J. Bangladesh Acad. Sci. 47(2); 181-193: December 2023 DOI:https://doi.org/10.3329/jbas.v47i2.67851

Journal of Bangladesh Academy of Sciences

Journal homepage: http://www.bas.org.bd/publications/jbas.html

Research Article

In silico analysis of deleterious SNPs of human *DCDC2* gene and their impacts on subsequent protein-protein interactions

Nure Asma Lata, Nahid Parvez and Sumaiya Farah Khan*

Department of Genetic Engineering and Biotechnology, Faculty of Life and Earth Science, Jagannath University, Dhaka, Bangladesh

ARTICLE INFO

Article History

Received: 25 July 2023
Revised: 08 August 2023
Accepted: 14 August 2023

Keywords: DCDC2, Dyslexia, Microtubules, Single nucleotide polymorphism (SNP), Computational analysis

ABSTRACT

DCDC2 is a clinically significant protein causing a number of neurological disorders and, hence, is an important protein for analysis. In this study, multiple tools were employed to identify missense SNPs that are harmful to the protein itself and destabilize the interaction of this protein with tubulin subunits. After analyzing all 378 missense SNPs retrieved from the dbSNP database, thirteen were found to have harmful effects on the protein, which are L20P, R23L, G25W, G25R, D26E, I36N, G60E, P68S, G83R, T174I, L179R, R186G, V208E. Among these, four SNPs- T174I, L179R, R186G, and V208E were suggested to be significantly destabilizing for the interaction of the C-DC domain with microtubule, and three SNPs- L20P, D26E, and G83R for the interaction of N-DC domain with microtubule. Based on the total ΔΔG value, SNP R186G and L20P seem most destabilizing for the interaction of the C-DC and N-DC domains. These SNPs are found to affect the protein negatively by analysis using several computational tools. Genetic association and protein-protein interaction studies focused on these SNPs can reveal new findings about dyslexia or other neurodevelopmental disorders.

Introduction

DCDC2 (Doublecortin Domain Protein 2) is a member of the superfamily of proteins that include doublecortin domains (DC). This domain was discovered in an X-linked gene called *Doublecortin (DCX),* which is found to be associated with neural development (Liu et al., 2012). The protein encoded by this gene, *Doublecortin (DCX),* is crucial for brain growth and the formation of the layers of the cerebral cortex in the embryonic stage (Kim et al., 2003; Manka and Moores, 2020). Most of the understanding of doublecortin-related genes' molecular function (or malfunction) comes from the DCX protein and its aptitude for microtubule binding (Dijkmans et al., 2010). DCX are traditional microtubule (MT) associated proteins (MAPs), having two conserved tandems in two terminal domains that contribute to tubulin-binding over evolutionary time (Reiner et al., 2006) and is linked with MT stabilization, which regulates the cytoskeleton and is essential for neuronal migration, differentiation, maturation, and postmitotic neurons (Kim et al., 2003; Conde and Caceres, 2009). DCX can be divided into two functional parts: the N-terminal 30 kDa part is the protein's microtubule (MT) binding portion and comprises two homologous, 11 kDa DC domains. DCX contributes to brain development by stabilizing microtubules in the leading process of migrating neurons and other neuronal processes (Horesh et al., 1999). Recently, electron microscopy has shown that a single DC

*Corresponding author: <sumaiyafarah@yhaoo.com>

domain binds to microtubules at the vertex of four tubulin subunits in the groove between protofilaments (Fourniol et al., 2010).

Microtubule-associated proteins (MAPs) are indispensable to neurons at the embryonic stage, administering neuronal migration, dendrites, and axon development (Gleeson et al., 1998). Numerous studies on the effects of DCX on microtubules in neurons have revealed cooperative binding effects, modifications to MT structure, and control over molecular motors (Gleeson et al., 1999; Moores et al., 2006). Several studies reported the dysfunction of MT binding by mutant DCX consequences in inappropriate brain development and evolving diseases like lissencephaly, dyslexia, etc. (Moores et al., 2004; Gabel et al., 2011; Bechstedt and Brouhard, 2012).

α/β-tubulin heterodimers that assemble into protofilaments (PFs) make up microtubules. Doublecortin preferentially induces the construction of 13 protofilament microtubules and binds to them (Tilney et al., 1973; Chaaban and Brouhard, 2017). The nucleation and stabilization of this physiological 13-PF architecture are highly influenced by doublecortin. DCX contains N-terminal (N-DC) and C-terminal domain (C-DC), which have a considerable contribution to microtubule binding via PFs (Bechstedt and Brouhard, 2012; Bechstedt et al., 2014; Burger et al., 2016). MT nucleation seems to be facilitated by the conformationally plastic C-DC module, which also appears to stabilize tubulin-tubulin interactions while N-DC conducts MT stabilization (Manka and Moores, 2020). However, DC domain tandem is necessary for this function because DC domains alone do not induce MT polymerization (Taylor et al., 2000). Missense changes in the DC domains can be divided into two groups: those that occur at surface residues and those at hydrophobic residues ensconced in the heart of the ubiquitin fold. It is anticipated that the latter group will result in DCX misfolding. Bechstedt showed three consequences: loss of 13-PF selectivity, loss of cooperative contacts, and decreased microtubule binding are the first two kinds of biochemical defects in mutant DCX, while decreased turbidity assay performance is the third in their study (Bechstedt and Brouhard, 2012). Other studies reported that point mutation in N-DC, C-DC, and tandem alters DCX binding to MT or forms inappropriate fold and contributes to disease mechanism (Kim et al., 2003; Manka and Moores, 2020). Over and above deafness in humans was caused by a missense mutation in the doublecortin protein's C-terminal region (Grati et al., 2015).

