Biochemistry and Biophysics Reports 29 (2022) 101218

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

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

In-silico analysis of nonsynonymous genomic variants within *CCM2* gene reaffirm the existence of dual cores within typical PTB domain

Akhil Padarti [a,b], Ofek Belkin [a], Johnathan Abou-Fadel [a], Jun Zhang [a,*]

[a] Departments of Molecular & Translational Medicine (MTM), Texas Tech University Health Science Center El Paso (TTUHSCEP), El Paso, TX, 79905, USA
[b] Department of Neurology, University of South Alabama School of Medicine, 5795 USA N Dr., Mobile, AL, 36608, USA

ARTICLE INFO

Keywords:
In-silico analysis
Tertiary structure
Superimposition of protein structures
Amino acid substitution
Single nucleotide polymorphisms (SNPs)
Nonsynonymous single nucleotide polymorphisms (nsSNPs)

ABSTRACT

Purpose: The objective of this study is to validate the existence of dual cores within the typical phosphotyrosine binding (PTB) domain and to identify potentially damaging and pathogenic nonsynonymous coding single nuclear polymorphisms (nsSNPs) in the canonical PTB domain of the CCM2 gene that causes cerebral cavernous malformations (CCMs).

Methods: The nsSNPs within the coding sequence for PTB domain of human CCM2 gene, retrieved from exclusive database searches, were analyzed for their functional and structural impact using a series of bioinformatic tools. The effects of mutations on the tertiary structure of the PTB domain in human CCM2 protein were predicted to examine the effect of nsSNPs on the tertiary structure of PTB Cores.

Results: Our mutation analysis, through alignment of protein structures between wildtype CCM2 and mutant, predicted that the structural impacts of pathogenic nsSNPs is biophysically limited to only the spatially adjacent substituted amino acid site with minimal structural influence on the adjacent core of the PTB domain, suggesting both cores are independently functional and essential for proper CCM2 PTB function.

Conclusion: Utilizing a combination of protein conservation and structure-based analysis, we analyzed the structural effects of inherited pathogenic mutations within the CCM2 PTB domain. Our results predicted that the pathogenic amino acid substitutions lead to only subtle changes locally, confined to the surrounding tertiary structure of the PTB core within which it resides, while no structural disturbance to the neighboring PTB core was observed, reaffirming the presence of independently functional dual cores in the CCM2 typical PTB domain.

1. Introduction

Scaffold proteins have essential roles in various pivotal cellular signaling cascades [1]. One recurring domain shared by scaffolding proteins is the phosphotyrosine binding (PTB) domain [2,3]. As the second-largest family of phosphotyrosine recognition domains, the PTB domain was shown to evolve from pleckstrin homology (PH) domains, as both PTB and PH domains are structurally and functionally comparable sharing the PH superfolds [4]. Contrary to the singular binding pocket in PH and FERM domains, our lab discovered that the full-length typical PTB domain contains two equal, unique, versatile, and independent binding pockets (PTB dual cores), allowing the domain to bind to multiple NPXY motifs simultaneously present in the cytoplasmic tails of membrane receptors [5]. In this report, we utilized confirmed genetic data to validate the existence of PTB dual cores with biological phenotypes and corresponding genetic data, which can help refine PTB domain binding partners within the cell.

CCM2 is a PTB domain-containing protein that binds to CCM1 through either one [6] or two [7] NPXY motifs. CCM1 and CCM3 bind to CCM2, which serves as a docking site, to form the CCM signaling

Abbreviations: PTB, phosphotyrosine binding; PH, pleckstrin homology; CSC, CCM signaling complex; CCMs, cerebral cavernous malformations; PDB, protein data bank; nsSNP, nonsynonymous single nucleotide polymorphism; INDELs, insertions/deletions; PTCs, premature termination codons; SIFT, Sorting Intolerant From Tolerant; PANTHER, Protein ANalysis THrough Evolutionary Relationship; PROVEAN, Protein Variation Effect Analyzer; POLYPHEN-2, Polymorphism Phenotyping; HOPE, Have (y)Our Protein Explained; CUPSAT, Cologne University Protein Stability Analysis Tool; MAF, minor allele frequency; I-TASSER, the iterative threading assembly refinement.

* Corresponding author. Department of Molecular and Translational Medicine (MTM), Texas Tech University Health Science Center El Paso, 5001 El Paso Drive, El Paso, El Paso, TX, 79905, USA.

E-mail address: jun.zhang2000@gmail.com (J. Zhang).

https://doi.org/10.1016/j.bbrep.2022.101218
Received 3 December 2021; Received in revised form 11 January 2022; Accepted 21 January 2022

complex (CSC) in combination with other CSC members, and plays a key role in multiple essential cellular processes [8]. To validate the existence of two functional binding pockets in the CCM2 typical PTB domain, which is used to bind CCM1, our focus in this report will be solely on CCM1/CCM2 interactions via the CCM2 typical PTB domain. Several mutations that disrupt the CCM1/CCM2 interaction have been implicated in cerebral cavernous malformations (CCMs), an autosomal dominant disease with incomplete penetrance, indicating even individuals that carry the same mutation in the same family have variable clinical outcomes; some will develop CCMs while others will never show any clinical CCM symptoms. Mutations on both alleles of one of the CCM genes have been reported only in CCM lesions [8]. While it has been experimentally shown that both PTB cores (Core1 and Core2) in CCM2 are independently capable of binding to NPXY motifs on CCM1 [9], the biological relevance of dual PTB cores remains elusive. To date, more than 150 distinct pathogenic *CCM1/CCM2/CCM3* germline mutations causing CCMs (OMIM 116860) have been implicated in 87–98% of familial CCMs [10]. Approximately 80 genomic variants have been reported in the CCM2 gene, and half of them are nonsense loss of function mutations or frameshifting-insertions/deletions (INDELs) leading to premature termination codons (PTCs) [11].

