulated to unravel the

complicated heterogeneity, mutations, and functional variants associated with different diseases. The need for well-defined and organized strategies yielding tremendous genetic information is imperative in this research field. The identification of novel pathogenic variants is a serious challenge with a wide variety of variants being available. It obliges the scientists to identify potentially damaging variants keeping in view all available evidences. For this purpose, a variety of computer-aided bioinformatic tools are available that have the potential to predict the likely pathogenic ability of the variants [7].

To find out and predict whether an SNP will lead to disease or not, computational tools designed for bioinformatic analysis can be used. Many projects have been carried out to analyze and predict the effect of various genetic variations using various computational tools to understand the genetic basis of different diseases. According to an estimate, there are more than four million proven single nucleotide polymorphism (SNPs) in the human genome. The vast volume of the available genetic variation aids in analyzing the basis of different diseases through bioinformatic tools. These tools are a great source of functional analysis of human genome [8].

The current study was carried out to find novel mutations of the gene SPG7 related to HSP through mitochondrial dysfunction with the help of *in silico* tools.

2. Materials and Methods

2.1. Data Retrieval

To study and identify mutations that lead to HSP, the variants of SPG7 gene associated with HSP through mitochondrial dysfunction were retrieved via *in silico* analysis using gnomAD (Genome Aggregation Database). Protein name was used to obtain the list of variants. Different filters such as the loss of function, non-synonymous, missense and heterozygous were applied on the gnomAD site [9] to get the desired variant list. The variant list was then downloaded in VCF format. The protein sequence of paraplegin protein was retrieved from NCBI.

2.2. Variant Selection

The variants obtained from gnomAD were filtered further on the basis of their allelic frequency. To select a cut-off value, the allelic frequencies of 100 pathogenic mutations related to the autosomal recessive form of HSP were collected from Variation Viewer of NCBI. An allelic frequency of <0.002 was used to filter out further variants in gnomAD. The variants below 0.002 were subjected to CADD analysis.

2.3. CADD Analysis

CADD [10] was used for scoring the variants based on their deleterious mutations. Various filters were applied to get the desired variants. A cut-off value was applied for C-score (PHRED score >15) and variants were further selected and analyzed with a variety of bioinformatic tools. The analysis was aimed at finding pathogenic mutations of paraplegin protein in the missense and splicing sites.

2.4. Missense Variant Analysis

After C-scoring, to analyze the missense variants, a filter was applied in the CADD file to retrieve only the missense variants of paraplegin protein that were later analyzed with PhD-SNP^{g} [11], PredictSNP2 [12], UMD-predictor [13], SNP&GO [14] and PROVEAN [15] for

further validation of results. The variant list was uploaded in either CSV or VCF file format on the online server and each tool showed the output based on its algorithmic criteria. The pathogenic variants obtained were further compared in ClinVar [16] to find out whether they had been reported in any previous article or research work.

2.5. 3D Modelling of Paraplegin

To check the effect of certain mutations on protein structure, function and stability, the modelling of protein was performed through I-TASSER. Structural models of protein were obtained through I-TASSER. Based on C-score and RMSD, the model with a high confidence level was selected. C-score basically shows the confidence level used for measuring the quality of all predicted structural models of protein [17]. For visualization purpose, Chimera was used.

2.6. Stability Analysis

To analyze the effect of mutations at amino acid level on the stability and functional properties of paraplegin protein, I-Stable was used [18]. Using this tool, protein sequence and amino acid substitution were added one by one for each variant as input format to subsequently study the positive or negative stabilizing effect of each variant on the protein. Pathogenic mutations can destabilize the normal structural and/or functional activity of the protein.

2.7. Splicing Variant Analysis

The analysis of splicing SNPs of paraplegin protein was performed using several computer-aided tools (Human Splice Finder, Spice and Spliceman). Mutations in the splicing regions have the potential to break or add a new splice site, thus disrupting the normal splicing process. To find novel mutations, the mutations obtained in this study were checked with already reported mutations in ClinVar.

