In-silico Study of the Uncertainly Significant VCP Variants reveal major Structural, Functional fluctuations leading to potential Disease-Based Associations

Tathagata Das[1,2*]

[1] Centre for Genetic Disorders, Institute of Science, Banaras Hindu University, Varanasi 221005, India

[2] Department of Chemistry & Chemical Biology, Indian Institute of Technology (Indian School of Mines), Dhanbad 826004, India

[*] Corresponding Author

E-Mail Address: **22mt0404@iitism.ac.in**

Abstract

Mutations in VCP are responsible for a plethora of fatal diseases like Amyotrophic Lateral Sclerosis, Inclusion body myopathy with Paget disease of bone and frontotemporal dementia type 1, Spastic paraplegia, Charcot-Marie-Tooth disease type 2Y, Dementia, and Osteitis Deformans to name a few. Till now, studies on disease phenotype relationship, structural, and functional modifications of the protein's stability, phenotype, molecular dynamics, and post-translational modifications haven't been performed. This in-silico study investigates the variants of VCP (R95C, R95G, A160P, R191P, R191Q) which have conflicting interests of pathogenicity and hence are often depreciated and lack data. In addition, this study screens the study cohort and the fatal diseases linked to all of these variants. Interestingly through various computational tools and disease-based population studies, it was found that these variants are often found in patients linked with fatal diseases. The protein-protein interaction showed UFD1 has a direct association with VCP. The physicochemical parameters showed that A160P had the highest fluctuations of all the variants. By VCP's molecular dynamics simulations, its variation, deviation, compactness, flexibility, hydrogen bonds, and solvent accessibility were all comparatively impacted due to the changes caused by the variants. Box-plot and Principal component analysis of the MD simulations visualized the changes in the wild type of the protein. Research in wet labs and screening of patient cohorts is necessary to investigate further the diseases linked with these variants and will be valuable to establish their specific roles and can further be used as biomarkers for fatal rare diseases.

Keywords

VCP, In-Silico study, disease phenotype relationship, Post-translational modifications, Molecular dynamics simulations.

Introduction

The Valosin-containing protein is a hexameric protein of the ATPases associated with the diverse cellular activities (AAA) family, it has critical functions in the quality control of the protein and signalling pathways. It has roles in a plethora of cellular physiology and is directly connected to various other proteins and their functioning. The VCP/p97 system has been deemed an important factor in the ubiquitin system and intracellular pathways. It has its role in mitosis, involved in the fragmentation and reassembly of the Golgi stacks, which is the link to its involvement in the making of the endoplasmic reticulum. The concentration of the protein was found to be 34 micrograms/liter via mass spectrometry. The protein is additionally involved in stalled replication forks, and DNA damage response, where it is called into action for double-stranded breaks and limits the mutations occurring due to DNA damage, along with that it is the negative regulator of the production of type I interferon. Due to the involvement of VCP in all the above-mentioned roles it has a direct effect on genome stability and degenerative diseases.

VCP can cause various diseases if it gets mutated majorly for missense mutations, like Inclusion body myopathy with Paget disease of bone and frontotemporal dementia type 1, Amyotrophic lateral sclerosis, Spastic paraplegia, Frontotemporal dementia and/or amyotrophic lateral sclerosis 6, Inclusion body myopathy with Paget disease of bone and frontotemporal dementia, etc. These recorded diseases are only in the case of, as quoted "pathogenic" or "likely pathogenic" missense variants in dbSNP. However, disease information is variable with no specified information for Clinvar Accession which has clinical significance as "Pathogenic likely Pathogenic" or "Conflicting interpretations of pathogenicity". The frequency of the variants in these studies was only screened in a small cohort with rs1554668805 having no frequency data in dbSNP. With the growing number of missense variants in proteins, it should be known which mutations are causing the particular phenotype leading to degenerative diseases. The most prevalent method to confirm and cross-confirm variations is sequencing, Next comes the criteria of classifying the variation under SNP or mutation by population genetics study, but with the growing number of variants it is becoming arduous, not all variants are responsible for change in protein function which may lead to a disease. Hence this study focuses on giving a

prominent profile of VCP variants that fall under uncertain significance of pathogenicity such that during wet lab research these variants can be classified and further proven to be used as biomarkers for certain rare diseases.

In-silico tools provide data on conserved regions throughout evolution, structure, protein-protein interactions, favourable conformations, ligand binding sites, location, and function of the proteins to name a few. Gain of function and loss of function mutations are also predicted through the stabilization energy of the protein and are highly relatable if the conformational energy is changing due to a mutation. There were about 7979 variations reported of VCP at the time of this study of which 414 variations were missense variants and 18 were missense-pathogenic. It was identified that out of the total 7979 variants, 2 fell under missense-pathogenic likely pathogenic with inconsistent pathogenicity profile, and 1 variant was under conflicting interpretations of pathogenicity, which had uncertain clinical significance. Several bioinformatic tools were utilized to study the interactions, pathogenicity, and stability of the protein. The physicochemical, structural, functional, and evolutionary features were studied and analysed of the missense-pathogenic likely pathogenic and of variants with conflicting interpretations of pathogenicity. Diseases associated with VCP were scrutinized through various disease and variant databases to find which are the variants reported to cause disease, It was observed that rs121909332 had a total of 9 SNPs and 17 variant disease association evidence, for rs121909334 had 2173 total SNPs with various diseases combined and 46 variant disease association with strong evidence of R191Q causing autosomal dominantly inherited ALS, Inclusion Body Myopathy with Paget's disease of bone and Frontotemporal Dementia (IBMPFD)[1], Paget disease of bone, familial ALS, FTLD-TDP type IV, and Braak stage five Parkinson's disease[2], next for rs1554668805 interestingly no record was found with no publications with evidence in NCBI or other resources[3].

Materials and Methods

Dataset Preparation

The NCBI dbSNP database was screened to obtain the various variants of VCP. A total of 7979 results were obtained without any filter. Upon putting filter for only missense mutations 414 rsID results were obtained, benign-missense mutations had 4 results, likely benign-missense gave 8 results, pathogenic-missense gave 18 results, likely pathogenic-missense was 12, pathogenic likely pathogenic-missense had 2 rsID in it and only 1 result was obtained for conflicting interpretations of pathogenicity which fell under missense classification[4]. In this in-silico research rsID of pathogenic-likely pathogenic-missense and conflicting interpretations of pathogenicity-missense variants were considered to determine the significance of pathogenicity, and the deleterious nature of the VCP variants were obtained through several bioinformatics tools and algorithms. rs121909332, rs121909334, and rs1554668805 were analyzed here and the transcript isoforms were filtered out. Hence, in total 5 amino acid changes were scrutinized[3].

Protein-Protein Interaction and Disease Pathology of VCP

The interaction study presented the empirical data on the interaction networks based on the protein-protein interactions of the VCP protein. Hence to study the interaction of VCP with other proteins protein atlas database was used. Additionally, the pathology of the protein was studied to detect its expression in various diseases. The human protein atlas was used to focus on the expression profiles in human tissues and proteins[5]. The pathology information was based on mRNA and protein expression data from around 17 different forms of human cancers to find the precise marker where VCP is involved[6].

Pathogenicity Analysis

Online algorithm-based tools were used to determine, and predict the nature of pathogenicity of the VCP variants. The tools used were SIFT[7], Polyphen2[8], Panther[9], SNP&GO[10], PhDSNP[11], MetaRNN[12],

Mutpred2[13]. These tools require the wild-type protein sequence and the specific amino acid change site, then it gives the pathogenicity level or score of the specific variant. More details can be studied from[14].

Protein Stability Analysis

A plethora of tools were used to understand the stability profile of the variants. Gibbs free energy-based algorithms were used like Maestro[15], Cupsat (Cologne University protein stability analysis tool)[16], mCSM (mutation cutoff scanning matrix)[17], I-mutant[18], SDM[19], Dynamut[20], MUPro[21], DUET[17]. All these tools provided ddG values with the score determining whether the VCP variants in consideration are stabilizing or destabilizing. Further methodology details can be found in[22].

Evolutionary Conservation

The Consurf web server was used to study evolutionary conservation. The consurf server is a bioinformatics tool for scrutinizing the evolutionary conservation of amino acids in this case, based on phylogenetic relations between homologous sequences, in this study, Uniref90 and Uniprot databases were used separately. The evolutionary conservation of each of the five variants was analyzed against the conservation 9-pointer scale (Figure- 4) [23]. The evolutionary rate of a particular amino acid is directly dependent on its structural and functional importance. The specific conservation score by position was computed using the Bayesian and Machine Learning algorithms.

Stereochemical and Physicochemical Analysis

To precisely determine the secondary structure of the VCP protein, Psipred workbench was used. The particular VCP variants were analysed using the secondary structure prediction obtained[24]. Discovery Studio was used to induce the mutational changes in the wild-type protein and to draw the Ramachandran plots[25] for each of the five mutations plus the wild-type. All the plots were individually inspected to find whether the variants of concern fell in the allowed or the disallowed regions.

The understanding and determination of the physicochemical properties of a protein allows to establish specifications, and changes of the mutant variants when compared to the wild-type protein. It is necessary to know the physicochemical conditions of the receptor during the design of the ligand drug. Hence to study the physicochemical property of the wild-type VCP protein and compare it with the mutant variations ProtScale was used. It detects, computes, and shows the specific profile of an amino acid on the given protein. For this research, the hydropathocity, hydrophobicity, polarity, alpha helix, and beta-sheet plots were drawn for each of the variants, and their respective changes were noted and shown[24]. Expasy Protparam was used for the computation of several physical and chemical parameters like GRAVY value, Instability index, and aliphatic index of the wild-type protein as a reference against the mutants.

Genome and OMIM Investigation

The UCSC genome browser was used to visualize the genomic data of VCP[26]. The involvement of the variants with various diseases was procured from OMIM and SNPedia[27] [28]. The Online Mendelian Inheritance in Man was used to gather comprehensive knowledge about the VCP gene and its genetic phenotypes which are associated with diseases[29]. The three rsIDs of the five variants were used to study the severity of the diseases in association or which have been screened[30].

Post-Translational Modifications

Post-translational modifications of proteins are important to enhance the functional diversity of the proteins by the addition of various functional groups. It is a necessity to understand the post-translational modifications when analyzing the various functionalities of proteins, this eventually puts a transparent view of potential diseases possible due to a protein and its prevention, and treatment. For this research article, various PTM sites were investigated. Netphos 3.1 (Phosphorylation)[31], NetNGlyc 1.0, NetOGlyc 4.0

(Glycosylation)[32], GPS-SNO (S-nitrosylation)[33], GPS-SUMO (Sumoylation)[34], PrePs (Prenylation)[35], GPS-PAIL (N-acetylation)[36], GPS-Palm (Palmitoylation)[37], GPS-Uber (Ubiquitination)[38], GPS-MSP (Methylation)[39], KprFunc (Propionylation)[40].

Molecular Dynamics Simulations

Molecular dynamics simulations were used to study the conformational transition of the VCP protein when mutations were induced into it. It provides a direct connection between the structure and dynamics via the vast exploration of the conformational energy landscape which is accessible to the protein molecules. The GROMACS software was utilized to investigate VCP both in the wild type plus the mutant variants. For this study, GROMOS96 43a1 was used as the forcefield. Considering the dissolvable water surrounding the protein model spc18 was utilized. The framework utilized was the steepest descent approach to minimize the parameters of energy, equilibrated at 310 Kelvin temperature, 1 bar pressure using the canonical ensemble (NVT) (constant number of particles, temperature, and volume) and the (NPT) outfit (constant number of particles, temperature, pressure)[41]. Additionally in MD run parameters simulation time was rose to 50 nsec, MD integrator Leap-frog was used and 5000 frames were used per simulation. Root-mean square deviation (RMSD), Root-mean square fluctuation (RMSF), Radius of gyration (RoG), H-bonds between protein and water, Surface accessible solvent area (SASA), were all calculated and plotted.

Results

Protein-Protein Interaction and Disease Pathology of VCP

The interactions of the VCP protein showed that it majorly interacts with 48 proteins. The subcellular location of the protein was found to be in the Nucleoplasm, cytosol and the predicted location being intracellular. VCP was found to have low tissue specificity, it belongs to the protein class of enzymes and transporters. VCP has low tissue, and low immune cell specificity too. The pathology of the VCP protein was studied to understand its expression in various diseases. The pathology analysis of VCP was done for breast cancer, carcinoid, cervical cancer, colorectal cancer, endometrial cancer, glioma, head and neck cancer, liver cancer, lung cancer, lymphoma, melanoma, ovarian cancer, pancreatic cancer, prostate cancer, renal cancer, skin cancer, stomach cancer, testis cancer, thyroid cancer, and urothelial cancer. For all these diseases VCP was not prognostic but it was found that renal cancer and endometrial cancer had high expression for VCP, and the protein was prognostic for these diseases as well, the p-value for both was found to be less than 0.001. Hence VCP can be considered as a prognostic marker for both renal cancer and endometrial cancer.