In general, amino acid substitutions can affect protein function through altered folding/stability, aberrant post-translational modifications, disruption of domain functions, and/or affecting sites of interaction [12]. Experimental methods for protein structure determination have been established with X-ray crystallography and NMR spectroscopy, generating substantial structural data. With an enormous amount of protein structure information in protein data bank (PDB), various computational approaches have been developed to model the 3D structure of a protein with a known amino acid sequence, a type of *in-silico* analysis [13–15]. Since it is unclear whether both CCM2 PTB cores are necessary and essential for CCM2 functionality, we believe that the effect of a nonsynonymous single nucleotide polymorphism (nsSNP) at the protein level should be studied with *in-silico* methodology to predict the effect of single amino acid substitutions during CCMs pathogenesis, and to validate the existence of functional dual PTB cores in CCM2. This will be accomplished through phenotype/genotype correlation during CCMs pathogenesis to evaluate the functionality of CCM2 PTB cores.

2. Materials and methods

Identification of nonsynonymous genomic variants within the CCM2 gene. To thoroughly investigate the clinical relevance of the nonsynonymous genomic variants within the CCM2 gene, we searched well-known databases from HGMD (the Human Gene Mutation Database, http://www.hgmd.cf.ac.uk), ClinVar (a public archive with interpretations of clinically relevant variants, https://www.ncbi.nlm.nih.gov/clinvar/), gnomAD (the Genome Aggregation Database, https://gnomad.broadinstitute.org), the 1000 Genomes Project (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes), dbSNP (the Single Nucleotide Polymorphism database, https://www.ncbi.nlm.nih.gov/snp/), OMIM (the Online Mendelian Inheritance in Man, https://www.omim.org), Angioma Alliance (https://www.angioma.org/), and ESP (NHLBI Exome Sequencing Project, https://evs.gs.washington.edu/EVS/).

Bioinformatic tools for in-silico analysis. In order to precisely predict the impact of nonsynonymous genomic variants within the CCM2 protein, we utilized well-known bioinformatic tools, which can be categorized into three major groups: protein conservation-based, protein structure-based analysis [16], or a combination thereof, with a selected array of *in-silico* predictive algorithms to evaluate the genomic variants [16,17]. For evolutionary conservative approaches, there are multiple programs such as SIFT (Sorting Intolerant From Tolerant, https://sift.bii.a-star.edu.sg/www/Extended_SIFT_chr_coords_submit.html) [18], PANTHER (Protein ANalysis THrough Evolutionary Relationship, http://www.pantherdb.org/tools/) [19], and MUTATION ASSESSOR (http://mutationassessor.org/r3/). For homolog modelling (sequence similarities/alignment), we utilized PROVEAN (Protein Variation Effect Analyzer, https://provean.jcvi.org/index.php) [20]. For protein structure/function using an evolutionary conservation-based approach, we selected POLYPHEN-2 (Polymorphism Phenotyping Ver. 2.0, https://genetics.bwh.harvard.edu/pph2/) [21], and MUTATION TESTER (http://www.mutationtaster.org/) [22]. For protein structure stability measurements, we utilized MUPRO (http://mupro.proteomics.ics.uci.edu/), I-MUTANT (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) [23], HOPE (Have (y)Our Protein Explained, https://www3.cmbi.umcn.nl/hope/method/), and CUPSAT (Cologne University Protein Stability Analysis Tool, http://cupsat.tu-bs.de/) [24] with different parameters to define nonsynonymous variants. The minor allele frequency (MAF) was acquired for each nsSNP from the SNP database. MAF represents the incidence of the gene variant in the general population. We hypothesize if MAF of one nsSNP is less than the overall prevalence of symptomatic CCMs in the general population (0.04%), this nsSNP is likely to be pathogenic.

Bioinformatic tools for tertiary structural modeling. Currently, there is only one X-ray crystallography structure of CCM2 PTB domain (bound with an NPXY motif ligand) deposited in PDB (4WJ7) [6]. To better serve our purpose, we utilized MODELLER (https://salilab.org/modeller/9.16/release.html), which uses homology modeling with *ab initio* methods producing solutions that satisfy a set of spatial rules derived from probability density functions and statistical analysis of PTB domain containing protein structures, deposited in PDB [25]. One limitation of using MODELLER, is that the program depends on the deposited x-ray structure data of the CCM2 PDB domain, which is bound to a ligand in the C-terminus, resulting in the predicted structure of CCM2 PTB domain being truncated in the C-terminus. Therefore, MODELLER was unable to generate nsSNPs encompassing the C-terminus of the CCM2 PTB structure (such as the A179S mutant). Comparatively, we also used an integrated platform I-TASSER (The iterative threading assembly refinement, https://zhanglab.dcmb.med.umich.edu/I-TASSER/), which is an automated protein structure and functional prediction software based on the sequence-to-structure-to-function paradigm from multiple threading alignments to perform iterative structural assembly simulations [26]. Three-dimensional (3D) atomic models generated by either I-TASSER or MODELLER were then visualized and compared by molecular visualization software, PYMOL (http://www.pymol.org/), CHIMERA (http://www.cgl.ucsf.edu/chimera) [25], and RASWIN (http://www.openrasmol.org/). This process was performed for wildtype CCM2 and identified CCM2 mutants. The tertiary structure of the two proteins were superimposed and analyzed for any structural differences.

3. Results

3.1. Define genomic variants within the PTB domain of CCM2

Total number of genomic variants within CCM2 PTB domain. By searching all available databases, 66 nsSNPs in 49 amino acid positions were identified in the PTB domain of CCM2, in addition, two in-frame deletions in exon 2 were also identified, making a total of 68 nonsynonymous genomic variants within the PTB domain (Suppl. Table 1). The interpretations of our *in-silico* analysis for all nsSNPs are shown (Table 1, Suppl. Table 1). Among known genetic mutants and our selected candidate mutants, there is consistent agreement among predicted results through our various *in-silico* analysis (Table 1).

3.2. Genomic variants with in-frame deletions leading to conformational changes in both PTB cores

Large Exon 2 deletion [p.(Pro11_Lys68del)]. This 58 amino acid in-frame deletion was one of the first discovered genomic variants in CCM2 mutation screening [27,28], and has been consistently observed [29,30]. Experimental data showed this in-frame deletion abolishes the

Table 1

Pathogenic nsSNPs in both cores of CCM2 PTB domain. The known pathogenic and several potential pathogenic nsSNPs of CCM2 PTB domain are presented. The nsSNPs were chosen based on genetic results from familial CCM cases and high probability of the predicted pathological nature of the mutation. All 66 recorded mutations within CCM2 PTB domain are shown in the Suppl.mentary Table 1. The secondary structural motif and the core location of each substitution is shown. The SNP nomenclature and MAF for known mutations are also shown if available. Each nsSNP were further evaluated with various *in-silico* tools for pathogenicity. The reference for each reported pathogenic nsSNP in human genetic study evidenced as phenotype/genotype correlation is also provided.