2.8. Clashes and Contacts

Clashes / contacts were also analyzed based on VDW radii to check if the introduced mutation showed an unfavorable interaction with nearby residues. Contacts tells that if there is any direct interaction, either polar or non-polar, between the residues which may not be necessarily favorable [19].

2.9. Hydrophobicity Analysis

Hydrophobicity analysis was performed according to kdHydrophobicity scale using Chimera [20].

3. Results

3.1. 3D structure of Paraplegin

Structural models of paraplegin were obtained through I-TASSER. Among the five models produced through I-TASSER, the first was selected as the best model of spartin protein (Figure 1). For SPG7, the estimated C-score was -2.06, estimated TM-score was 0.47±0.15 and estimated RMSD was 13.5±4.0Å

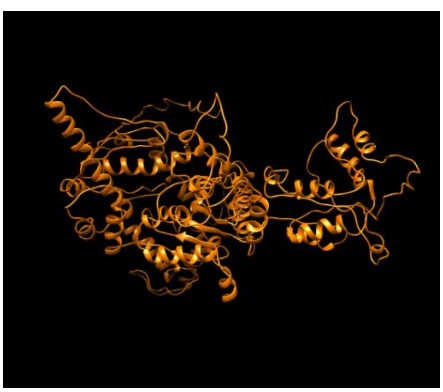

Figure 1. 3D model of paraplegin protein generated through I-TASSER

3.2. Missense Variant Analysis of SPG7

Variants of SPG7 gene were retrieved through gnomAD. A total of 2021 variants were obtained. An allelic frequency filter (< 0.002) was applied and a total of 1977 variants were further uploaded in CADD. After CADD analysis, missense and PHRED score filter (≥15) were applied and 463 variants were obtained.

3.2.1. Deleterious missense SNPs identified through different bioinformatic tools. The obtained variants were further analyzed using different bioinformatic tools including SNPs&GO, PhD-SNPg, PREDICTSNP2, PROVEAN and UMD-Predictor. A specific set of filters including deleterious, pathogenic, and probably pathogenic filters was applied to obtain highly pathogenic variants among a variety of variants. Thirty-two variants were obtained. The purpose of using these filters was to obtain those variants that were predicted to have an effect on the activity of the gene using all five *in silico* tools (Table 1).

3.2.2. Stability predictions. The variants were further analyzed through I-Stable to check the effect of these mutations on the stability of protein. These mutations have the potential to cause destabilization in the normal structural and functional properties of the protein. The I-Stable tool shows whether the introduced mutation will decrease the stability of the protein or not. The filter of "decrease" was applied and only the variants with a high potential to decrease the stability of protein were selected. Nine variants out of thirty-two listed in the variant list were finally obtained (Table 2, Figure 2).