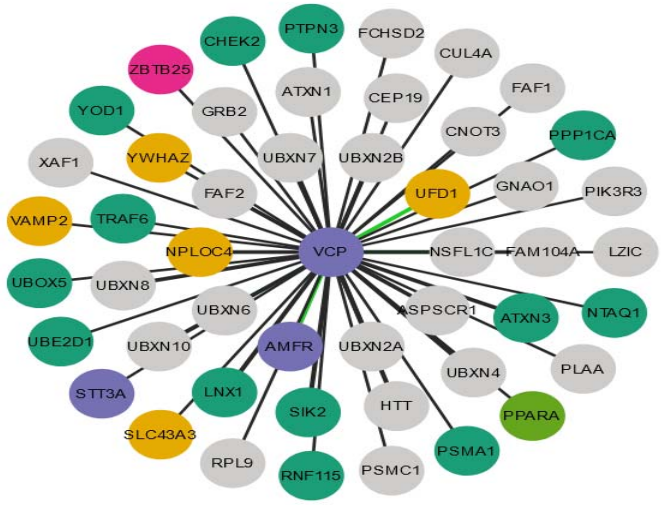


Figure 1: Protein-protein interaction of the VCP protein **(Colour - Yes)**

The green edge showed a direct association with UFD1. The colour code of the protein depicts the classes of the protein VCP interacts with. Dark green being only enzymes, violet for enzymes and transporters, pink being transcription factors, light green for transcription factors and transporters, yellow ochre for only transporters, and grey for others.

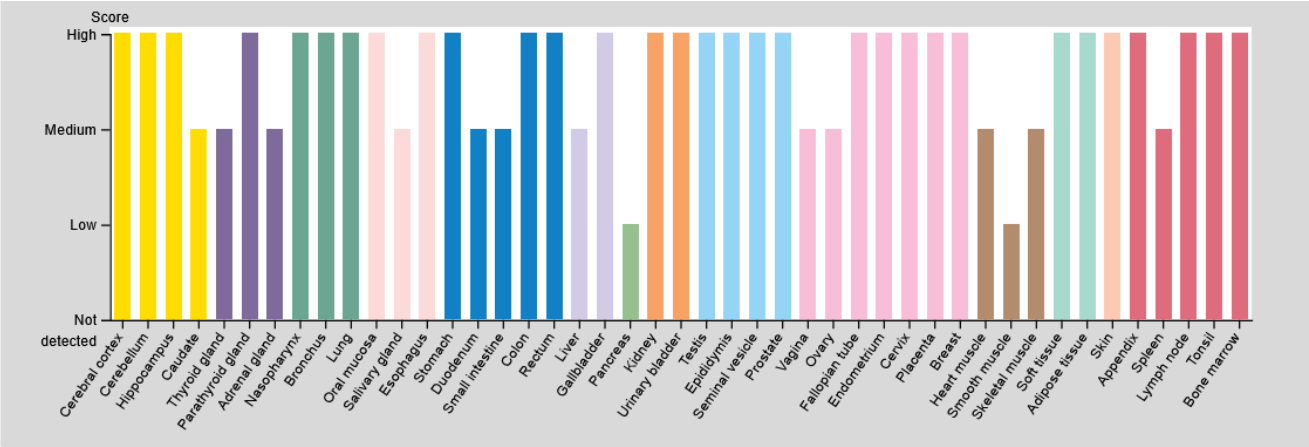


Figure 2: Expression of the VCP protein. **(Colour - No)**

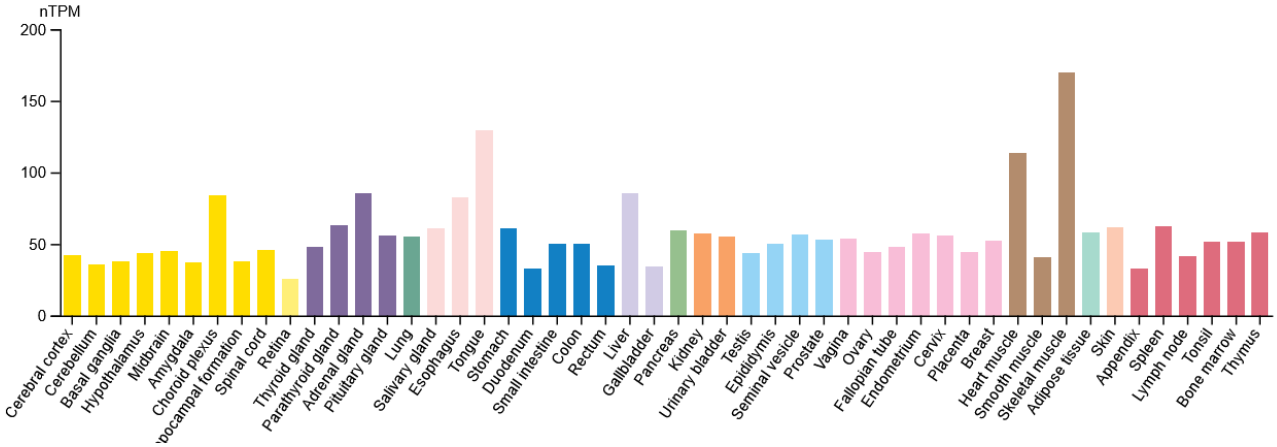


Figure 3: RNA tissue specificity. VCP showed to have low tissue specificity. **(Colour - No)**

Pathogenicity Prediction

The four variants of VCP which fall under the criteria of Pathogenic and likely pathogenic R95C, R95G (rs121909332), and R191P, R191Q (rs121909334) were analyzed for pathogenicity using various bioinformatics tools to establish the deleteriousness of the variants and to garner concrete insights to know whether the mutations can lead to fatal causes. The variant with the conflicting interest of pathogenicity was also considered to get a better idea such that the A160P (rs1554668805) mutation can fall under a proper, significant class of being deleterious or benign.

rsID	Amino Acid change	SIFT	POLYPHEN2	PANTHER	SNP&GO	PHDSNP	MetaRNN
	(Pathogenic likely Pathogenic)						
rs121909332 283C>T	R95C	Not tolerated	Probably damaging (Score – 0.993)	Probably damaging (Preservation time – 1629 my)	Disease (Reliability Index -7)	Disease Polymorphism (Reliability Index -5)	0.7522
283C>G	R95G	Tolerated	Probably damaging (Score – 0.999)	Probably damaging (Preservation time – 1629 my)	Disease (Reliability Index -2)	Neutral Polymorphism (Reliability Index -2)	0.9404
rs121909334 572G>C	R191P	Not tolerated	Probably damaging (Score – 1.0)	Probably damaging (Preservation time – 1629 my)	Disease (Reliability Index -8)	Disease Polymorphism (Reliability Index -4)	0.9019
572G>A	R191Q	Tolerated	Probably damaging (Score – 1.0)	Probably damaging (Preservation time – 1629 my)	Disease (Reliability Index -1)	Neutral Polymorphism (Reliability Index -3)	0.8911
	(Conflicting interpretations of pathogenicity)						
rs1554668805 478G>C	A160P	Tolerated	Benign (Score - 0.002)	Probably damaging (Preservation time – 1039 my)	Disease (Reliability Index -5)	Disease Polymorphism (Reliability Index -0)	0.5813

Table 1: Prediction of pathogenicity of the VCP variants

rsID	Amino Acid change	MutPred2 score	Remarks	Affected PROSITE and ELM Motifs
	(Pathogenic likely Pathogenic)			
rs121909332	R95C	0.429	-	-
	R95G	0.448	-	-
rs121909334	R191P	0.912	-	ELME000002, ELME000100, ELME000108, ELME000155
Molecular mechanisms with P-values <= 0.05			Probability	P-value
Gain of Intrinsic disorder			0.30	0.05
Gain of Loop			0.29	0.01
Altered Transmembrane protein			0.28	6.5e-04
Gain of B-factor			0.27	9.8e-03
Loss of Relative solvent accessibility			0.27	0.02
Altered Metal binding			0.13	0.05
	R191Q	0.609	-	ELME000002, ELME000100, ELME000108
Molecular mechanisms with P-values <= 0.05			Probability	P-value
Altered Transmembrane protein			0.29	3.8e-04
Gain of B-factor			0.25	0.02
Loss of Relative solvent accessibility			0.25	0.03
rs1554668805	A160P	0.928	-	ELME000137, PS00008
Molecular mechanisms with P-values <= 0.05			Probability	P-value
Altered Ordered interface			0.32	0.01
Gain of ADP-ribosylation at R159			0.22	0.02
Altered Transmembrane protein			0.21	4.2e-03
Gain of Allosteric site at F163			0.19	0.05

Table 2: The table depicts the Mutpred2 pathogenicity results of the VCP variants.

Protein Stability Prediction

The stability prediction of the mentioned VCP variants was done using various in-silico tools. The stability prediction of proteins is quintessential to understanding their molecular origins and what changes can cause the protein to be unstable, which may lead to a disease phenotype.

rsID	Amino Acid change	Maestro		Cupsat (kcal/mol)	mCSM (kcal/mol)	I-mutant (kcal/mol)	SDM	Dynamut (kcal/mol)	MUPro (kcal/mol)	DUET
Pathogenic likely pathogenic		ddG (kcal/mol)	Confidence prediction (kcal/mol)							
rs121909332	R95C	0.199 (Destabilizing)	0.923	-0.04 Destabilizing	-0.92 Destabilizing	-1.29 Destabilizing	-0.190 Destabilizing	-0.006 Destabilizing	-1.38 Destabilizing	-0.976 Destabilizing
	R95G	0.509 (Destabilizing)	0.870	0.35 Stabilizing	-1.246 Destabilizing	-1.81 Destabilizing	0.100 Stabilizing	-0.800 Destabilizing	-2.22 Destabilizing	-1.161 Destabilizing
rs121909334	R191P	0.271 (Stabilizing)	0.925	0.26 Stabilizing	-0.518 Destabilizing	-1.40 Destabilizing	-1.330 Destabilizing	0.371 Stabilizing	-1.22 Destabilizing	-0.796 Destabilizing
	R191Q	0.170 (Destabilizing)	0.932	-0.47 Destabilizing	-0.786 Destabilizing	-1.31 Destabilizing	-0.900 Destabilizing	-0.092 Destabilizing	-0.52 Destabilizing	-0.809 Destabilizing
(Conflicting interpretations of pathogenicity)										
rs1554668805	A160P	-0.110 (Stabilizing)	0.849	1.31 Stabilizing	-0.696 Destabilizing	-2.21 Destabilizing	-1.850 Destabilizing	-0.112 Destabilizing	-1.21 Destabilizing	-0.777 Destabilizing

Table 3: Prediction of protein stability on VCP variants

Evolutionary Conservation

The evolutionary conservation of the protein was checked using the Consurf web server and the data was checked for all the variants. The homologues for the evolutionary conservation were taken from two protein databases UniRef90 and UniProt separately.

Figure 4: Consurf Evolutionary conservation scale **(Colour - Yes)**

rsID	Amino Acid change	Conservation Score	Localization of residue	Role of residue
	Pathogenic likely Pathogenic			
rs121909332	R95C	5	Buried	Not available
	R95G	5	Buried	Not available
rs121909334	R191P	4	Exposed	Functional
	R191Q	4	Exposed	Functional
	(Conflicting interpretations of pathogenicity)			
rs1554668805	A160P	2	Buried	Not available

Table 3: Evolutionary conservation of VCP variants with homologues from UniRef90 database

rsID	Amino Acid change	Conservation Score	Localization of residue	Role of residue
	Pathogenic likely Pathogenic			
rs121909332	R95C	3	Exposed	Not available
	R95G	3	Exposed	Not available
rs121909334	R191P	3	Exposed	Not available
	R191Q	3	Exposed	Not available
	(Conflicting interpretations of pathogenicity)			
rs1554668805	A160P	1	Buried	Not available

Table 4: Evolutionary conservation of VCP variants with homologues from UniProt database