Mutation	Exon number	PTB Core	Secondary structure	SNP nomenclature	MAF	SIFT	MUTATION ASSESSOR	PANTHER	CUPSAT	MUPRO	HOPE	I-MUTANT-2.0	PROVEAN	POLYPHEN-2	MUTATION TESTER	References
I76T	3	Core1	β1-α2	rs756431644	8.00E-06	Functional	low	probably benign	Destabilising	Decrease stability	mutation can disturb this domain and abolish its function	Destabilising	Deleterious	probably damaging	Disease-Causing	
A98T	4	Core1	α2	rs780867674	1.70E-05	Tolerated	low	probably damaging	Stabilising	Decrease stability	mutation can disturb this domain and abolish its function	Destabilising	Neutral	benign	POLYMORPHISM	
A111P	4	Core1	β2	rs750889112	4.00E-06	Tolerated	medium	probably damaging	Destabilising	Decrease stability	might disturb the core structure of this domain.	Stabilising	Deleterious	probably damaging	Deleterious	[35]
L113P	4	Core1	β2	rs11552377	0.06	Functional	low	probably benign	Stabilising	Decrease stability	might disturb the core structure of this domain.	Destabilising	Deleterious	likely damaging	Deleterious	[34]
L115R	4	Core1	β2	N/A	N/A	Functional	medium	probably damaging	Destabilising	Decrease stability	residue is located near a highly conserved position	Destabilising	Deleterious	probably damaging	Deleterious	[6]
V120I	4	Core1	β3	rs11552377	0.175	Tolerated	low	probably benign	Stabilising	Decrease stability	polymorphism	Destabilising	Neutral	benign	Benign	[40]
V120D	4	Core1	β3	rs745788686	4.00E-06	Functional	medium	probably benign	Destabilising	Decrease stability	might be damaging to the protein and abolish its function	Destabilising	Deleterious	probably damaging	Deleterious	[36]
A141T	4	Core2	β5	rs1562908094	8.00E-06	Functional	medium	probably damaging	Destabilising	Decrease stability	might be damaging to the protein and abolish its function	Destabilising	Deleterious	likely damaging	Deleterious	
R146W	4	Core2	β5	rs769929401	4.80E-05	Functional	medium	probably damaging	Destabilising	Decrease stability	residue is located near a highly conserved position	Destabilising	Deleterious	benign	Disease-Causing	
L152 M	4	Core2	β6	rs760117074	8.00E-06	Functional	medium	probably damaging	Destabilising	Decrease stability	might be damaging to the protein and abolish its function	Destabilising	Neutral	likely damaging	Deleterious	
V154G	4	Core2	β6	rs141353947	4.00E-06	Functional	medium	probably damaging	Destabilising	Decrease stability	might be damaging to the protein and abolish its function	Destabilising	Deleterious	probably damaging	Deleterious	

(*continued on next page*)

Table 1 (*continued*)

Mutation	Exon number	PTB Core	Secondary structure	SNP nomenclature	MAF	SIFT	MUTATION ASSESSOR	PANTHER	CUPSAT	MUPRO	HOPE	I-MUTANT-2.0	PROVEAN	POLYPHEN-2	MUTATION TESTER	References
L155P	4	Core2	β6	rs373239614	4.00E-06	Functional	medium	probably damaging	Stabilising	Decrease stability	might be damaging to the protein and abolish its function	Destabilising	Deleterious	probably damaging	Deleterious	[6]
A179S	5	Core2	α3	rs373136857	1.60E-05	Tolerated	neutral	probably benign	Destabilising	Decrease stability	might be damaging to the protein and abolish its function	Destabilising	Neutral	benign	POLYMORPHISM	

interaction between CCM1/CCM2, indicating this portion of the peptide sequence plays an important role in this interaction [28]. Our analysis indicates that the deletion encompasses a portion of the β1 strand within Core1 (Fig. 1A, Suppl. Video 1A). β-Sheet 1 contains only 3 β-strands in the deletion mutant, one fewer than the wildtype. The majority of the PTB domain remains overlapping between the mutant and the wildtype PTB domain, however, the α3 helix in Core2 is surprisingly distorted indicating that this large deletion results in overwhelming structural changes to both Core1 and Core2 of the CCM2 PTB domain.

In-frame four amino acid deletions within Exon 2 [p.(Lys65_Lys68del)]. Another smaller four amino acid in-frame deletion in the same region of the PTB domain was also reported in two Japanese familial CCM cases [31], indicating the importance of this region for the PTB domain. In this deletion, only the N-terminal portion of Core1 is affected with intact upstream flanking sequences. Our analysis indicate that these four amino acids are in the β1 strand between Core1 and Core2 (Fig. 1B, Suppl. Video 1B). Similarly in the deletion mutant, β-Sheet 1 contains only 3 β-strands, one fewer than the wildtype, while β-Sheet 2 contains 2 β-strands, two fewer than the wildtype. While the majority of the PTB domain between the wildtype and mutant remains aligned, the α3 helix is misaligned in the mutant, similar to the large exon 2 deletion. It is interesting that this four amino acid deletion alters the structure of the β-barrel more severely than the large 58 amino acid deletion, suggesting that these four amino acids may be essential for Core1 function. The structural changes resulting from this deletion encompasses both Core1 and Core 2 of the PTB domain.