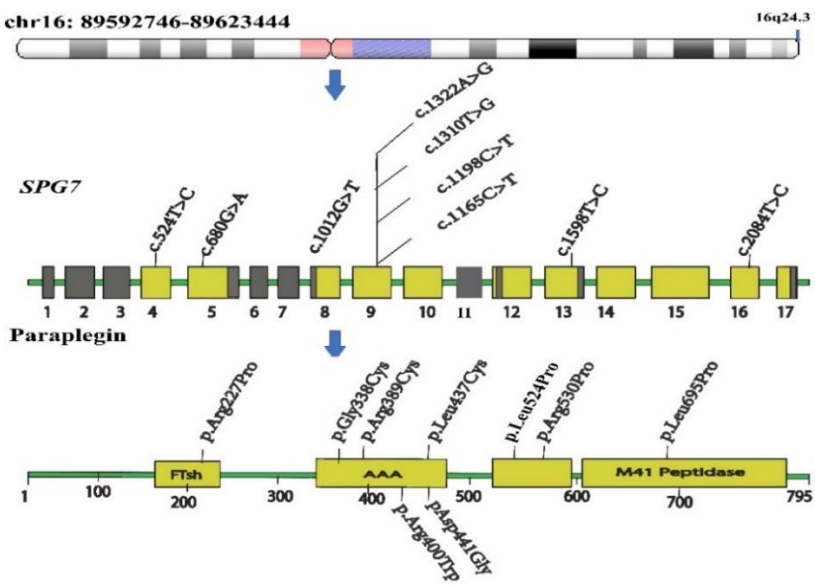

Figure 2. List of mutations of paraplegin from cDNA (SNV) to protein level (amino acid substitution). Gene position on the chromosome is 16q24.3. SPG7 gene constitutes three main domains: FTsh domain (144-237), AAA domain (341-481) and M41 domain (562-745)

Table 1. List of High-Risk Variants of Paraplegin (*SPG7*) after Applying the Cut-off Values (Phdphd-Snp^g ≥ 0.9, PROVEAN ≤ -4.00, Predict-SNP2 ≥ 0.8, SNP&Gos ≥ 0.6, UMD-Predictor ≥ 80)
(Abbreviations: D = Deleterious, Dis = Disease, P = Pathogenic, Pre = Prediction, S = Score)

Chr	Pos	R	A	Substitution	PhD-SNPg		PROVEAN		PredictSNP2		SNP&GO		UMD		PHRED
					Pre	S	Pre	S	Pre	S	Pre	S	Pre	S	
16	89620349	T	C	p.Leu695Pro*	P	0.988	D	-6.25	D	1	Dis	0.853	84	P	17.79
16	89620279	G	A	p.Gly672Arg*	P	0.996	D	-7.49	D	1	Dis	0.849	93	P	18.39
16	89620261	G	C	p.Gly666Arg*	P	0.982	D	-7.49	D	1	Dis	0.863	99	P	18.51
16	89620241	A	G	p.Tyr659Cys	P	0.97	D	-7.82	D	1	Dis	0.843	90	P	18.56
16	89614447	T	C	p.Leu530Pro*	P	0.996	D	-6.44	D	1	Dis	0.812	84	P	22.7
16	89613160	G	A	p.Gly515Glu	P	0.977	D	-7.78	D	1	Dis	0.838	100	P	22.8
16	89613142	T	G	p.Leu509Arg	P	0.997	D	-5.74	D	1	Dis	0.8	93	P	22.9
16	89611095	C	T	p.Thr455Met	P	0.988	D	-5.58	D	1	Dis	0.806	93	P	23.5
16	89599042	A	G	p.Asp441Gly*	P	0.993	D	-6.58	D	1	Dis	0.825	99	P	23.7
16	89599030	T	G	p.Leu437Arg*	P	0.994	D	-6	D	1	Dis	0.853	93	P	23.8
16	89598969	C	T	p.Arg417Cys	P	0.974	D	-7.79	D	1	Dis	0.824	99	P	24.1
16	89598937	A	G	p.Tyr406Cys	P	0.994	D	-8.34	D	1	Dis	0.867	93	P	24.2
16	89598943	A	G	p.Asp408Gly	P	0.995	D	-6.6	D	1	Dis	0.859	100	P	24.2
16	89598918	C	T	p.Arg400Trp	P	0.954	D	-7.23	D	1	Dis	0.813	96	P	24.3
16	89598891	C	T	p.Arg391Trp*	P	0.993	D	-7.77	D	1	Dis	0.841	93	P	24.5
16	89598885	C	T	p.Arg389Cys*	P	0.991	D	-7.7	D	1	Dis	0.842	100	P	24.5
16	89598382	G	A	p.Cys353Tyr	P	0.996	D	-10.09	D	1	Dis	0.856	100	P	25