Stereochemical and Physicochemical Analysis

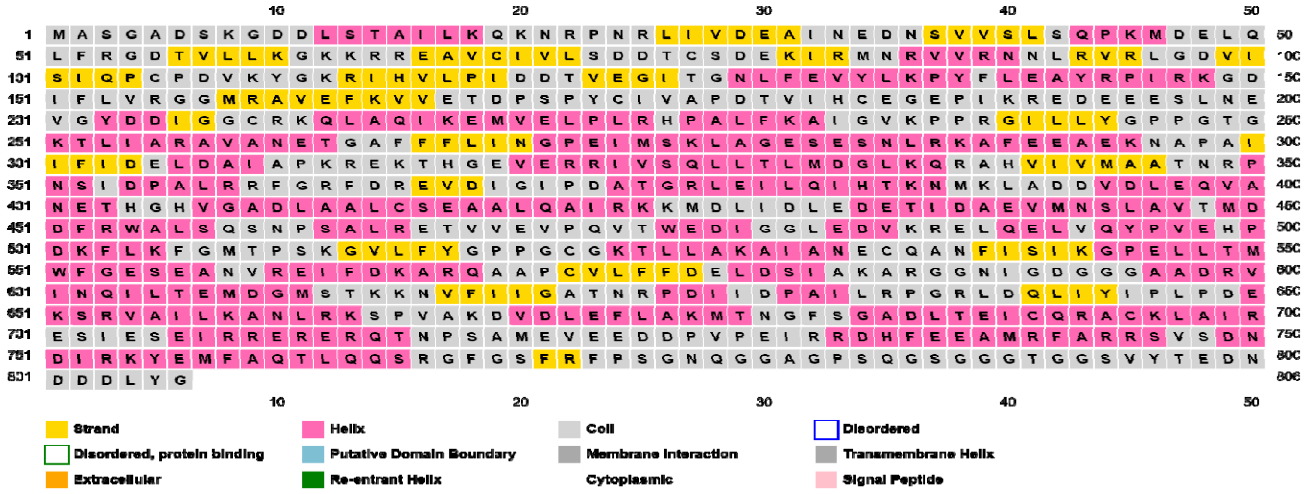


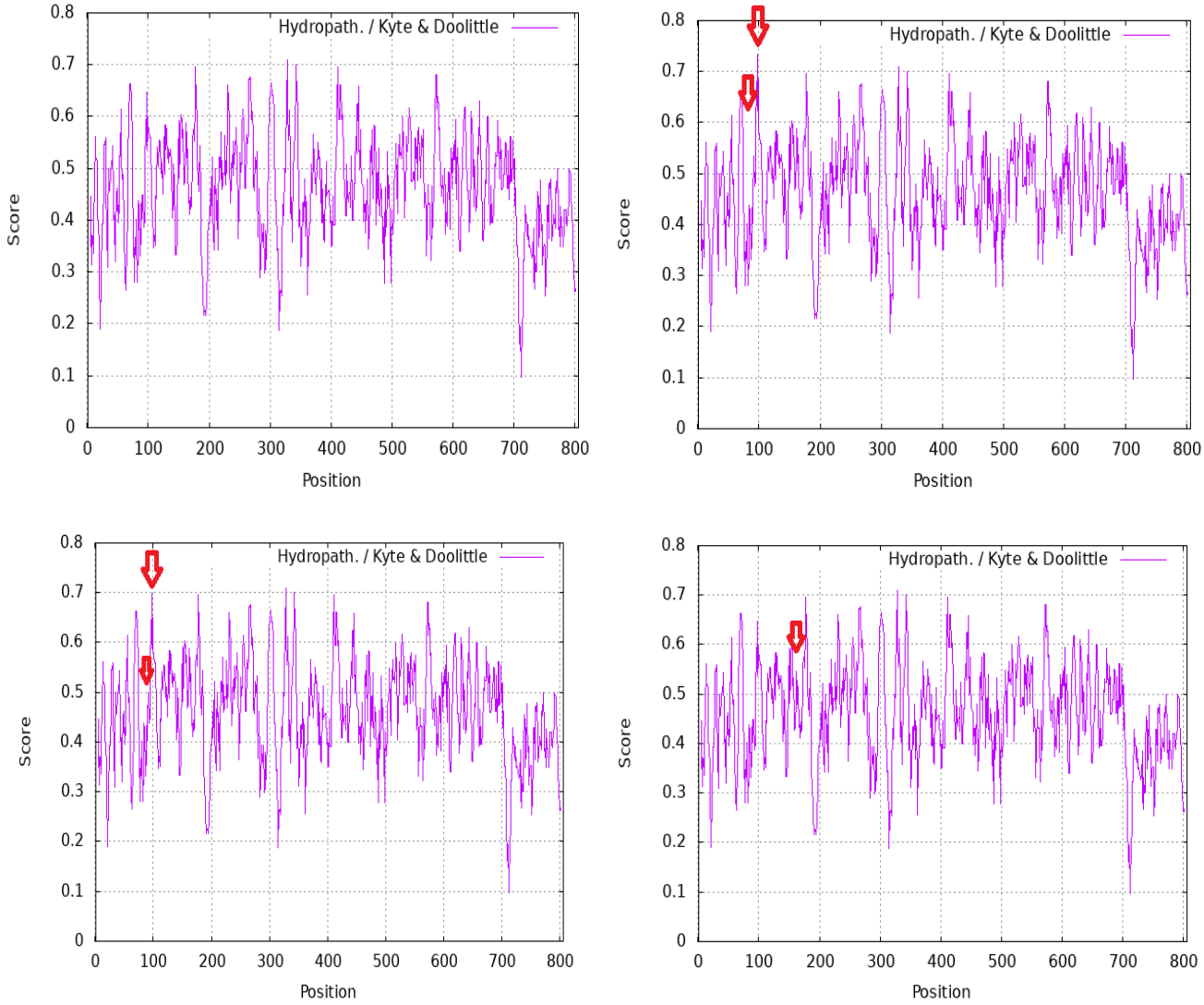
Figure 5: Depiction of the secondary structure of the protein in detail. **(Colour - Yes)**

It was observed that the mutations of interest fall in the following categories, position 95 (Strand potential formation of Beta-sheet), position 160 (Strand potential formation of Beta-sheet), and 191 (Coil).

The stereochemical quality of the protein structure was analyzed through the Ramachandran plot. Ramachandran plot of wild-type VCP protein showed above 95% of the residues fell in the favoured region.

It was noticed that rs121909332 when changed from Arginine to Cysteine (R95C) was under the disallowed region with the coordinates (144.27:0.00). Similarly, on the introduction of G in place of R in the 95th position the residue appeared in the disallowed region with the same coordinates as previously (144.27:0.00). For rs1554668805, when 160th residue changed from Alanine to Proline, the A160P residue appeared in the verge of allowed and disallowed regions in position (-71.00:-87.80). rs121909334, with Arginine replaced by Proline in the 191st position, led to the residue in the unfavourable region (-71.00:-98.57). Similarly, for R191Q, Glutamine was in the unfavourable region (148.69:125.37).

Certain important scales are used to describe a protein's function and activity, like Hydropathicity, which has an important role in drug design, biomolecular interactions of the proteins, Hydrophobicity, Polarity, the polarity pattern is of critical importance for the functioning of the protein, Alpha helix, and Beta-Sheet. Hence, to check whether the wild type of the VCP protein is getting influenced due to the mutations, all the variants were thoroughly scrutinized to find changes using the protein scales.



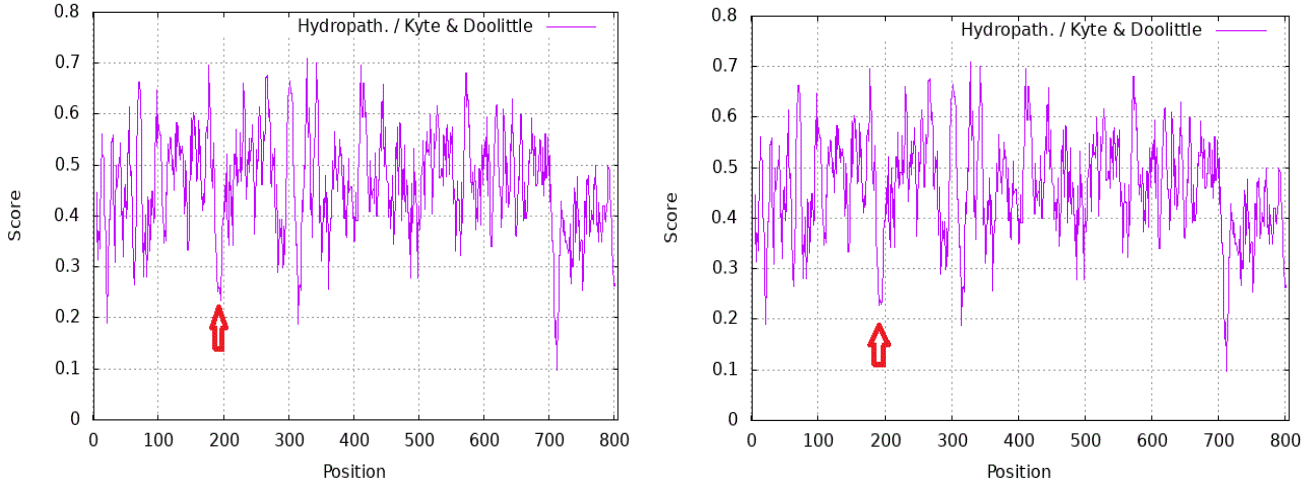
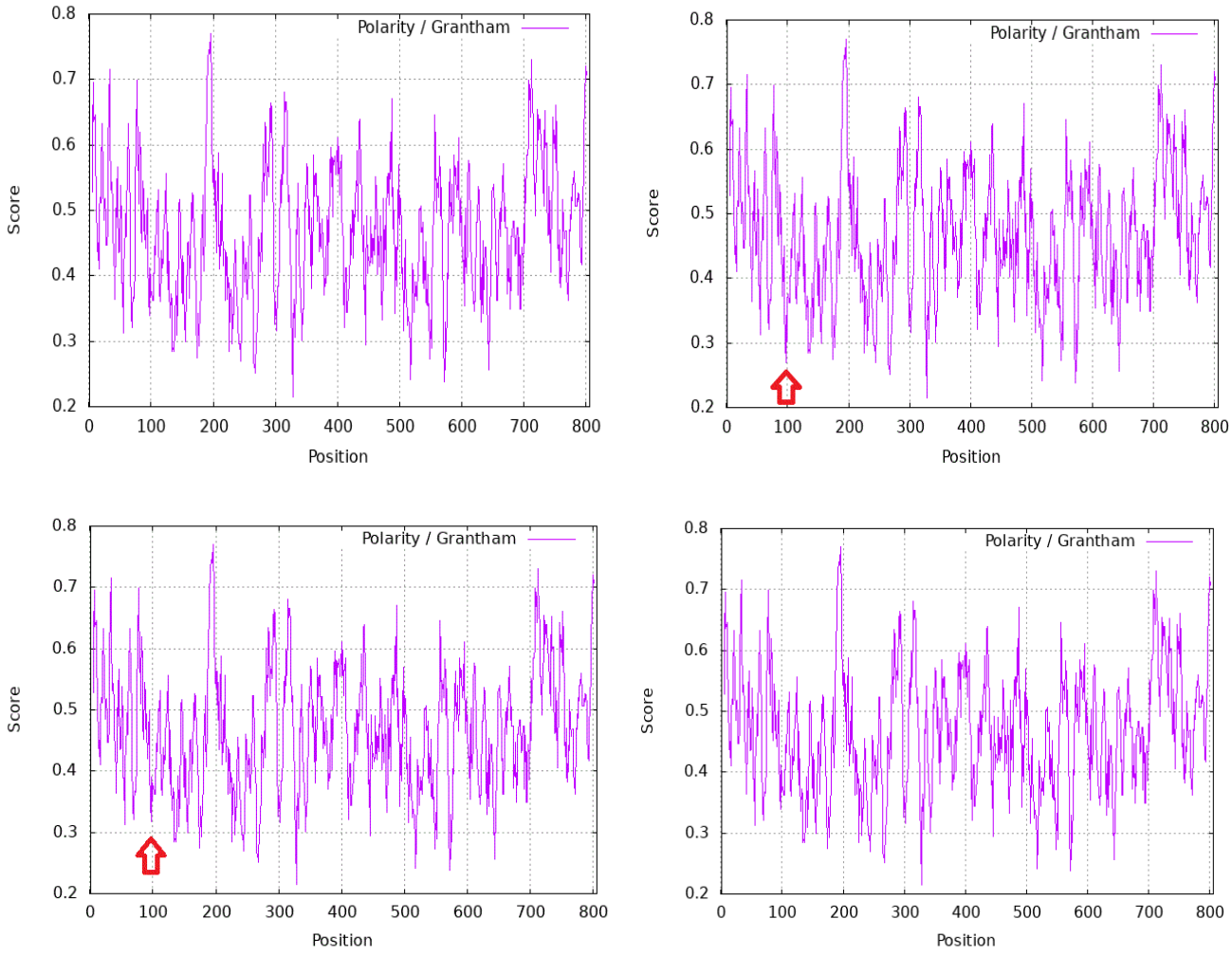


Figure 6: Hydropathicity plots of VCP Wild type and the five variants with the position of change marked with red arrow A) Wild type B) R95C C) R95G D) A160P E) R191P F) R191Q **(Colour - No)**