3.3. Genomic variants with nsSNPs only have local effects

nsSNPs in Core2. CCM disease associated with nsSNPs in Core2, such as L198R and L213P, were among the first discovered genomic variants identified in CCM mutation screening [32–34]. Six nsSNPs/mutations (all MAFs <4e-4, symptomatic CCM incidence in general population) found in Core2 of CCM2 are modelled with I-TASSER and MODELLER (A141T, R146W, L152 M, V154G, L155P, A179S, Fig. 2, Suppl. Fig 1, Suppl. Video 2). As aforementioned, MODELLER generated the tertiary structures with a truncated C-terminus, which resulted in being unable to generate a valid structure for the A179S mutant (C-terminal). A141T (Fig. 2, Suppl. Fig. 1A, Suppl. Videos 2A, 2E) is predicted to be a destabilising/unfavorable, and rare nsSNP (MAF <10e-5). This mutation is in the β5 strand, with tertiary structure changes encompassing β5 and a portion of the β5/β6 loop, due to the polar side chain of threonine relative to alanine in the wildtype. L152 M (Fig. 2, Suppl. Fig. 1B, Suppl. Videos 2B, 2F) and V154G (Fig. 2, Suppl. Fig. 1C, Suppl. Videos 2C, 2G) are two other destabilising/unfavorable, and rare nsSNPs (MAF< 10e-4) encompassing the β6 strand and these structural perturbations in both mutations are limited to the β6 strand only. The amino acid side chains of both CCM2 mutants are more polar than the side chains of the wildtype CCM2. L155P (Fig. 2, Suppl. Fig. 1D, Suppl. Videos 2D, 2H) is a known CCM2 mutation resulting in CCM phenotype [6]. This mutation is in the β6 strand, resulting in structural changes in the β6 strand and β6/β7 loop. While leucine and proline both have hydrophobic side chains, proline results in a peptide backbone twist resulting in local structural disturbance. All six Core2 nsSNPs/mutations (A141T, R146W, L152 M, V154G, L155P, A179S) are predicted to show only local structural alterations, but none of these mutations lead to backbone distortion or have structural effects on the neighboring Core1 of CCM2 PTB domain (Fig. 2).

nsSNPs in Core1. Interestingly, more pathogenic nsSNPs have been reported in the N-terminus of the PTB domain corresponding to Core1 than Core2. Six nsSNPs/mutations in Core1 have been modelled with I-TASSER and MODELLER (I76T, A98T, A111P, L113P, L115R, V120D) (Fig. 3, Suppl. Fig. 2, Suppl. Video 3). A111P (Fig. 3, Suppl. Fig. 2A, Suppl. Videos 3A, 3E) is a pathogenic mutation in the β2 strand (β-sheet 1), with minimal overlap between the amino acid side chains in both tertiary structure modalities [35]. There is misalignment in the β2, β2/3

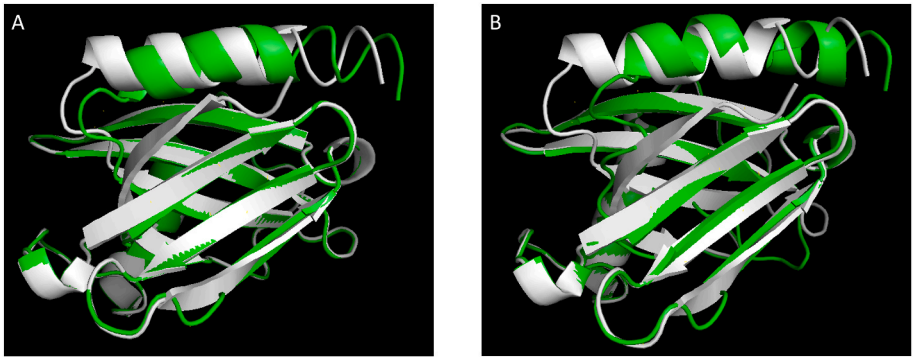

Fig. 1. N-terminus in-frame deletion in Core 1 leads to conformational changes in both cores of CCM2 PTB domain. Deletions were modelled by I-TASSER and alignments were visualized using PYMOL. wildtype CCM2 (white) and mutant (green) are shown. The α3 helix is shown superior and α2 helix is shown inferior, while β-sheet 1 is shown anterior to β-sheet 2. **A)** 58 amino acid deletion in core1 mutant is illustrated. **B)** In-frame 4 amino acid deletion mutant (65-KEVK-68) is illustrated. Although the β-sheet overlaps between the wildtype and mutant, the α3 helix is shown to be mismatched indicating conformational change in the C-terminus. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