Chr	Pos	R	A	Substitution	PhD-SNPg		PROVEAN		PredictSNP2		SNP&GO		UMD		PHRED
					Pre	S	Pre	S	Pre	S	Pre	S	Pre	S	
16	89598369	G	A	p.Gly349Arg	P	0.999	D	-7.62	D	1	Dis	0.833	100	P	25.2
16	89598370	G	T	p.Gly349Val	P	0.991	D	-8.54	D	1	Dis	0.848	100	P	25.2
16	89598355	G	A	p.Gly344Asp	P	0.989	D	-6.64	D	1	Dis	0.845	90	P	25.3
16	89598336	G	T	p.Gly338Cys	P	0.992	D	-8.47	D	1	Dis	0.834	93	P	25.4
16	89592798	G	C	p.Arg227Pro	P	0.987	D	-6.21	D	1	Dis	0.847	100	P	27.1
16	89590561	T	C	p.Leu175Pro*	P	0.995	D	-6.61	D	1	Dis	0.829	84	P	28.4

Asterisk (*) sign with amino acid substitutions indicates that the given mutations have already been reported in ClinVar, while all other mutations are novel.

Table 2. High-Risk Pathogenic Mutation Identified through *in Silico* Tools Causing Decrease in The Stability of Paraplegin (*SPG7*)

Chr	Pos	Ref	Alt	Substitution	i-Mutant 2.0	DDG	MUpro	Conf. Score	iStable
16	89598336	G	T	p.Gly338Cys	Decrease	-1.36	Decrease	-0.0666	Decrease
16	89599030	T	G	p.Leu437Arg	Decrease	-1.05	Decrease	-0.0123	Decrease
16	89598918	C	T	p.Arg400Trp*	Decrease	-0.39	Decrease	-0.77	Decrease
16	89599042	A	G	p.Asp441Gly	Decrease	-0.59	Decrease	-0.105	Decrease
16	89598885	C	T	p.Arg389Cys	Decrease	-1.01	Decrease	-0.7189	Decrease
16	89592798	G	C	p.Arg227Pro	Decrease	-0.65	Decrease	-0.4989	Decrease
16	89620349	T	C	p.Leu695Pro*	Decrease	-1.64	Decrease	-0.9667	Decrease
16	89614447	T	C	p.Leu530Pro	Decrease	-1.4	Decrease	-1	Decrease
16	89590561	T	C	p.Leu175Pro*	Decrease	-1.81	Decrease	-1	Decrease

Asterisk (*) indicates the mutations already reported in the ClinVar, all other mutations are novel.

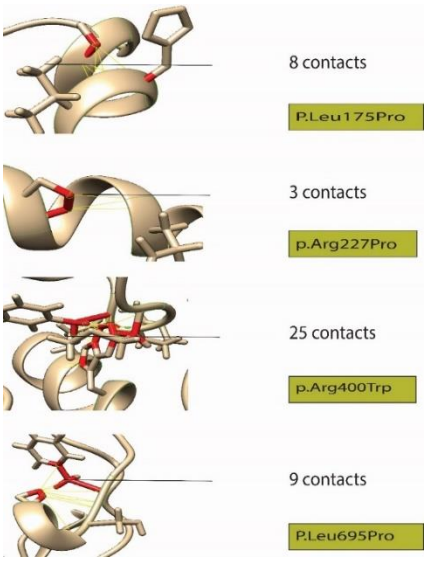

Figure 3. Analysis of clashes / contacts found as a result of amino acid substitiution of paraplegin (SPG7)

3.3. Clashes and Contacts

Clashes and contacts were analyzed to study the interaction of mutant residue with the neighboring residues using Chimera. Four missense mutations exhibited interaction with neighboring residues. These interactions have the potential to destabilize the structure of protein, thus leading to the disruption of the normal protein structure (Figure 3).