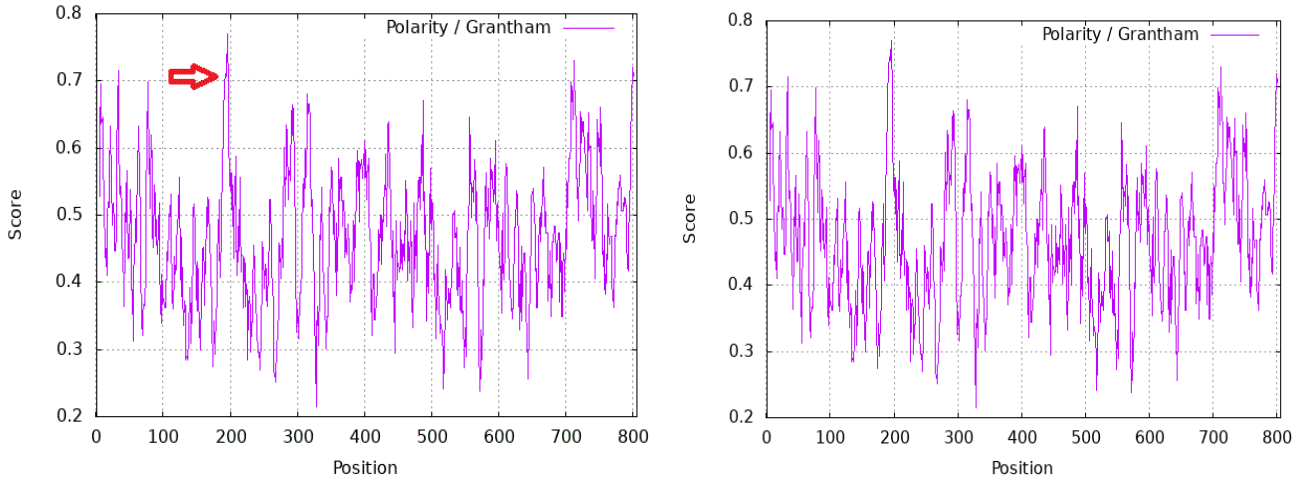
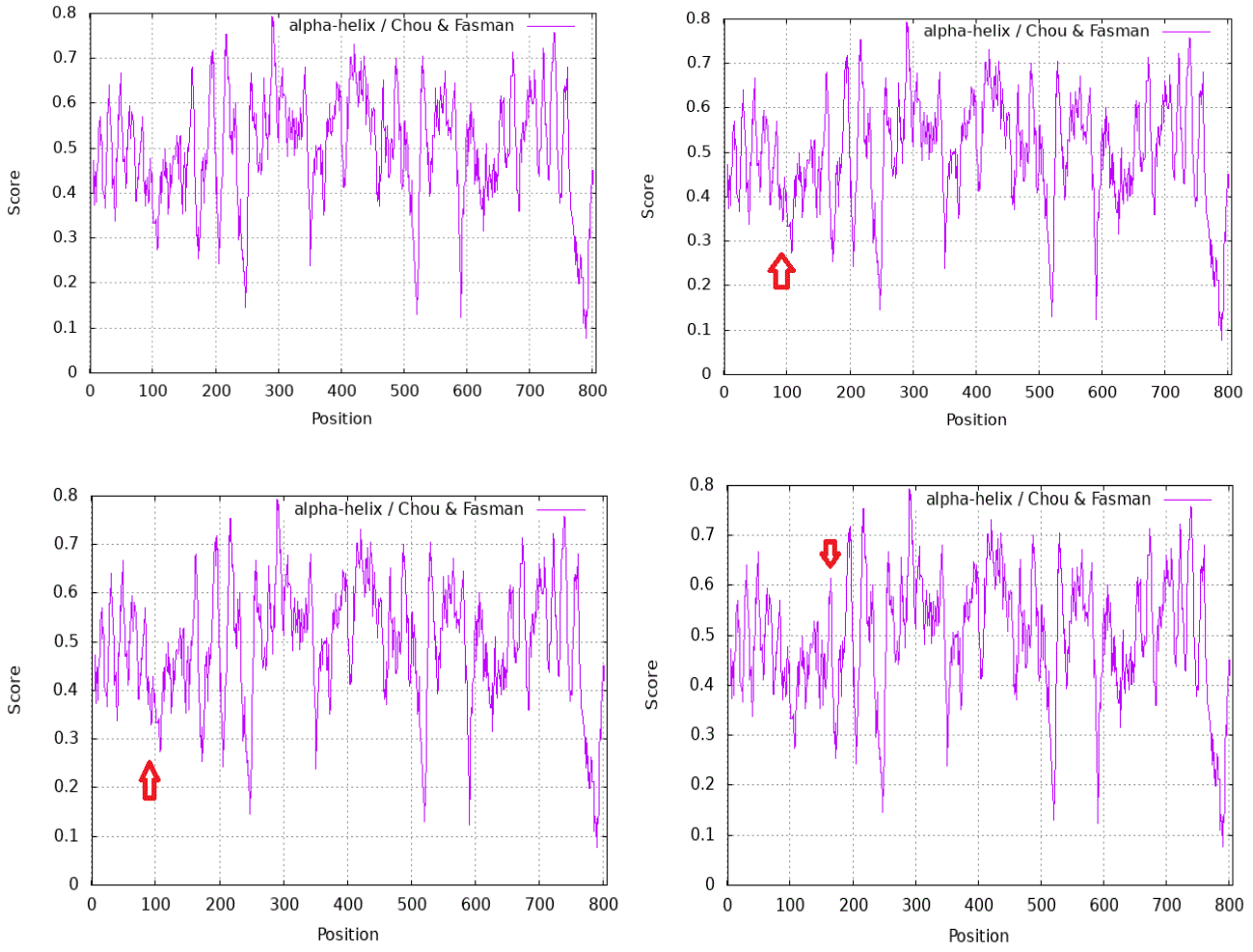
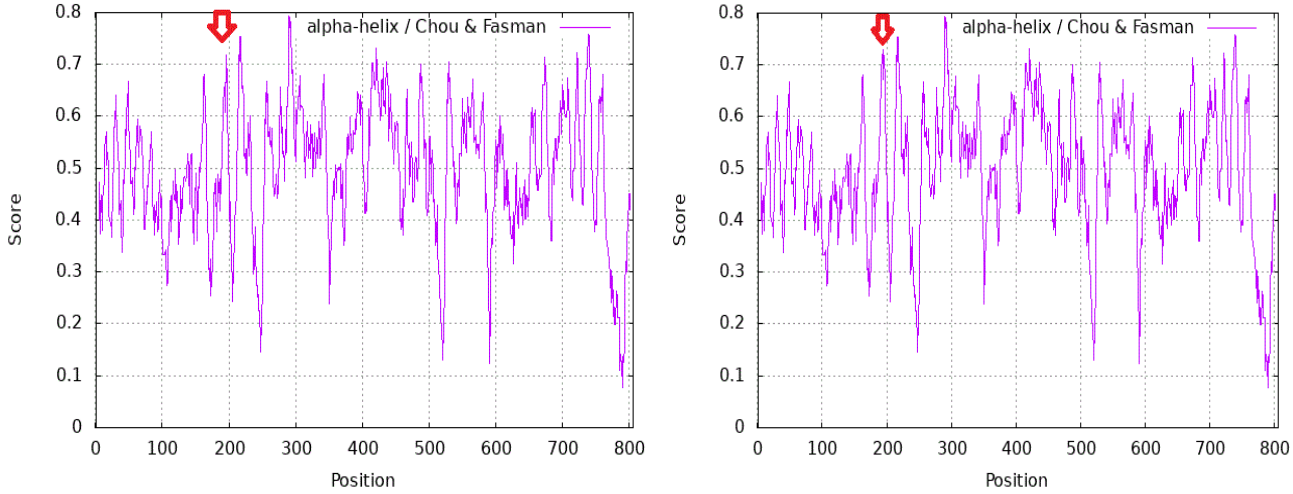
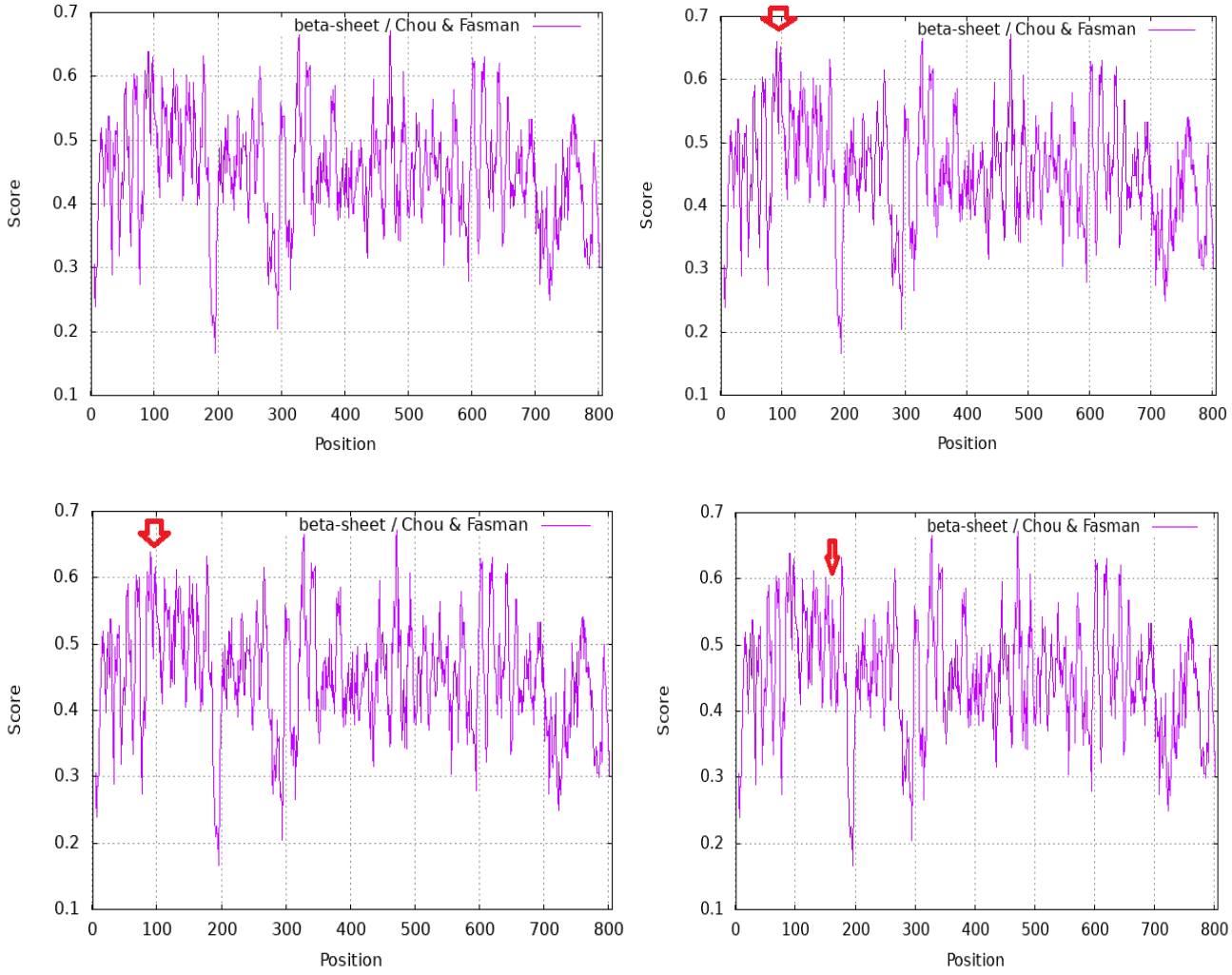


Figure 7: Polarity plots of VCP Wild type and the five variants with the position of change marked with a red arrow and with exception in A160P & R191Q where there's no change in Polarity as compared to the Wild type. A) Wild type B) R95C C) R95G D) A160P E) R191P F) R191Q **(Colour - No)**





<u>Figure 8:</u> Alpha-Helix plots of VCP Wild type and the five variants with the position of change marked with red arrow A) Wild type B) R95C C) R95G D) A160P E) R191P F) R191Q. **(Colour - No)**



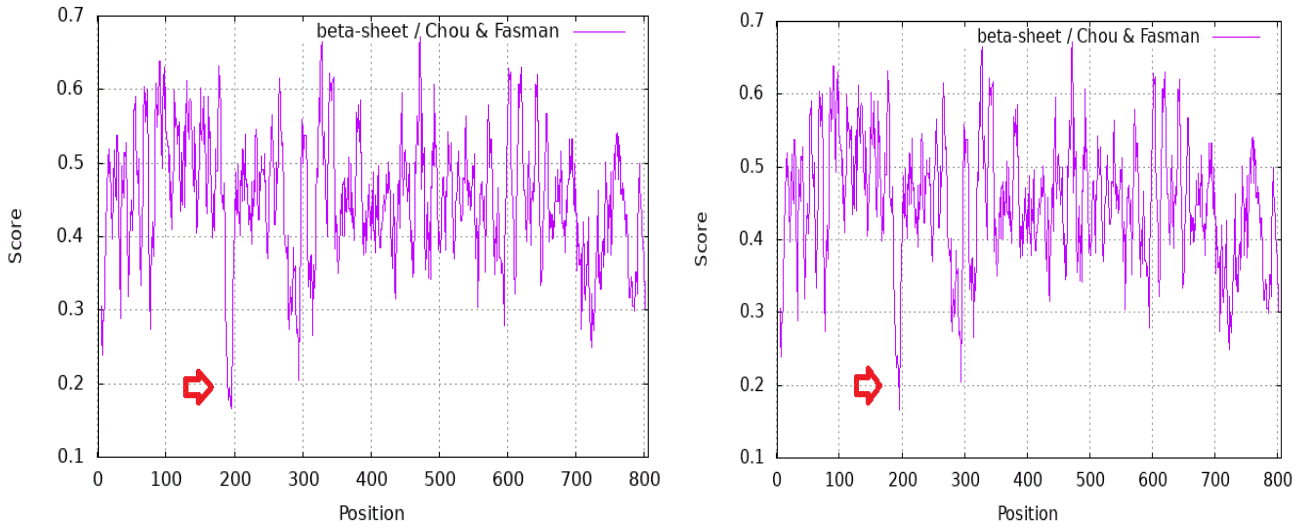


Figure 9: Beta-Sheet plots of VCP Wild type and the five variants with the position of change marked with red arrow. Starting from the top left. A) Wild type B) R95C C) R95G D) A160P E) R191P F) R191Q. **(Colour - No)**

Genome and OMIM Investigation

The UCSC genome browser was screened for the Valosin-containing protein under Human Assembly GRCH38/hg38 (December 2013 release). The position of VCP was observed to be in the position of p13.3 in chromosome 9: 33200001-36300000. Under Gene and gene predictions GencodeV43, NCBI RefSeq, and Gencode Versions (of all new and previous Gencode releases) were in visible respectively.