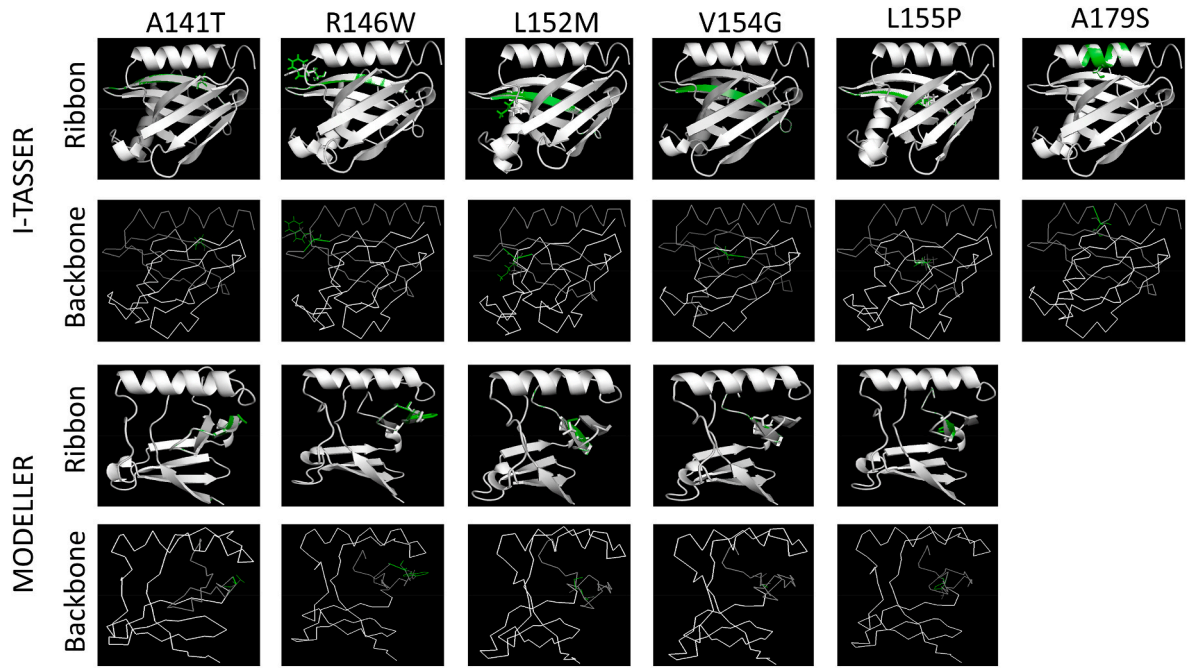

Fig. 2. nsSNPs in Core 2 lead to local disturbance of substituted amino acids without perturbing 3D conformation in Core 1 of CCM2 PTB domain in I-TASSER and MODELLER. For I-TASSER 3D structure orientation (Row 1, 2), the α3 helix is shown superior and the α2 helix is shown inferior, while β-sheet 1 is shown anterior to β-sheet 2. For MODELLER 3D structure orientation (Row 3, 4), the α2 helix is shown superior, while the full β-sheet 1 is shown inferior and the partial β-sheet 2 is shown on the right. For the ribbon models (Row 1, 3), the nsSNP mutant (green) is superimposed on the wildtype (white) in the ribbon conformation while the substituted amino acids are shown in stick configuration to demonstrate local conformational change. For the backbone models (Row 2, 4), the same nsSNP is superimposed on the wildtype in backbone conformation with mutant amino acids in line conformation to explore any possible peptide backbone distortion. For the backbone models (Row 2, 4), each core is shown in the different color: wildtype core1 (white), wildtype core2 (gray), mutant core1 (red), and mutant core2 (green) to highlight the two PTB cores (both red and green backbone from mutant can only be visualized if there is a distortion between mutant and wildtype). Several 3D PDB images are also provided as Suppl.ments. The nsSNPs are A141T (first column), R146W (second column), L152 M (third column), V154G (fourth column), L155P (fifth column), A179S (last column). The C-terminal portion (α3 helix and partial β-sheet 2) of the CCM2 PTB domain model is absent due to the CCM2 x-ray crystallographic structure data in the PDB database which emphasizes ligand binding in the C-terminal PTB core2 and for the same reason, A179S mutant is unable to be generated with MODELLER. Only subtle local disturbance was seen surrounding the substituted amino acids and no amino acid backbone distortion was observed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

loop, and α1/β2 loop between the two structures. Proline is more sterically bulky than alanine and possesses a larger local twist in the peptide backbone, resulting in a shortened β2 strand. L113P (Fig. 3, Suppl. Fig. 2B, Suppl. Videos 3B, 3F) is a known mutation that results in familial CCMs [34] that is present in the β2 strand, with predicted structural perturbations in the β1/β2 strands and β2/β3 loop. The side chain of proline results in a backbone peptide chain twist resulting in delayed formation of the β2 strand. L115R (Fig. 3, Suppl. Fig. 2C, Suppl. Videos 3C, 3G), a known pathological CCM mutation [6], resides in the β2 strand, with all predicted structural differences limited to the β2 strand and β2/β3 loop. Arginine has a charged side chain and all hydrophobic interactions with wildtype leucine are severed. V120D (Fig. 3, Suppl. Fig. 2D, Suppl. Videos 3D, 3H), a known mutation resulting in a Japanese familial CCM pedigree [36], is located in the β3 strand, with all structural differences limited to the β3 strand and β3/4 loop. Similar to L115R, V120D modelling predicts inhibition of all hydrophobic interactions of valine with the charged amino acid side chain of aspartate. In sum, like Core2, all six nsSNPs/mutations within Core1 (I76T, A98T, A111P, L113P, L115R, V120D) are predicted to show only local structural alterations, but none of these mutations lead to backbone distortion or have structural disruption on the adjacent Core2 of the PTB domain (Fig. 3).

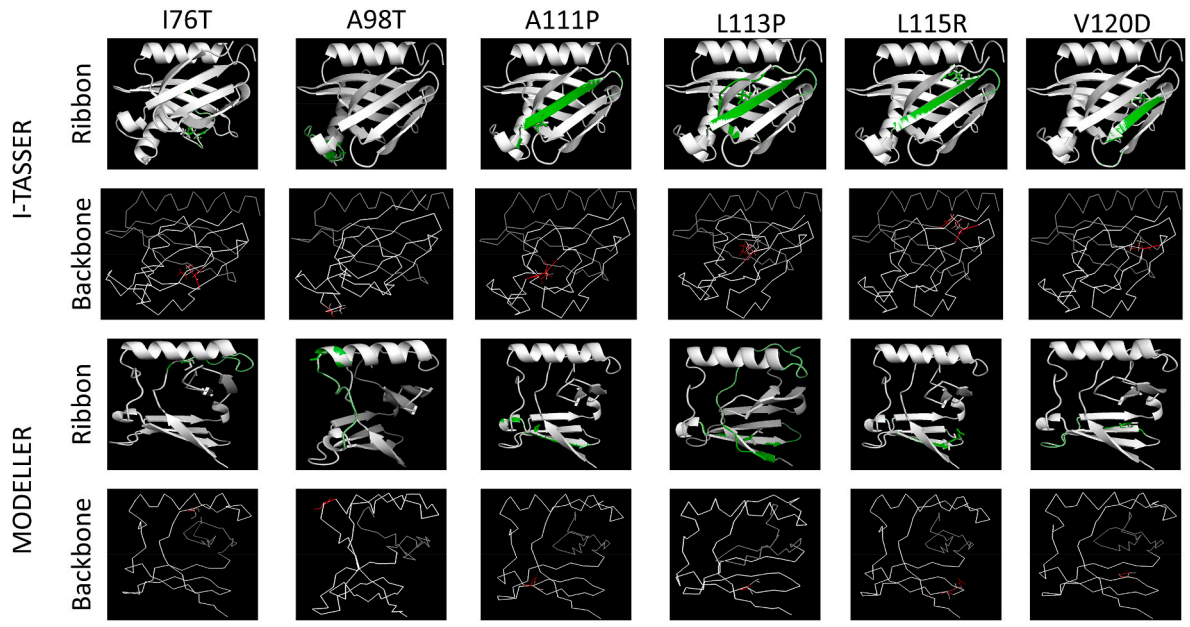

Fig. 3. nsSNPs in Core 1 lead to local disturbance of substituted amino acids without perturbing 3D conformation in Core 2 of CCM2 PTB domain in I-TASSER and MODELLER. . Diagram layout is similar to Fig. 2, I-TASSER (Row 1, 2) and MODELLER
(Row 3, 4) 3D structure orientation, ribbon (Row 1, 3) and backbone (Row 2, 4) are displayed in the same fashion. The nsSNPs are I76T (first column), A98T (second column), A111P (third column), L113P (fourth column), L115R (fifth column), V120D (last column). Similar to Core2 mutants, only subtle local disturbance was seen surrounding the substituted amino acids and no amino acid backbone distortion was observed.