3.4. Hydrophobicity Analysis

Chimera was used to carry out hydrophobicity analysis. In this process, amino acid residues were given characteristic features and properties according to kdHydrophobicity. Values were assigned to each amino acid residue based on the hydrophobicity scale generated by Kyte and Doolittle. A color was allocated to each amino acid according to its hydrophilic or hydrophobic nature (blue color: most hydrophilic, white: neutral, orange / red: most hydrophobic), along with a value according to the hydropathy scale (hydrophobicity value: more positive, hydrophobic: more negative, hydrophilic) (Table 3, Figure 4)

3.5. Splicing SNPs Identified through HSF, SPICE and SPLICEMAN

The variants were retrieved through gnomAD and a total of 2403 variants were obtained. After applying allelic frequency and PASS filter (< 0.002), 2351 variants were left which were analyzed through CADD score. CADD score filter (≥ 15) was applied and a total of 148 variants were obtained which were further analyzed by applying various filters (Canonical-Splice, Splice donor and Splice acceptor). Ultimately, 23 variants were obtained at the end. The variants were further analyzed through different bioinformatic tools namely Human Splice Finder, Spicev2.15 and Spliceman to check the effect of the mutation on the splicing regions of paraplegin (Table 4).

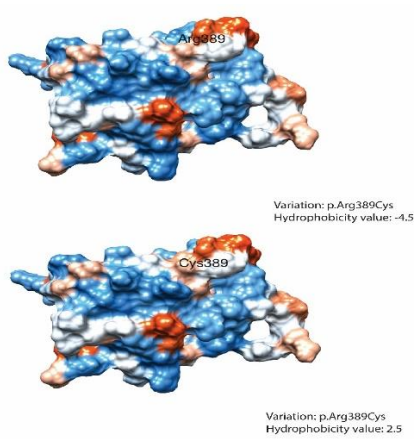

Figure 4. An example of hydrophobicity analysis of AAS with color depiction ranging from blue to orange to red, exhibiting the effect of mutation on the hydropathy properties of protein

Table 3. Hydrophobicity Values of AAS of Paraplegin (*SPG7*) According to Kyte-Doolittle Hydrophobicity Scale

Chr	Pos	Ref	Alt	AAS	Hydrophobicity Value	
					Wild	Mutant
16	89590561	T	C	p.Leu175Pro	3.8	-1.6
16	89592798	G	C	p.Arg227Pro	-4.5	-1.6
16	89598336	G	T	p.Gly338Cys	-0.4	2.5
16	89598885	C	T	p.Arg389Cys	-4.5	2.5
16	89598918	C	T	p.Arg400Trp	-4.5	-0.9
16	89599030	T	G	p.Leu437Arg	3.8	-4.5
16	89599042	A	G	p.Asp441Gly	-3.5	-0.4
16	89614447	T	C	p.Leu530Pro	-4.5	-1.6
16	89620349	T	C	p.Leu695Pro	-4.5	-1.6

Table 4. Analysis of The Effect of Mutations on the Splicing Sites of Paraplegin (*SPG7*) through Various Bioinformatic Tools

Chr	Pos	Ref	Alt	Consequence	Spliceman		SPICE						HSF	
					Splicemanl1 distance	Ranking (L1)	SSF_wt	SSF_mut	MES_wt	MES_mut	SPiCEprobability	SPiCEinter_2thr	WT score	CV Variation (%)
16	89595988	G	C	c.861+1G>C	37791	81%	74.43	0	7.4	-0.87	1	high	82.4	-32.56
16	89597089	A	G	c.862-2A>G	35137	68%	90.22	0	9.4	1.44	1	high	89.5	-32.35
16	89590413	G	C	c.377-1G>C	33813	61%	78.2	0	5.04	-3.03	1	high	83.09	-2.21
16	89597217	G	A	c.987+1G>A	32812	56%	78.98	0	4.51	-3.67	1	high	83.39	-32.19
16	89613064	A	G	c.1450-2A>G*	35709	71%	85.92	0	6.82	-1.14	1	high	86.98	-33.28
16	89592877	G	T	c.758+1G>T	34295	64%	76.63	0	6.93	-1.57	1	high	66.43	-18.53