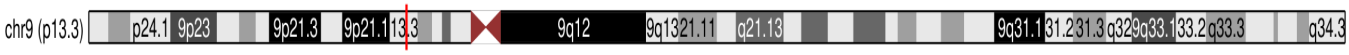


Figure 10: Chromosome 9 (p13.3) **(Colour - No)**

All the rsID of VCP through UCSC and OMIM were analyzed to search for the specific rsIDs of Pathogenic likely Pathogenic and Conflicting interpretations of pathogenicity which were obtained from dbSNP, to understand the extent of severity and the possible associated diseases with it.

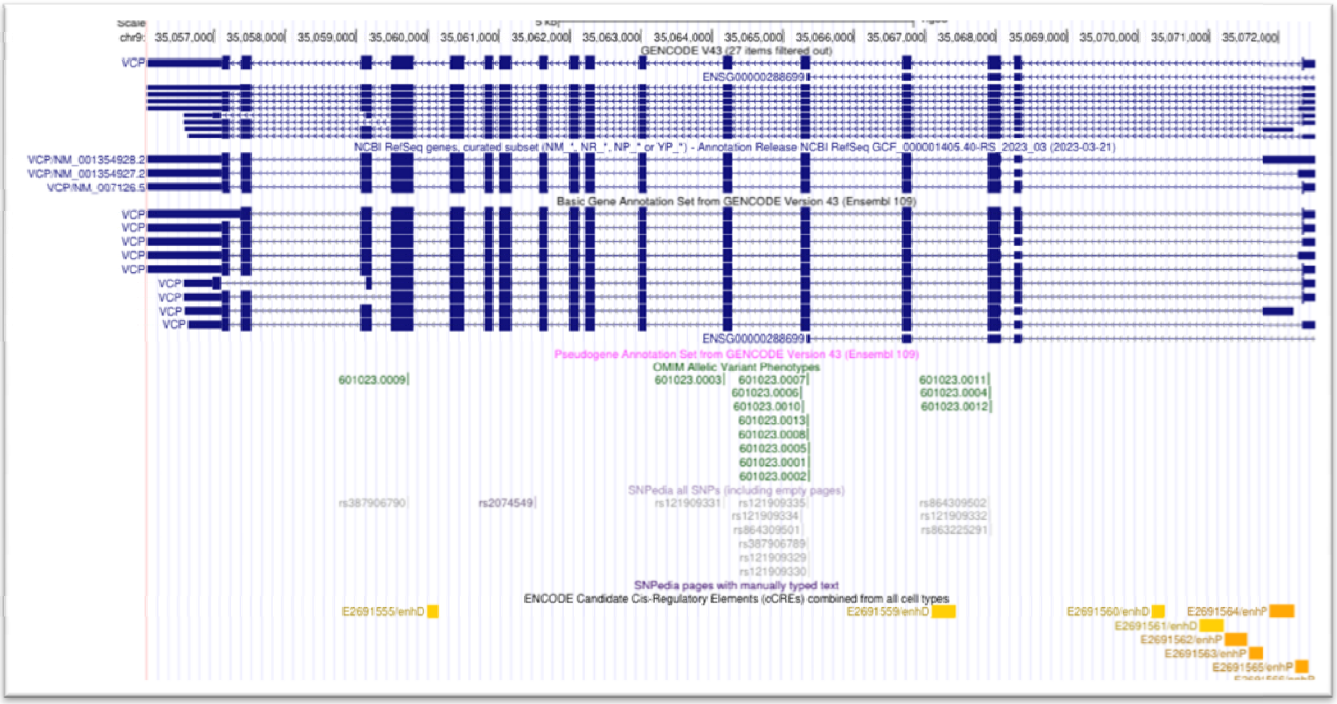


Figure 11: Data from UCSC showing VCP under Gencode, NCBI RefSeq, Ensemble along with OMIM variants, SNPedia, and ENCODE Candidate Cis-Regulatory Elements (cCREs). **(Colour - No)**

Upon analysis, rs121909334 (R191Q) with OMIM allelic variant 601023.0006 and position chr9:35065255-35065255 (Band - 9p13.3) was found to be associated with Inclusion Body Myopathy with Early-Onset Paget disease and Frontotemporal Dementia 1. Similarly, rs121909332 (R95G) with OMIM allelic variant 601023.0004 and position chr9:35067910-35067910 (Band - 9p13.3) was also found to be associated with Inclusion Body Myopathy with Early-Onset Paget disease and Frontotemporal Dementia 1.

Under SNPedia and all SNPs, rs121909332 and rs121909334 were found but with no concrete results. According to SNPedia the empty pages are autogenerated by bots and more details are to be added, there isn't enough information about these variants available, this may be due to lack of screening and prove of pathogenicity and disease pathology. Details regarding the variants R95C (rs121909332), R191P (rs121909334), and A160P (rs1554668805) were not found from UCSC combined with the OMIM. In ClinVar variant R95G was deemed as Pathogenic. The Human Assembly GRCH38/hg38 (December 2013 release) showed the ENCODE Candidate Cis-Regulatory Elements (cCREs), which are essential for the regulation of a gene.

Post Translational Modifications

	Glycosylation	Methylation	N-acetylation	Palmitoylation	Phosphorylation	Propionylation	Prenylation	S-nitrosylation	Sumoylation	Ubiquitinylation
Apoptosis			●					●		
Cell cycle division										●
Cell signalling				●				●		
Chromosome Assembly									●	
Chromosome maintenance		●						●		
DNA repair										
Extracellular Protein Export										
Immune Response										●
PPI	●				●	●	●			
Protein Activity	●							●		
Protein membrane							●		●	
Protein stability	●		●			●		●	●	●
Protein trafficking	●			●	●					
Protein thermodynamics	●									
Transcriptome		●								

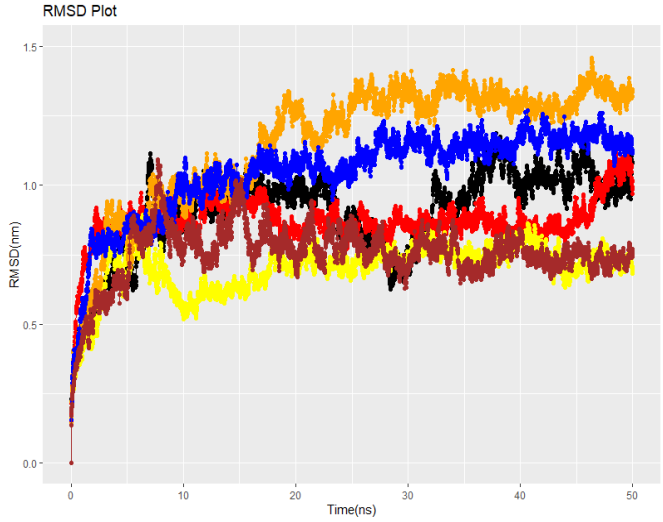
Table 5: Post-translational modification's impact on the functional aspects in a cell. The circles depict the functional aspects the PTMs are responsible for.

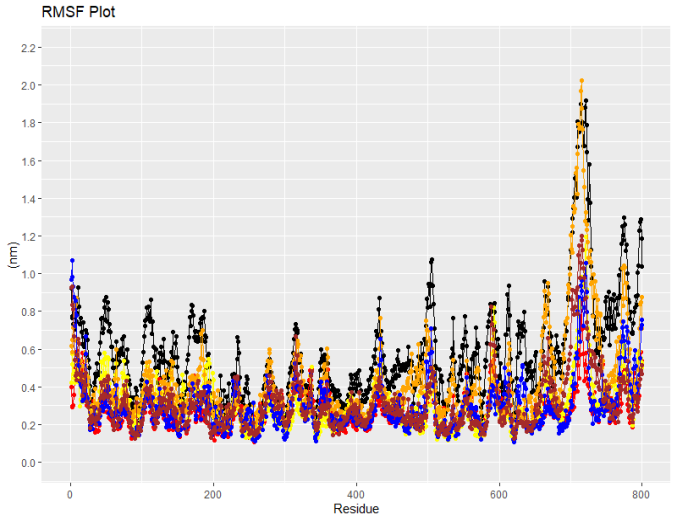
	Glycosylation	Methylation	N-acetylation	Palmitoylation	Phosphorylation	Propionylation	Prenylation	S-nitrosylation	Sumoylation	Ubiquitinylation
Apoptosis			●					R95C		
Cell cycle division										●
Cell signalling				●				R95C		
Chromosome Assembly									A160P, R191P, R191Q	

Chromosome maintenance		✓						R95C		
DNA repair										
Extracellular Protein Export										
Immune Response										✓
PPI	✓				R95C, R95G	✓	✓			
Protein Activity	✓							R95C		
Protein membrane							✓		A160P, R191P, R191Q	
Protein stability	✓		✓			✓		R95C	A160P, R191P, R191Q	✓
Protein trafficking	✓			✓	R95C, R95G					
Protein thermodynamics	✓									
Transcriptome		✓								

Table 6: Impacts on Post-translational modification due to the VCP variants. The variants responsible for changes in the functional aspects are accordingly mentioned in the table.

Molecular Simulations





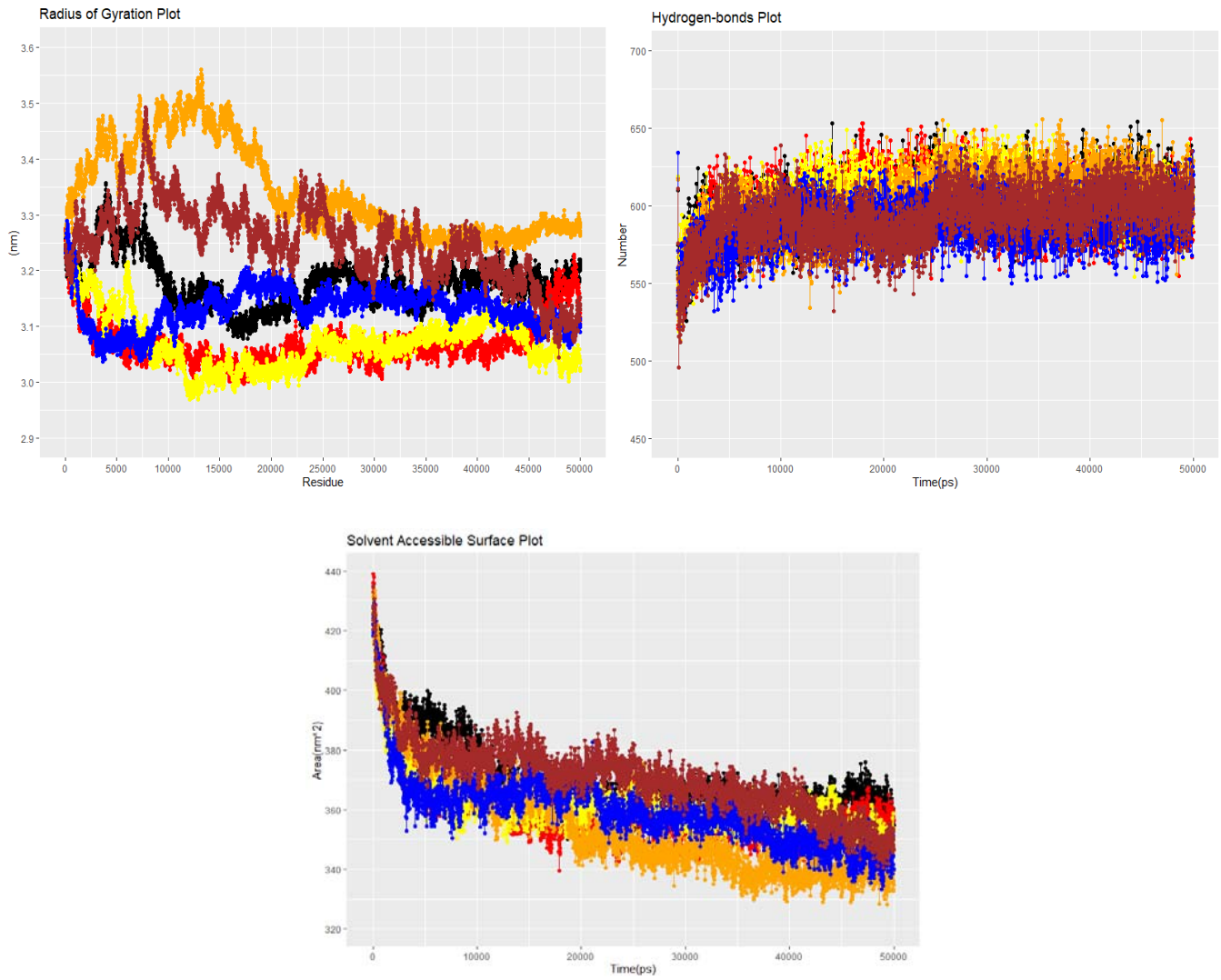


Figure 12: Molecular Simulation results of the VCP Wild type and the mutant variants A) RMSD B) RMSF C) Radius of Gyration (RoG) D) Hydrogen bonds E) Solvent Accessible Surface Area (SASA). Colour code: Wild- Black, R95C- Red, R95G- Yellow, A160P-Orange, R191P- Blue, R191Q- Brown. **(Colour - Yes)**

Discussion

In this research, various algorithms were used to confirm and shed light on the pathogenicity profile of VCP variants that don't fall under clear-cut brackets of pathogenic or benign. This investigational study can be used to predict the severity of the mutants and it eases the process to reach a definitive conclusion in mutational studies in both wet and dry labs.