In-silico analysis of a common genomic variant within the CCM2 PTB domain questions its role as a causative mutation. It is interesting to note that nsSNP (rs11552377, (c.358G > A, NM_ 001167935, p.Val120Ile)) has been reported as either pathogenic [29,36–40] and associated with increased risk of CCMs [41,42], or predicted to be at the "benign" end of the CCM phenotypic spectrum [37]. However, some data suggest this nsSNP affects splicing [41,43], despite its allele frequency (MAF range 0.13–0.17, Table 1) being much higher than the incidence of CCMs. Furthermore, various *in-silico* testing methodologies of V120I consistently concluded it is benign/polymorphic (Table 1), challenging previous mutational screening reports. Our *in-silico* analysis indicates V120I mutation is in the β3 strand in β-sheet 1 of Core1 (Suppl. Figs. 3, 4, Suppl. Video 4). The associated tertiary structure perturbation is limited to the β3 strand and β3/4 loop, which is the more flexible region of the PTB domain. The side chains of both valine and isoleucine are overlapping, similar sterically and equally hydrophobic, maintaining the hydrophobic side chain interactions. Therefore, we conclude this mutation (V120I) is polymorphic and not pathogenic. Further mutation screening in newly described novel coding exons of CCM2 [9] might be warranted for these familial CCM cases.

4. Discussion

In-silico analysis of genetic mutations among CCM genes. Since identification of causative genes of CCMs, there have been many attempts to utilize *in-silico* analysis with bioinformatics tools to interpret the genetic variants identified within CCM genes [29,30,34,41,42,44–47]. However, the majority of these targeted genetic variants among three known CCM genes are either nonsense mutations or frame-shift mutations, making the outcomes of the *in-silico* analysis irrelevant to protein structural investigations. To date, one protein tertiary structure of CCM2 PTB domain binding to NPXY motif has been deposited in PDB (4WJ7), determined by x-ray crystallography [6]. Although this work was primarily focused on the traditionally recognized PTB functional pocket, Core2, it provided the foundation for one of our protein structural simulation programs, probability density functions-based MODELLER [25], making our novel integrated *in-silico*/structural simulation analysis more thorough. Furthermore, recent efforts for *in-silico* analysis of Core1 in a genetic variant that is co-segregated in a large Chinese CCM pedigree [35] provide us with additional strong evidence for supporting this methodology. On the structural level, each amino acid substitution frequently resulted in perturbations within the tertiary structure of the PTB domain. These perturbations were due to disruption of the electrostatic/hydrophobic interactions, yet the structural alteration was limited to the surrounding region of the single amino acid substitution within the PTB domain. The two in-frame deletions are the exception to this observance. Both in-frame N-terminal deletions resulted in alteration of both Core1 and Core2, strongly suggesting that the β1 strand/α2 helix at the N terminus may be equally essential for maintaining structural stability of the entire PTB domain as the C terminus, in contrast with previous reports [6,48]. Our combined structure and *in-silico* bioinformatics analyses provided strong evidence that nsSNPs in Core1 did not disturb neighboring Core2 and vice versa, indicating that both binding pockets have independent functional roles in their interactions with NPXY motifs, and dysfunction of either binding pocket is sufficient to initiate pathogenesis of CCMs.

In-silico analysis provides novel candidates for future CCM2 mutational screening and assists in excluding some nsSNPs as causal mutations. A significant portion of the CCM patients have no known causative mutations identified [41]. It may be difficult to differentiate between non-pathogenic and pathogenic nsSNPs in these patients. *In-silico* methodology can be used to identify high risk nsSNPs as potential causal mutations for these patients while ruling out certain nsSNPs as benign polymorphisms. Our *in-silico* analysis revealed that one missense mutation, V120I, currently considered a causal mutation of CCMs, is in fact a relatively common polymorphism and unlikely to result in phenotype. Similarly, this methodology also helped us to identify several potential candidate pathological nsSNPs along with known pathogenic nsSNPs previously reported (Table 1). Those candidate pathogenic nsSNPs can be further evaluated through MAF and their performance using *in-silico* analytical tools. In sum, our integrated structural and *in-silico* analysis will be useful in future CCM mutational screenings to identify pathogenic mutations and excluding normal variants (Suppl. Table 1).

5. Conclusion

This analysis predicted that both PTB-cores in CCM2 have independent functional binding pockets and mutations in either binding pocket can result in CCM phenotype without disrupting the conformation of the neighboring core, validating our PTB dual core theory. One important limitation to our methodology is analyzing the extent of cooperative binding between the two PTB cores. It is rather likely that cooperative binding between the two PTB cores exists, optimizing the binding ability of the CCM2 PTB domain. However, the extent of this cooperative binding cannot be evaluated through *in-silico* analysis. Nonetheless with the advent of robust computational tools, *in-silico* analysis is a proven valid and robust methodology for analyzing protein domains. Our integrated approach can also be utilized to verify potential pathogenic mutations in subsequent mutant screening analyses. Future efforts will be made to further explore the dual core nature of the PTB domains and to identify potential unique cellular partners for each binding pocket.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Structural/crystallographic, amino acid sequence data for PTB cores, computational models, and genetic polymorphisms (nsSNPs) data are either available in public database, or in Suppl.mental materials, or provided by correspondent author upon request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2022.101218.

References

[1] T. Kaneko, R. Joshi, S.M. Feller, S.S. Li, Phosphotyrosine recognition domains: the typical, the atypical and the versatile, Cell communication and signaling, CCS 10 (2012) 32.