16	89598311	G	A	c.988-1G>A*	34233	64%	80.84	0	4.97	-3.78	1	high	84.68	-34.19
16	89614410	G	A	c.1553-1G>A	33968	62%	87.91	0	5.74	-3.01	1	high	54.55	+22.31
16	89619386	G	A	c.1780-1G>A	34955	67%	80.8	0	6.36	-2.39	1	high	46.24	+62.59
16	89598310	A	G	c.988-2A>G	35306	69%	80.84	0	4.97	-2.99	1	high	54.55	+22.31
16	89614409	A	G	c.15532A>G*	35112	68%	87.91	0	5.74	-2.22	1	high	88.53	-32.7
16	89613169	G	A	c.1552+2dupT	30624	45%	77.38	0	8.61	0.42	1	high	54.55	+14.25
16	89577001	G	A	c.286+1G>A	31794	51%	85	0	7.76	-0.42	1	high	89.21	-30.08
16	89613169	G	T	c.1552+2dupT	30624	45%	77.38	0	8.61	0.1	1	high	84.69	-31.69
16	89619545	T	A	c.1936+6_1936+8delGAG	32723	56%	84.36	0	8.3	0.11	1	high	49.15	+58.88
16	89620200	A	G	c.19372A>G*	37197	79%	84.31	0	7.25	-0.7	1	high	92.49	-31.29
16	89576897	G	T	c.184-1G>T	37529	80%	95.65	0	11.88	3.28	1	high	93.93	-30.82
16	89576896	A	T	c.184-2A>T	39870	92%	95.65	0	11.88	3.51	1	high	93.93	-30.82
16	89595883	A	G	c.759-2A>G	35041	68%	87.01	0	9.51	1.55	1	high	89.54	-32.33
16	89579446	G	A	c.376+1G>A	34153	63%	76.56	0	7.84	-0.34	1	high	78.08	-34.38
16	89579446	G	T	c.376+1G>T*	34795	66%	76.56	0	7.84	-0.66	1	high	41.15	+65.22
16	89579446	G	C	c.376+1G>C	30178	43%	76.56	0	7.84	-0.43	1	high	78.08	-34.38
16	89592735	A	G	c.619-2A>G	37927	82%	86.12	0	9.68	1.73	1	high	89.4	-3.45

The threshold value is set as 65 in HSF. For a mutation, if the wild type score is above the set value and the variation score is below -10%, then it will break the splice site. If the wild type score is below the set value and variation score is above +10%, the mutation will produce a new splice site. Asterisk (*) sign with amino acid substitutions indicates that the given mutations have already been reported in ClinVar, while all other mutations are novel.

4. Discussion

HSP is clinically and genetically heterogeneous group of disorders associated with lower limb weakness followed by progressive spasticity and with the passage of time, the condition may worsen. Currently, there is no particularly effective therapy for HSP, although there are drugs (anti-spastic effect) such as diazepam and baclofen that play an important role in lessening the complications associated with the disease (including pain and fractures), as well as improving the lives of patients [21, 22]. Today, there is a dire need to understand the pathophysiology associated with this disease as well as a detailed analysis of the genetic mutation leading to HSP.

The importance of identifying novel variants in the field of genetic testing related to a particular disease is increasing day by day. It will ultimately bring prominent changes by enhancing the sensitivity of the genetic testing process. Although the most commonly used approach for this purpose is gene sequencing; however, due to the vast variety of newly found variants it is really difficult to analyze the pathogenicity of each variant. Also, all variants do not alter the functional properties of the protein [23, 24].