The research was conducted with five variants of VCP R95C, R95G, A160P, R191P, R191Q obtained from the NCBI dbsnp database, where R95C, R95G (both rs121909332) and R191P, R191Q (both rs121909334) fell under the criteria of Pathogenic like Pathogenic, whereas A160P (rs1554668805) fell under the brackets of Conflicting interpretations of Pathogenicity. Upon literature review and analysis, it was found that these variants have been found in individuals with various diseases. rs121909332 was found to be involved in patients suffering from Inclusion Body Myopathy with Early-Onset Paget Disease with or without

Frontotemporal Dementia 1, rs121909334 was seen to be involved with Amyotrophic Lateral Sclerosis, Amyotrophic Lateral Sclerosis 14 with or without Frontotemporal Dementia, Charcot-Marie-Tooth disease type 2Y, Inclusion Body Myopathy with Early-Onset Paget Disease with or without Frontotemporal Dementia 1, Dementia, Osteitis Deformans, Inclusion Body Myopathy with Paget's disease of bone and Frontotemporal Dementia (IBMPFD), Parkinson's disease[1,2]. But no information could be found for rs1554668805, in regards to its relation with a disease. The variants were found to be variably conserved, with the lowest conservation score for A160P.

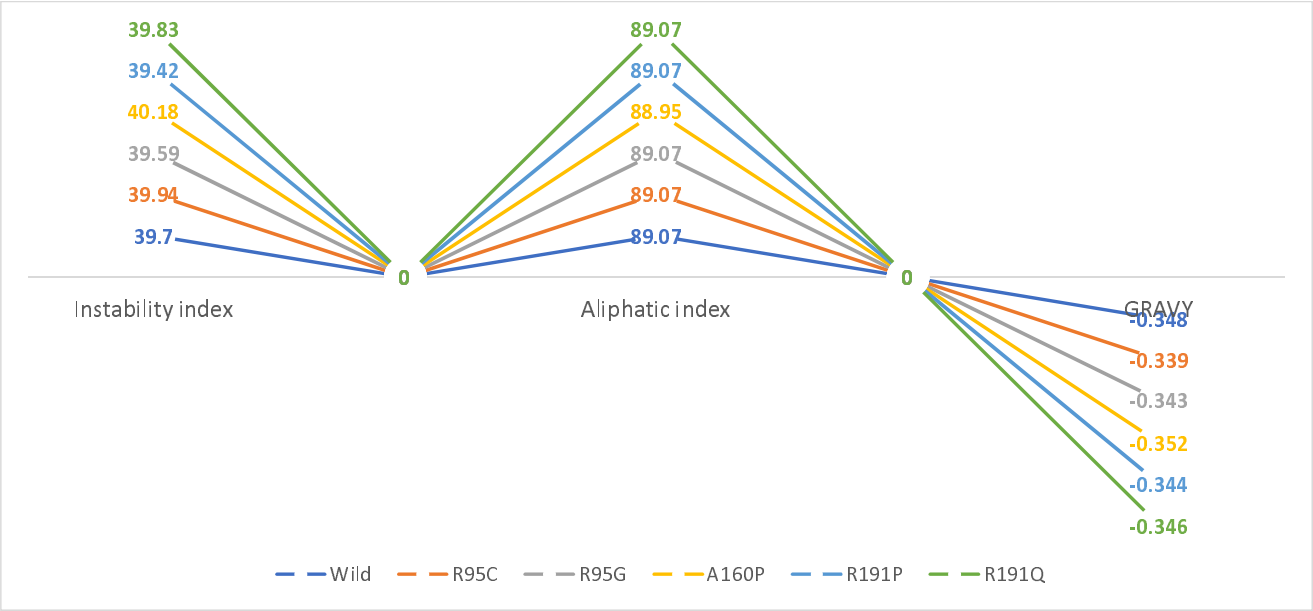


Figure 13: Variation in the Instability Index, Aliphatic Index, and GRAVY values of the mutants in comparison to the Wild type. **(Colour - Yes)**

It was seen from the indices above that the values were quite near to the Wild type VCP value, in the Instability Index, the highest variation was seen in A160P with a value of 40.18, for the aliphatic index A160P had the lowest value, that is 88.95 within all, in terms of GRAVY as well A160P had the most deviation with the lowest value -0.352.

All amino acids have their geometry, size, charge, hydropathicity, hydrophobicity, Polarity, Alpha-helix, and Beta-Sheet values. The variants often differ in these in comparison to the wild type of the protein. In the case of R95C and R95G, the wild-type residue forms a salt bridge with Aspartic Acid at position 98, and with Glutamic Acid at positions 194, 196, and 200, the difference in charge may affect the interactions made in the general case by the wild type. The backbone of amino acids remains the same always, the changes in properties come due to the change in the side-chain of the mutant from the wild type. As seen for R95C the side chain of the mutant residue is smaller and neutral as compared to the wild-type residue which is larger and positively charged. In the scenario for R191P and R191Q the mutant residue was found to be smaller and neutral than the wild-type residue which was larger and positively charged. The wild type formed a salt bridge

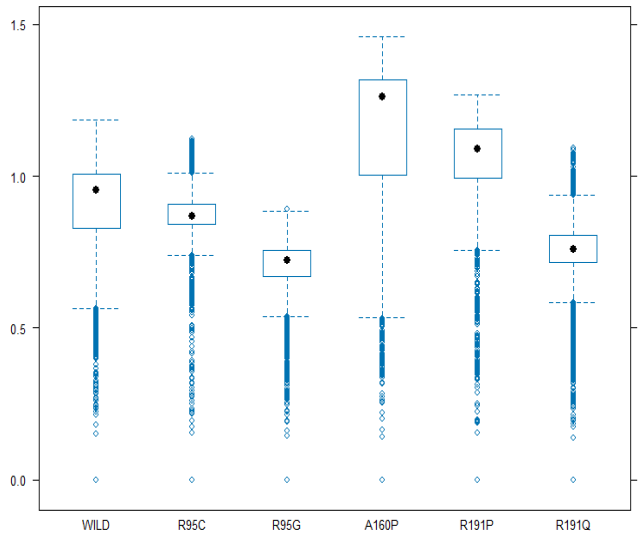
with Glutamic Acid in positions 162, 192, 194, and 195. The difference in charge will disturb the interactions made by the wild-type residue. Similarly, for R191Q, the mutant residue was found to be smaller and neutral than the wild-type residue which was larger and positively charged. The variants R95C, R95G, and R191P, were found to be more hydrophobic than the wild type. The wild type formed a salt bridge with Glutamic Acid in positions 162, 192, 194, and 195. The difference in charge will disturb the interactions made by the wild-type residue. For A160P, the mutant residue is bigger than the wild type and was found to be impacting the protein function, unlike the other variants much information couldn't be found on A160P variation. Considering all the data it can cause possible loss of external interactions of the protein[42]. By Interpro, Gene Ontology, and Broad Gene Ontology, the mutated residue is situated in a domain that is heavily part of the activity of the protein and in contact with other important residues. Due to this a mutation in that position can affect the proper functioning of the protein.

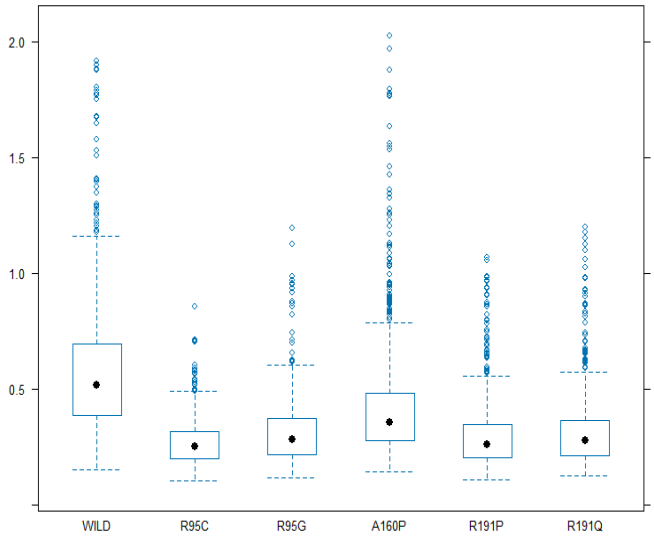
Considering the OMIM analysis, according to Weihl et al. (2006) for Inclusion Body Myopathy with Paget Disease of Bone and Frontotemporal Dementia 1 (IBMPFD1) there was an increase in the conjugates of ubiquitin which were diffused and aggregated and hence the improper function of the endoplasmic reticulum related degradation was observed, additionally, disproportionate Endoplasmic reticulum structure was noticed when cells were transfected with R95G (OMIM ID: 601023.0004)[43]. It was shown by Ju et al. that upon mutations in the VCP gene, the proper clearance of the aggregated proteins was disturbed. R191Q (OMIM ID: 601023.0006) was found in patients suffering from frontotemporal dementia along with/ amyotrophic lateral sclerosis-6 (FTDALS6).

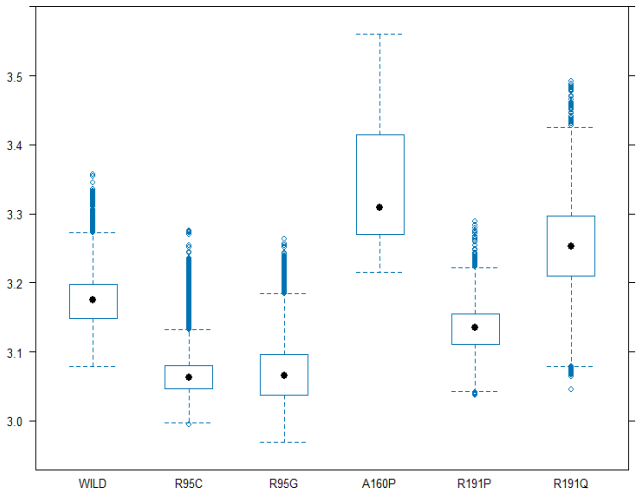
Autosomal dominant inclusion body myopathy with early-onset Paget disease and frontotemporal dementia 1 (IBMPFD1) was identified in 1 out of 13 families due to a transversion of C to G position 283 of the nucleotide VCP gene, eventually led to the substitution of Arginine with Glycine in the 95th position of the protein sequence.

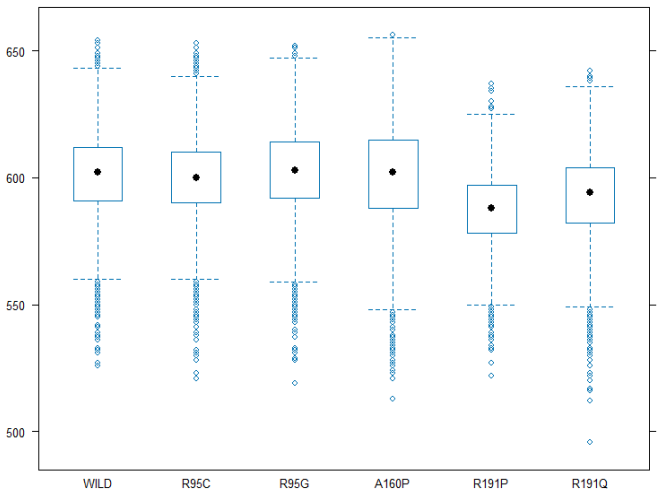
G to C transversion in the 572nd position of the VCP gene was seen in one family out of a total of 13 families screened which led to change in the 191st position to Glutamine from Arginine. R191Q mutation was identified as heterozygous in 4 affected members of the same Italian family who had FTDALS6. The affected individuals in adulthood had phenotypes of ALS, none had symptoms of Paget's disease and one of the other patients had mild frontotemporal dementia. In addition, phenotypes of IBMPFD were seen in two distinct men aged around fifty, one had Scapuloperoneal weakness while not having any facial involvement, with increased creatine kinase, and the other person had facial weakness, girdle weakness in the shoulder and anterior foreleg, plus had increased creatine kinase by 4 times. Further investigations showed myopathic patterns and dystrophic changes, later mild dysexecutive syndrome was found in one of the patients, but there was no evidence of Paget's disease. These findings tend to bring a vast spectrum of the disease-associated causes of VCP mutations.