[2] B. Margolis, J.P. Borg, S. Straight, D. Meyer, The function of PTB domain proteins, Kidney Int. 56 (1999) 1230–1237.

[3] J. Zhang, R.E. Clatterbuck, D. Rigamonti, D.D. Chang, H.C. Dietz, Interaction between krit1 and icap1alpha infers perturbation of integrin beta1-mediated angiogenesis in the pathogenesis of cerebral cavernous malformation, Hum. Mol. Genet. 10 (2001) 2953–2960.

[4] A. Padarti, J. Abou-Fadel, J. Zhang, Resurgence of Phosphotyrosine Binding Domains: Structural and Functional Properties Essential for Understanding Disease Pathogenesis, Biochimica et Biophysica Acta - General Subjects, 2021. In press.

[5] J. Zhang, A. Padarti, X. Jiang, J. Abou-Fadel, Redefining PTB domain into independently functional dual cores, Biochem. Biophys. Res. Commun. 524 (2020) 595–607.

[6] O.S. Fisher, W. Liu, R. Zhang, A.L. Stiegler, S. Ghedia, J.L. Weber, T.J. Boggon, Structural basis for the disruption of the cerebral cavernous malformations 2 (CCM2) interaction with Krev interaction trapped 1 (KRIT1) by disease-associated mutations, J. Biol. Chem. 290 (2015) 2842–2853.

[7] J. Zhang, D. Rigamonti, H.C. Dietz, R.E. Clatterbuck, Interaction between krit1 and malcavernin: implications for the pathogenesis of cerebral cavernous malformations, Neurosurgery 60 (2007) 353–359. ; discussion 359.

[8] A. Padarti, J. Zhang, Recent advances in cerebral cavernous malformation research, Ves. Plus 2 (2018).

[9] X. Jiang, A. Padarti, Y. Qu, S. Sheng, J. Abou-Fadel, A. Badr, J. Zhang, Alternatively spliced isoforms reveal a novel type of PTB domain in CCM2 protein, Sci. Rep. 9 (2019) 15808.

[10] S. Spiegler, M. Rath, S. Hoffjan, P. Dammann, U. Sure, A. Pagenstecher, T. Strom, U. Felbor, First large genomic inversion in familial cerebral cavernous malformation identified by whole genome sequencing, Neurogenetics 19 (2018) 55–59.

[11] L. Yang, J. Wu, J. Zhang, A novel CCM2 gene mutation associated with cerebral cavernous malformation, Front. Neurol. 11 (2020) 70.

[12] D.G. Knorre, N.V. Kudryashova, T.S. Godovikova, Chemical and functional aspects of posttranslational modification of proteins, Acta Nat. 1 (2009) 29–51.

[13] B.I. Dahiyat, *In-silico* protein design: fitting sequence onto structure, Methods Mol. Biol. 316 (2006) 359–374.

[14] S. Abdulazeez, S. Sultana, N.B. Almandil, D. Almohazey, B.J. Bency, J.F. Borgio, The rs61742690 (S783N) single nucleotide polymorphism is a suitable target for disrupting BCL11A-mediated foetal-to-adult globin switching, PLoS One 14 (2019), e0212492.

[15] S. AbdulAzeez, J.F. Borgio, *In-silico* computing of the most deleterious nsSNPs in HBA1 gene, PLoS One 11 (2016), e0147702.

[16] R.C. Caswell, M.M. Owens, A.C. Gunning, S. Ellard, C.F. Wright, Using structural analysis *in-silico* to assess the impact of missense variants in MEN1, J. Endocr. Soc. 3 (2019) 2258–2275.

[17] M. Seifi, M.A. Walter, Accurate prediction of functional, structural, and stability changes in PITX2 mutations using *in-silico* bioinformatics algorithms, PLoS One 13 (2018).

[18] P. Kumar, S. Henikoff, P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, Nat. Protoc. 4 (2009) 1073–1081.

[19] P.D. Thomas, A. Kejariwal, N. Guo, H. Mi, M.J. Campbell, A. Muruganujan, B. Lazareva-Ulitsky, Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools, Nucleic Acids Res. 34 (2006) W645–W650.

[20] Y. Choi, A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, Bioinformatics 31 (2015) 2745–2747.

[21] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, Nat. Methods 7 (2010) 248–249.

[22] J.M. Schwarz, D.N. Cooper, M. Schuelke, D. Seelow, MutationTaster2: mutation prediction for the deep-sequencing age, Nat. Methods 11 (2014) 361–362.

[23] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005) W306–W310.

[24] V. Parthiban, M.M. Gromiha, D. Schomburg, CUPSAT: prediction of protein stability upon point mutations, Nucleic Acids Res. 34 (2006) W239–W242.

[25] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[26] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat. Methods 12 (2015) 7–8.

[27] C.L. Liquori, M.J. Berg, F. Squitieri, T.P. Leedom, L. Ptacek, E.W. Johnson, D. A. Marchuk, Deletions in CCM2 are a common cause of cerebral cavernous malformations, Am. J. Hum. Genet. 80 (2007) 69–75.

[28] S. Stahl, S. Gaetzner, K. Voss, B. Brackertz, E. Schleider, O. Surucu, E. Kunze, C. Netzer, C. Korenke, U. Finckh, M. Habek, Z. Poljakovic, M. Elbracht, S. Rudnik-Schoneborn, H. Bertalanffy, U. Sure, U. Felbor, Novel CCM1, CCM2, and CCM3 mutations in patients with cerebral cavernous malformations: in-frame deletion in CCM2 prevents formation of a CCM1/CCM2/CCM3 protein complex, Hum. Mutat. 29 (2008) 709–717.

[29] G. Nardella, G. Visci, V. Guarnieri, S. Castellana, T. Biagini, L. Bisceglia, O. Palumbo, M. Trivisano, C. Vaira, M. Scerrati, D. Debrasi, V. D'Angelo, M. Carella, G. Merla, T. Mazza, M. Castori, L. D'Agruma, C. Fusco, A single-center study on 140 patients with cerebral cavernous malformations: 28 new pathogenic variants and functional characterization of a PDCD10 large deletion, Hum. Mutat. 39 (2018) 1885–1900.