In this study, a list of variants of SPG7 gene associated with HSP was retrieved from gnomAD and was analyzed to find novel pathogenic missense and splicing variants of paraplegin protein. The allelic frequency filter was applied and gnomAD file was then uploaded in the CADD server. After CADD analysis, missense variants and splicing variants were sorted out by applying filters. The results for missense variants were further confirmed with several renowned bioinformatic tools used for missense variant analysis (SNP&GO, PHD-SNPg, PredictSNP2, UMD-Predictor and PROVEAN). After a detailed analysis of missense variants aimed at analyzing their pathogenicity, a total of 9 mutations were finally obtained which were predicted to have highly pathogenic effects by all *in silico* tools used in the analysis. The mutations of SPG7 were analyzed using NM_003119.4.

In this work, to analyze the effect of missense mutations on the stability of protein related to HSP through mitochondrial dysfunction, I-Stable tool was used. Clashes / contacts were further analyzed in Chimera to see the type of interaction between the substituted residue (after mutation) and the neighboring residues (Figure 3).

To analyze the impact of mutation on the splicing region of these mitochondrial genes, bioinformatic tools namely Spliceman, Human Splice Finder and SPICE were used. After thorough analysis, 25 mutations were identified in case of SPG7 through CADD analysis. All these mutations have the potential to produce a deleterious effect in case of the autosomal recessive form of HSP through mitochondrial dysfunction.

There are various empirical techniques used to analyze the distribution of all twenty amino acids in different conformation for the formation of the three-dimensional structure of a specific protein. By devising a scheme, the probability of secondary structures of protein to adopt a general shape can be evaluated. Amino acids can be either hydrophilic and hydrophobic based on their particular chemical characteristics, for example, alanine (A), valine (V), leucine (L), proline (P), phenylalanine (F) and cysteine (Cys) are considered to be hydrophobic, while lysine and arginine are hydrophilic in nature, both

consisting of positive charge at the neutral pH [25]. SOAP is an online computer program which is written in C-language and it allocates a proper hydropathy value to every residue in a specific amino acid sequence. According to the scale, as the value increases, the more hydrophobic becomes the residue [26].

Pathogenic mutations found in the splicing region have the potential to disrupt the normal splicing process of the protein. They may be present either in the exons or the introns. The mutations either break or add a new splicing site or may turn on the cryptic ones, which ultimately affects the protein. They may further bind with the enhancers or silencers involved in the splicing process and also bring alteration in the structure of mRNA [27]. In this study, 23 splicing mutations were predicted to have a pathogenic effect on the normal splicing process of the paraplegin protein. Out of these 23, 18 mutations have not been reported in any previous work for their associated pathogenicity in causing HSP.

In this study, after an elaborated series of analysis, each variant was further analyzed in ClinVar to check the novelty of the work. This study found a total of 14 missense mutations and 18 splicing mutations which were considered to be highly pathogenic. Their pathogenicity was not reported previously in any kind of scientific work. Thus, they are considered novel.

5. Conclusion

HSP, a neurodegenerative disease, can be characterized with a number of disorders in which lower limb complexity associated with progressive spasticity is the prefatory symptom. The proteins associated with mitochondrial dysfunction are directly involved in causing HSP. In this study, mutation analysis of paraplegin protein (SPG7) was carried out through bioinformatic tools. Different bioinformatic tools were utilized to analyze the effect of missense and splicing mutations on the structure and stability of protein. The results showed that 14 missense mutations and 18 splicing mutations have the potential to cause the autosomal recessive form of HSP and its associated complications. The mutations reported in this work can be further analyzed through *in vitro* analysis.

References

[1] Depienne C, Stevanin G, Brice A, Durr A. Hereditary spastic paraplegias: an update. *Curr Opin Neurol*. 2007;20(6): 674–680.

[2] Giudice TL, Lombardi F, Santorelli FM, Kawarai T, Orlacchio A. Hereditary spastic paraplegia: clinical-genetic characteristics and evolving molecular mechanisms. *Exp Neurol*. 2014;261: 518–539.

[3] Shribman S, Reid E, Crosby AH, Houlden H, Warner TT. Hereditary spastic paraplegia: from diagnosis to emerging therapeutic approaches. *Lancet Neurol*. 2019;18(12): 1136–1146.