MD simulation results of the wild type and mutants are shown (Figure - 12) depicting the RMSD, RMSF, Radius of gyration, and Solvent accessible surface area. MD was utilized to get an understanding of a protein's structural and functional behaviour in the environment and provide data on real-time changes that are happening in the variants in comparison to the wild type and thus helped to understand the phenotypic changes which may happen due to differences in stability, conformation, flexibility. When RMSD analysis was used to compare to Wild type, R95C, R95G, and R191Q were stable, high fluctuation was seen in A160P, and moderate fluctuation in R191P. From the RMSF graph, all the variants looked to have lower flexibility than the wild type only A160P was comparably seen to have higher flexibility than the rest of the variants. The gain or loss of flexibility may directly impact the protein's functioning. RoG determines the compaction of the proteins, R95C, and R95G were more compact than wild unlike A160P which showed the lowest compaction out of all, and R191Q even had lower compaction than wild type. R191P was seen to be almost similar to the wild. The high value of RoG suggests high flexibility and even less stability. Hydrogen bonds contribute favourably to protein's stability, R191P had lower average H-bonds than wild, R95G, A160P, R95C, and R191Q. SASA predicts the solvent environment accessibility of the protein, only R191Q showed a little higher solvent accessibility than the wild compared to the other variants which showed lower solvent accessibility than the wild with A160P being the lowest of all. SASA is of intense interest as it has direct roles in functional studies and protein folding.


1.5
1.0
0.5
0.0
WILD
R95C
R95G
A160P
R191P
R191Q


2.0
1.5
1.0
0.5
WILD
R95C
R95G
A160P
R191P
R191Q


3.5
3.4
3.3
3.2
3.1
3.0
WILD
R95C
R95G
A160P
R191P
R191Q


650
600
550
500
WILD
R95C
R95G
A160P
R191P
R191Q

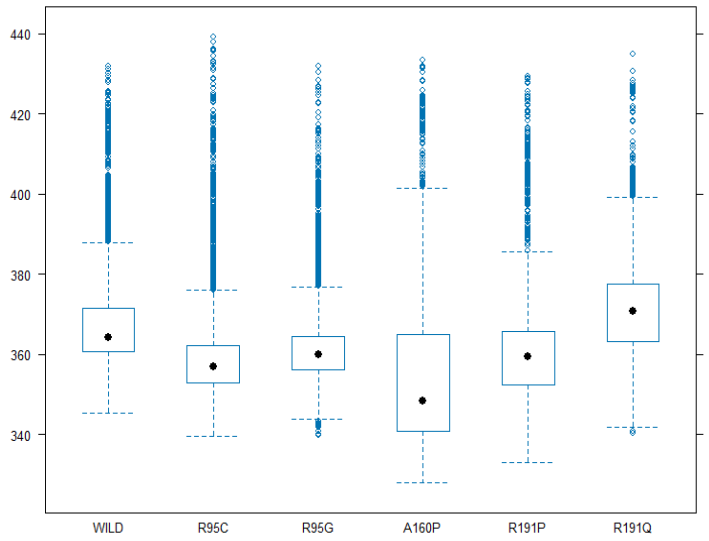
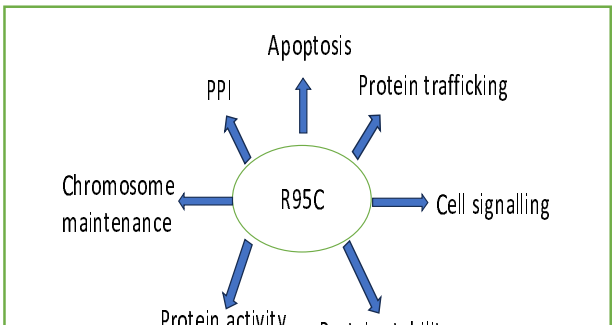
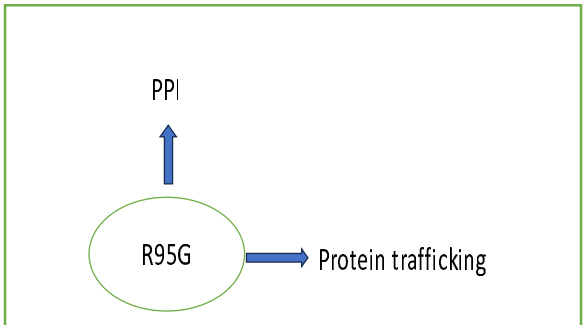


Figure 14: Detailed visualization of the Molecular dynamics data of VCP wild type and the variants. A) RMSD B) RMSF C) Radius of Gyration D) Hydrogen bonds E) Solvent Accessible Surface Area (SASA). **(Colour - No)**

The analysis of Post-translational modifications showed three parameters being affected due to R95C, R95G (affects phosphorylation), R95C (affects S-nitrosylation), A160P, R191P, R191Q (affects Sumoylation), respectively. It should be noted that all the post-translational modifications were checked against the VCP wild type. The changes in Phosphorylation seen in the variants R95C, R95G were found to be below the threshold, hence the changes mayn't affect the PTM of VCP to a very large extent. The regions impacted by S-nitrosylation in the VCP wild type were found to be in the amino acid positions 77, 209, and 695. Similarly, for Sumoylation the sites impacted were 8, 136, 164, 486, and 754 in the wild-type VCP variant. Consecutively, the same was checked for N-acetylation which showed changes in 8, 236, and 668, in Methylation change was in the amino acid number 8. Propionylation showed its impact on amino acid sites 60, 109, 502, and 677 in the VCP variant respectively. Hence, all the variants of VCP were found to be impacting any one or the other PTM out of a total of ten being considered in this study. The structure and the function therefore gets impacted due to the variants of VCP, affecting the biological processes. The variants may eminently regulate the functioning due to their impact on the Post-translational modifications.





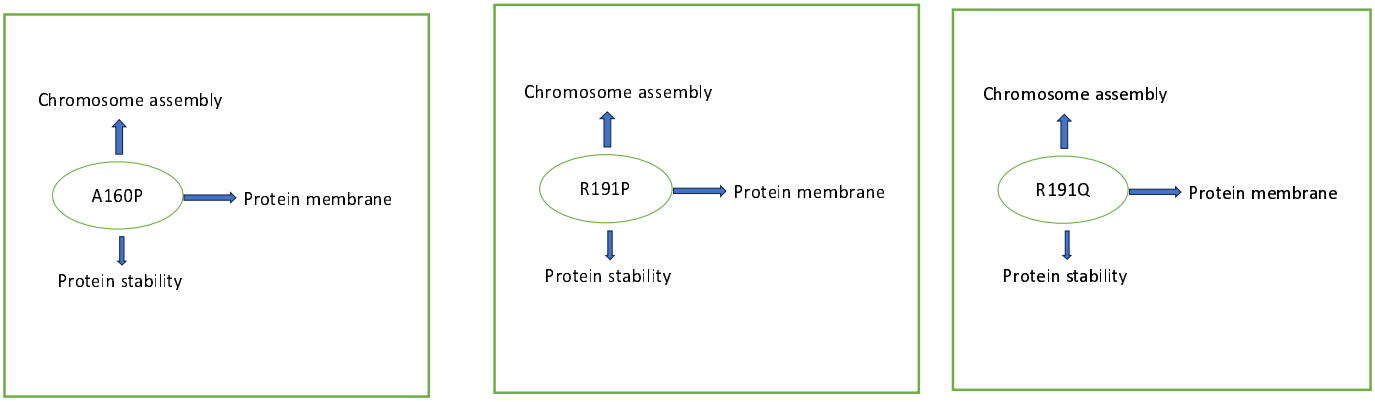
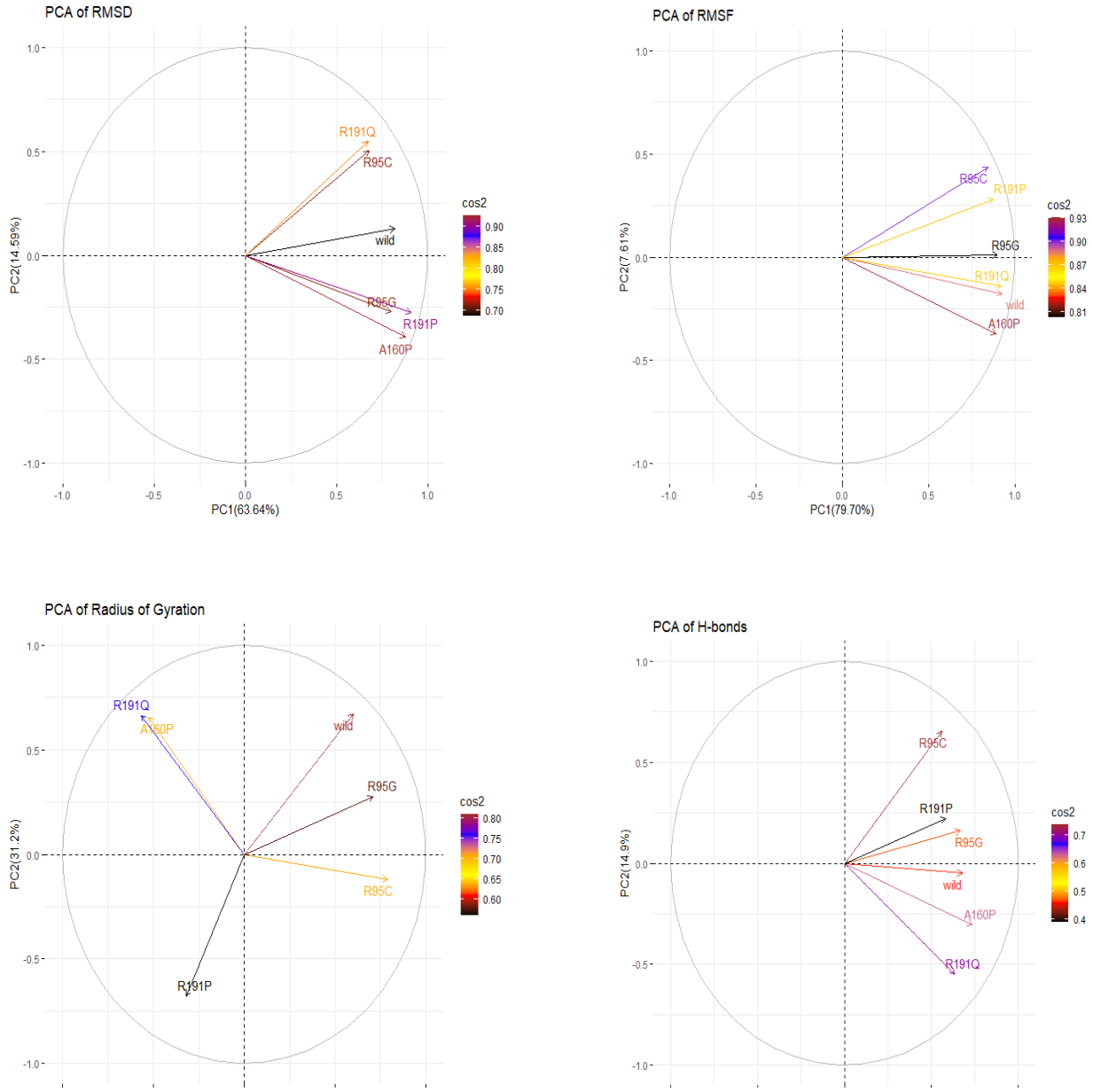


Figure 15: All the VCP variants responsible for changes in the various functional domains of the Post-Translational modifications. **(Colour - No)**



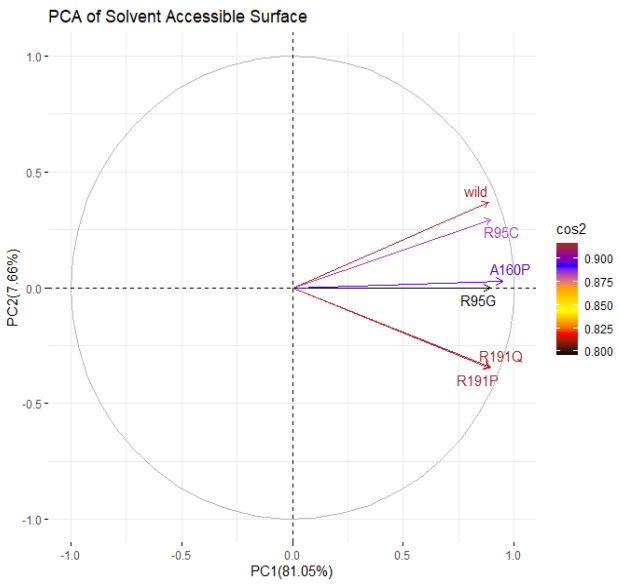


Figure 16: Principal Component Analysis (PCA) of all the VCP variants. A) RMSD B) RMSF C) Radius of Gyration D) Hydrogen Bonds E) Solvent Accessible Surface Area (SASA). **(Colour - Yes)**

PCA analysis was used to visualize the data in an exquisite way and for a better understanding of the MD simulations of the wild type and the variants (R95C, R95G, A160P, R191P, R191Q). A detailed explanation of PCA can be found in[44].