[30] C.D. Much, K. Schwefel, D. Skowronek, L. Shoubash, F. von Podewils, M. Elbracht, S. Spiegler, I. Kurth, A. Floel, H.W.S. Schroeder, U. Felbor, M. Rath, Novel pathogenic variants in a cassette exon of CCM2 in patients with cerebral cavernous malformations, Front. Neurol. 10 (2019) 1219.

[31] S. Tsutsumi, I. Ogino, M. Miyajima, T. Ikeda, N. Shindo, Y. Yasumoto, M. Ito, H. Arai, Genomic causes of multiple cerebral cavernous malformations in a Japanese population, J. Clin. Neurosci. 20 (2013) 667–669.

[32] C. Denier, S. Goutagny, P. Labauge, V. Krivosic, M. Arnoult, A. Cousin, A. L. Benabid, J. Comoy, P. Frerebeau, B. Gilbert, J.P. Houtteville, M. Jan, F. Lapierre, H. Loiseau, P. Menei, P. Mercier, J.J. Moreau, A. Nivelon-Chevallier, F. Parker, A. M. Redondo, J.M. Scarabin, M. Tremoulet, M. Zerah, J. Maciazek, E. Tournier-Lasserve, N. Societe Francaise de, Mutations within the MGC4607 gene cause cerebral cavernous malformations, Am. J. Hum. Genet. 74 (2004) 326–337.

[33] J.S. Zawistowski, L. Stalheim, M.T. Uhlik, A.N. Abell, B.B. Ancrile, G.L. Johnson, D. A. Marchuk, CCM1 and CCM2 protein interactions in cell signaling: implications for cerebral cavernous malformations pathogenesis, Hum. Mol. Genet. 14 (2005) 2521–2531.

[34] F. Riant, M. Cecillon, P. Saugier-Veber, E. Tournier-Lasserve, CCM molecular screening in a diagnosis context: novel unclassified variants leading to abnormal splicing and importance of large deletions, Neurogenetics 14 (2013) 133–141.

[35] G. Han, L. Ma, H. Qiao, L. Han, Q. Wu, Q. Li, A novel CCM2 missense variant caused cerebral cavernous malformations in a Chinese family, Front. Neurosci. 14 (2020) 604350.

[36] K. Ishii, N. Tozaka, S. Tsutsumi, A. Muroi, A. Tamaoka, Familial cerebral cavernous malformation presenting with epilepsy caused by mutation in the CCM2 gene: a case report, Medicine (Baltim.) 99 (2020), e19800.

[37] R. D'Angelo, C. Scimone, C. Rinaldi, G. Trimarchi, D. Italiano, P. Bramanti, A. Amato, A. Sidoti, CCM2 gene polymorphisms in Italian sporadic patients with cerebral cavernous malformation: a case-control study, Int. J. Mol. Med. 29 (2012) 1113–1120.

[38] Q. Du, Z. Shi, H. Chen, Y. Zhang, J. Wang, H. Zhou, Two novel CCM2 heterozygous mutations associated with cerebral cavernous malformation in a Chinese family, J. Mol. Neurosci. 67 (2019) 467–471.

[39] P.J. Dunn, B.H. Maher, C.L. Albury, S. Stuart, H.G. Sutherland, N. Maksemous, M. C. Benton, R.A. Smith, L.M. Haupt, L.R. Griffiths, Tiered analysis of whole-exome sequencing for epilepsy diagnosis, Mol. Genet. Genom. 295 (2020) 751–763.
[40] S. Pileggi, S. Buscone, C. Ricci, M.C. Patrosso, A. Marocchi, P. Brunori, S. Battistini, S. Penco, Genetic variations within KRIT1/CCM1, MGC4607/CCM2 and PDCD10/CCM3 in a large Italian family harbouring a Krit1/CCM1 mutation, J. Mol. Neurosci. 42 (2010) 235–242.
[41] C. Scimone, P. Bramanti, C. Alafaci, F. Granata, F. Piva, C. Rinaldi, L. Donato, F. Greco, A. Sidoti, R. D'Angelo, Update on novel CCM gene mutations in patients with cerebral cavernous malformations, J. Mol. Neurosci. 61 (2017) 189–198.
[42] C. Scimone, L. Donato, Z. Katsarou, S. Bostantjopoulou, R. D'Angelo, A. Sidoti, Two novel KRIT1 and CCM2 mutations in patients affected by cerebral cavernous malformations: new information on CCM2 penetrance, Front. Neurol. 9 (2018) 953.
[43] W.Q. Huang, C.X. Lu, Y. Zhang, K.H. Yi, L.L. Cai, M.L. Li, H. Wang, Q. Lin, C. M. Tzeng, A novel CCM2 gene mutation associated with familial cerebral cavernous malformation, Front. Aging Neurosci. 8 (2016) 220.
[44] R. D'Angelo, C. Alafaci, C. Scimone, A. Ruggeri, F.M. Salpietro, P. Bramanti, F. Tomasello, A. Sidoti, Sporadic cerebral cavernous malformations: report of further mutations of CCM genes in 40 Italian patients, BioMed Res. Int. 2013 (2013) 459253.
[45] E. Merello, M. Pavanello, A. Consales, S. Mascelli, A. Raso, A. Accogli, A. Cama, C. Valeria, P. De Marco, Genetic screening of pediatric cavernous malformations, J. Mol. Neurosci. 60 (2016) 232–238.
[46] E. Merello, M. Pavanello, A. Consales, S. Mascelli, A. Raso, A. Accogli, A. Cama, V. Capra, P. De Marco, Erratum to: genetic screening of pediatric cavernous malformations, J. Mol. Neurosci. 61 (2017) 132.
[47] M. Rath, S.E. Jenssen, K. Schwefel, S. Spiegler, D. Kleimeier, C. Sperling, L. Kaderali, U. Felbor, High-throughput sequencing of the entire genomic regions of CCM1/KRIT1, CCM2 and CCM3/PDCD10 to search for pathogenic deep-intronic splice mutations in cerebral cavernous malformations, Eur. J. Med. Genet. 60 (2017) 479–484.
[48] N. Sain, G. Tiwari, D. Mohanty, Understanding the molecular basis of substrate binding specificity of PTB domains, Sci. Rep. 6 (2016) 31418.