[4] Wali G, Kumar KR, Liyanage E, Davis RL, Mackay A-S, Sue CM. Mitochondrial function in hereditary spastic paraplegia: deficits in SPG7 but not SPAST patient-derived stem cells [published online ahead of print August 20, 2020]. *Front Neurosci*. https://www.frontiersin.org/articles/10.3389/fnins.2020.00820/full

[5] Zhang X, Zhang L, Wu Y, et al. Identification of novel compound

heterozygous SPG7 mutations-related hereditary spastic paraplegia in a Chinese family: a case report. *BMC Neurol.* 2018;18(1): 196.

[6] Koppen M, Metodiev MD, Casari G, Rugarli EI, Langer T. Variable and tissue-specific subunit composition of mitochondrial m-AAA protease complexes linked to hereditary spastic paraplegia. *Mol Cell Biol.* 2007;27(2): 758–767.

[7] Flanagan SE, Patch A-M, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers*. 2010;14(4): 533–537.

[8] Srinivasan E, Natarajan N, Rajasekaran R. TTRMDB: a database for structural and functional analysis on the impact of SNPs over transthyretin (TTR) using bioinformatic tools [published online ahead of print May 25, 2020]. *Comput Biol Chem.* 2020;87: 107290.

[9] Karczewski K, Francioli L. *The genome aggregation database (gnomAD)*. MacArthur Lab Web site.https://macarthurlab.org/2017/02/27/thegenomeaggregationdatabase-gnomad/ Updated February 27, 2017.

[10] Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47 (D1): D886–D894.

[11] Capriotti E, Fariselli P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* 2017;45(W1): W247–W252.

[12] Bendl J, Musil M, Štourač J, Zendulka J, Damborský J, Brezovský J. Predict SNP2: a unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS Comput Biol.* 2016;12(5): e1004962.

[13] Salgado D, Desvignes JP, Rai G, et al. UMD-predictor: a high-throughput sequencing compliant system for pathogenicity prediction of any human cDNA substitution. *Hum Mutat*. 2016;37(5): 439–446.

[14] Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs & GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics.* 2013;14(3): S6.

[15] Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *BMC Bioinf.* 2015;31(16): 2745–2747.

[16] Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2017;46(D1): D1062–D1067.

[17] Yang J, Zhang, Y. Protein structure and function prediction using I-TASSER. *Curr Protoc Bioinf.* 2015;52(1): 5.8.1–5.8.15. https://pubmed.ncbi.nlm.nih.gov/26678386/

[18] Chen CW, Lin J, Chu YW. iStable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinf.* 2013;14: S5.

[19] Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13): 1605–1612.

[20] Yang Z, Lasker K, Schneidman D-D, et al. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J Struct Biol.* 2012;179(3): 269–278.

[21] Finsterer J, Löscher W, Quasthoff S, Wanschitz J, Auer M-G, Stevanin G. Hereditary spastic paraplegias with autosomal dominant, recessive, X-linked, or maternal trait of inheritance. *J Neurol Sci.* 2012;318(1-2): 1–18.

[22] Klebe S, Stevanin G, Depienne C. Clinical and genetic heterogeneity in hereditary spastic paraplegias: from SPG1 to SPG72 and still counting. *Rev Neurol.* 2015;171(6-7): 505–530.

[23] Campuzano O, Allegue C, Fernandez A, Iglesias A, Brugada R. Determining the pathogenicity of genetic variants associated with cardiac channelopathies. *Sci Rep.* 2015;5: 7953.

[24] Young DL, Fields S. The role of functional data in interpreting the effects of genetic variation. *Mol Biol Cell*. 2015;26(22): 3904–3908.

[25] Scheiner S, Kar T, Pattanayak J. Comparison of various types of hydrogen bonds involving aromatic amino acids. *J Am Chem Soc.* 2002;124(44): 13257–13264.

[26] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157(1): 105–132.

[27] Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet.* 2018;59(3): 253–268.