Conclusion

Investigations were done with in-silico tools to study the R95C, R95G, A160P, R191P, and R191Q mutations in the VCP protein. By this study, all the mutations were found to be mildly pathogenic but variants like R95G (rs121909332), and R191Q (rs121909334) have heavy evidence against various diseases reported in patients. Unlike these two variants, disease reports against R95C, A160P, and R191P aren't to be found, especially rs1554668805 (A160P) which has no information in OMIM, this can depict either these variants aren't responsible for any fatal disease or more screening of patients are required to find the significant functional disease-based changes caused by these specific variants if any. All 5 variants of VCP showed indications of structural and functional changes in gene ontology, Protein expression, Protein interactions, PTM, and Molecular simulations but the variants were not in highly conserved regions. Hence, disease relation was further investigated, the major variants were found to be R95G (601023.0004) and A191Q (601023.0006), respectively with deep-phenotypic details. The molecular simulation plots further validated the precise malfunctioning of the VCP protein upon the induction of the variant amino acid changes especially rs1554668805 which had notable fluctuations in comparison to the wild and the other four variants. This analysis can be used in wet lab screening to further understand the disease relation and the possibility of finding biomarkers in the five variants, especially in those variants that show significant fluctuations but have no fatal disease-related findings and, frequency data available to date.

Abbreviations

GRAVY: Grand Average of Hydropathy

RMSD: Root Mean Square Deviation

RMSF: Root Mean Square Fluctuation

RoG: Radius of Gyration

H-Bonds: Hydrogen bonds

SASA: Solvent Accessible Surface Area

IBMPFD: Inclusion Body Myopathy with Paget's disease of bone and Frontotemporal Dementia

PCA: Principal Component Analysis

CRediT authorship contribution statement

All work regarding this study was done by Tathagata Das.

Declaration of competing interest

There is no known competing financial interest, personal relationships, or conflict of interest involved with this paper.

Ethical Approval

Not applicable

Acknowledgement

Banaras Hindu University, Varanasi, India for providing with internet services, IIT (ISM) Dhanbad, India for allowing to carry out research offsite and AICTE, for providing scholarship to Tathagata Das.

References

[1] P. Cloutier, M. Lavallée-Adam, D. Faubert, M. Blanchette, B. Coulombe, A Newly Uncovered Group of Distantly Related Lysine Methyltransferases Preferentially Interact with Molecular Chaperones to Regulate Their Activity, PLoS Genet. 9 (2013) e1003210. https://doi.org/10.1371/journal.pgen.1003210.

[2] E. Majounie, B.J. Traynor, A. Chiò, G. Restagno, J. Mandrioli, M. Benatar, J.P. Taylor, A.B. Singleton, Mutational analysis of the VCP gene in Parkinson's disease, Neurobiol Aging. 33 (2012) 209.e1-209.e2. https://doi.org/10.1016/j.neurobiolaging.2011.07.011.

[3] National Center for Biotechnology Information (NCBI), (1998). https://www.ncbi.nlm.nih.gov/ (accessed July 24, 2023).

[4] E.W. Sayers, E.E. Bolton, J.R. Brister, K. Canese, J. Chan, D.C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B.W. Trawick, K.D. Pruitt, S.T. Sherry, Database resources of the national center for biotechnology information, Nucleic Acids Res. 50 (2022) D20–D26. https://doi.org/10.1093/nar/gkab1112.

[5] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome, Science (1979). 347 (2015). https://doi.org/10.1126/science.1260419.

[6] The Protein Atlas, (2023). proteinatlas.org (accessed August 20, 2023).

[7] P.C. Ng, S. Henikoff, Predicting Deleterious Amino Acid Substitutions, Genome Res. 11 (2001) 863–874. https://doi.org/10.1101/gr.176601.

[8] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2, Curr Protoc Hum Genet. 76 (2013). https://doi.org/10.1002/0471142905.hg0720s76.

[9] P.D. Thomas, D. Ebert, A. Muruganujan, T. Mushayahama, L. Albou, H. Mi, <scp>PANTHER</scp>: Making genome-scale phylogenetics accessible to all, Protein Science. 31 (2022) 8–22. https://doi.org/10.1002/pro.4218.

[10] R. Calabrese, E. Capriotti, P. Fariselli, P.L. Martelli, R. Casadio, Functional annotations improve the predictive score of human disease-related mutations in proteins, Hum Mutat. 30 (2009) 1237–1244. https://doi.org/10.1002/humu.21047.

[11] E. Capriotti, R. Calabrese, R. Casadio, Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information, Bioinformatics. 22 (2006) 2729–2734. https://doi.org/10.1093/bioinformatics/btl423.

[12] C. Li, D. Zhi, K. Wang, X. Liu, MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning, Genome Med. 14 (2022) 115. https://doi.org/10.1186/s13073-022-01120-z.

[13] V. Pejaver, J. Urresti, J. Lugo-Martinez, K.A. Pagel, G.N. Lin, H.-J. Nam, M. Mort, D.N. Cooper, J. Sebat, L.M. Iakoucheva, S.D. Mooney, P. Radivojac, Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, Nat Commun. 11 (2020) 5918. https://doi.org/10.1038/s41467-020-19669-x.

[14] P. Ranjan, P. Das, Understanding the impact of missense mutations on the structure and function of the ***EDA*** gene in X-linked hypohidrotic ectodermal dysplasia: A bioinformatics approach, J Cell Biochem. 123 (2022) 431–449. https://doi.org/10.1002/jcb.30186.

[15] J. Laimer, H. Hofer, M. Fritz, S. Wegenkittl, P. Lackner, MAESTRO - multi agent stability prediction upon point mutations, BMC Bioinformatics. 16 (2015) 116. https://doi.org/10.1186/s12859-015-0548-6.

[16] V. Parthiban, M.M. Gromiha, D. Schomburg, CUPSAT: prediction of protein stability upon point mutations, Nucleic Acids Res. 34 (2006) W239–W242. https://doi.org/10.1093/nar/gkl190.

[17] D.E. V. Pires, D.B. Ascher, T.L. Blundell, mCSM: predicting the effects of mutations in proteins using graph-based signatures, Bioinformatics. 30 (2014) 335–342. https://doi.org/10.1093/bioinformatics/btt691.

[18] K.A. Bava, ProTherm, version 4.0: thermodynamic database for proteins and mutants, Nucleic Acids Res. 32 (2004) 120D – 121. https://doi.org/10.1093/nar/gkh082.

[19] C.L. Worth, R. Preissner, T.L. Blundell, SDM--a server for predicting effects of mutations on protein stability and malfunction, Nucleic Acids Res. 39 (2011) W215–W222. https://doi.org/10.1093/nar/gkr363.

[20] C.H. Rodrigues, D.E. Pires, D.B. Ascher, DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability, Nucleic Acids Res. 46 (2018) W350–W355. https://doi.org/10.1093/nar/gky300.

[21] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins: Structure, Function, and Bioinformatics. 62 (2005) 1125–1132. https://doi.org/10.1002/prot.20810.

[22] P. Ranjan, P. Das, An inclusive study of deleterious missense PAX9 variants using user-friendly tools reveals structural, functional alterations, as well as potential therapeutic targets., Int J Biol Macromol. 233 (2023) 123375. https://doi.org/10.1016/j.ijbiomac.2023.123375.

[23] B. Yariv, E. Yariv, A. Kessel, G. Masrati, A. Ben Chorin, E. Martz, I. Mayrose, T. Pupko, N. Ben-Tal, Using evolutionary data to make sense of macromolecules with a “face-lifted” ConSurf, Protein Science. 32 (2023). https://doi.org/10.1002/pro.4582.

[24] J.M. Walker, ed., The Proteomics Protocols Handbook, Humana Press, Totowa, NJ, 2005. https://doi.org/10.1385/1592598900.

[25] B. Tam, S. Sinha, S.M. Wang, Combining Ramachandran plot and molecular dynamics simulation for structural-based variant classification: Using TP53 variants as model, Comput Struct Biotechnol J. 18 (2020) 4033–4039. https://doi.org/10.1016/j.csbj.2020.11.041.

[26] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler, The Human Genome Browser at UCSC, Genome Res. 12 (2002) 996–1006. https://doi.org/10.1101/gr.229102.

[27] Online Mendelian Inheritance in Man, (n.d.). https://omim.org/ (accessed July 6, 2023).

[28] M. Cariaso, G. Lennon, SNPedia: a wiki supporting personal genome annotation, interpretation and analysis, Nucleic Acids Res. 40 (2012) D1308–D1312. https://doi.org/10.1093/nar/gkr798.

[29] J. Navarro Gonzalez, A.S. Zweig, M.L. Speir, D. Schmelter, K.R. Rosenbloom, B.J. Raney, C.C. Powell, L.R. Nassar, N.D. Maulding, C.M. Lee, B.T. Lee, A.S. Hinrichs, A.C. Fyfe, J.D. Fernandes, M. Diekhans, H. Clawson, J. Casper, A. Benet-Pagès, G.P. Barber, D. Haussler, R.M. Kuhn, M. Haeussler, W.J. Kent, The UCSC Genome Browser database: 2021 update, Nucleic Acids Res. 49 (2021) D1046–D1057. https://doi.org/10.1093/nar/gkaa1070.

[30] UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly, (n.d.). http://genome.ucsc.edu (accessed July 11, 2023).

[31] N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence, Proteomics. 4 (2004) 1633–1649. https://doi.org/10.1002/pmic.200300771.

[32] A. Faccenda, C.A. Bonham, P.O. Vacratsis, X. Zhang, B. Mutus, Gold Nanoparticle Enrichment Method for Identifying *S* -Nitrosylation and *S* -Glutathionylation Sites in Proteins, J Am Chem Soc. 132 (2010) 11392–11394. https://doi.org/10.1021/ja103591v.

[33] Y. Xue, Z. Liu, X. Gao, C. Jin, L. Wen, X. Yao, J. Ren, GPS-SNO: Computational Prediction of Protein S-Nitrosylation Sites with a Modified GPS Algorithm, PLoS One. 5 (2010) e11290. https://doi.org/10.1371/journal.pone.0011290.

[34] Q. Zhao, Y. Xie, Y. Zheng, S. Jiang, W. Liu, W. Mu, Z. Liu, Y. Zhao, Y. Xue, J. Ren, GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs, Nucleic Acids Res. 42 (2014) W325–W330. https://doi.org/10.1093/nar/gku383.

[35] P. Radivojac, V. Vacic, C. Haynes, R.R. Cocklin, A. Mohan, J.W. Heyen, M.G. Goebl, L.M. Iakoucheva, Identification, analysis, and prediction of protein ubiquitination sites, Proteins: Structure, Function, and Bioinformatics. 78 (2010) 365–380. https://doi.org/10.1002/prot.22555.

[36] W. Deng, C. Wang, Y. Zhang, Y. Xu, S. Zhang, Z. Liu, Y. Xue, GPS-PAIL: prediction of lysine acetyltransferase-specific modification sites from protein sequences, Sci Rep. 6 (2016) 39787. https://doi.org/10.1038/srep39787.

[37] W. Ning, P. Jiang, Y. Guo, C. Wang, X. Tan, W. Zhang, D. Peng, Y. Xue, GPS-Palm: a deep learning-based graphic presentation system for the prediction of *S* -palmitoylation sites in proteins, Brief Bioinform. 22 (2021) 1836–1847. https://doi.org/10.1093/bib/bbaa038.

[38] C. Wang, X. Tan, D. Tang, Y. Gou, C. Han, W. Ning, S. Lin, W. Zhang, M. Chen, D. Peng, Y. Xue, GPS-Uber: a hybrid-learning framework for prediction of general and E3-specific lysine ubiquitination sites, Brief Bioinform. 23 (2022). https://doi.org/10.1093/bib/bbab574.

[39] W. Deng, Y. Wang, L. Ma, Y. Zhang, S. Ullah, Y. Xue, Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins, Brief Bioinform. (2016) bbw041. https://doi.org/10.1093/bib/bbw041.

[40] K. Shui, C. Wang, X. Zhang, S. Ma, Q. Li, W. Ning, W. Zhang, M. Chen, D. Peng, H. Hu, Z. Fang, A. Guo, G. Gao, M. Ye, L. Zhang, Y. Xue, Small-sample learning reveals propionylation in determining global protein homeostasis, Nat Commun. 14 (2023) 2813. https://doi.org/10.1038/s41467-023-38414-8.

[41] de Groot, Robert A., J. Nadrchal, Physics Computing '92: Proceedings of the 4th International Conference , in: Physics Computing '92, WORLD SCIENTIFIC, 1992.

[42] H. Venselaar, T.A. te Beek, R.K. Kuipers, M.L. Hekkelman, G. Vriend, Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, BMC Bioinformatics. 11 (2010) 548. https://doi.org/10.1186/1471-2105-11-548.

[43] C.C. Weihl, S. Dalal, A. Pestronk, P.I. Hanson, Inclusion body myopathy-associated mutations in p97/VCP impair endoplasmic reticulum-associated degradation, Hum Mol Genet. 15 (2006) 189–199. https://doi.org/10.1093/hmg/ddi426.

[44] F. Kherif, A. Latypova, Principal component analysis, in: Mach Learn, Elsevier, 2020: pp. 209–225. https://doi.org/10.1016/B978-0-12-815739-8.00012-2.