The current study is focused on *DCDC2* of chromosome 6, which also has DC domains like DCX. Having DC domains, DCDC2 should also interact with microtubules and probably have functions in neural development. Studies reported mutation in our targeted gene *DCDC2* involved in renal-hepatic ciliopathy, neonatal sclerosing cholangitis with developmental delay, nephronophthisis, human recessive deafness, dyslexia, etc. (Gabel et al., 2011; Schueler et al., 2015; Girard, 2016; Srivastava et al., 2018; Syryn et al., 2021). Polymorphism in DCDC2 was reported to emerge in learning disability and speech difficulty (Lind et al., 2010; Zhong et al., 2013; Zhang et al., 2016).

No X-ray crystallography structure is available for the DCDC2 protein under study. The predicted tertiary structure of the DCDC2 protein is available in the AlphaFold database (ID: AF-Q9UHG0-F1), and it shows that this protein, like DCX protein, also contains two DC domains: one is N-terminal (17-100 aa) and another is C-terminal (139-222aa). Each of the domains should interact at the groove between two protofilaments. The rest of the region has no defined structural motif and is indicated as disordered. It has been stated that DC domain-containing proteins are involved in neural development through their interaction with microtubules. It is worth mentioning that association studies of the DYX2 locus on 6p21.3 have already identified altered DCDC2 as involved in common neurogenetic disorder reading disability (RD) or

dyslexia (Meng et al., 2011). The mentioned association studies identified 14 missense SNPs, each part of either N-terminal DC or C-terminal DC domains.

SNPs (Single nucleotide polymorphism), a molecular marker, refers to the variation or substitution of a single nucleotide at a specific position in DNA sequence among individuals (Vignal et al., 2002). SNPs can either be silent (synonymous), modify the encoded amino acids, or appear in non-coding areas. Missense SNPs result in the substitution of the wild-type amino acid. They are reported to be involved in the diversity of people, the evolution of the genome, and the most prevalent familial features, interindividual variations in drug response, as well as complicated and widespread illnesses like diabetes, obesity, hypertension, and mental health issues. They may have an impact on messenger RNA (mRNA) structure (stability), subcellular localization of mRNAs and/or proteins, and promoter activity (gene expression), and hence may result in illness (Shastry, 2009).

Though polymorphisms in DCDC2 cause several cases of diseases, no SNP-based computational analysis has been conducted yet. *Silico* SNP analysis will reveal risky SNP candidates' functional and structural importance. As a MAP-associated protein-coding gene, *DCDC2* has a distinct importance as a target of the experiment. The study will identify the most damaging and disease-causing missense SNP, alteration of protein stability, and evolutionary conservation to evaluate the potential for causing disease in the selected candidate. Moreover, maintaining protein-protein interactions is important for the usual function of the protein, and any mutation altering that interaction can cause disease. Accurate prediction of the effect of missense SNPs in protein binding is crucial for carrying out genome-wide studies. Therefore, in this study, the effect of missense SNPs on the protein-protein interaction of DCDC2 has also been explored.

Materials and Methods

Retrieving missense SNPs

All the required data of the human *DCDC2* gene, such as FASTA sequence and SNP, were retrieved from UniProtKB (https://www.uniprot.org/) and the National Centre for Biological Information (NCBI) (https://www.ncbi.nlm.nih.gov/) respectively. For this study, our main concern was missense SNPs. The missense SNP information, including protein accession number, missense rsID, allele change, position, and residue change, was collected from the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/nsNSSNP/).

Identifying the most damaging missense SNPs

The functional effect of missense SNPs of human *DCDC2* gene was identified using five tools, PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2010), PROVEAN [http://provean.jcvi.org/index.php] (Choi and Chan, 2015), SNPnexus (https://www.nsNSSNP-nexus.org/v4/) (Oscanoa et al., 2020), SNAP2 (https://www.rostlab.org/servces/snap/) (Johnson et al., 2008), Pon-P2 (http://structure.bmc.lu.se/PON-P2/) (Niroula et al., 2015).

PolyPhen-2 (Polymorphism Phenotyping v2) is an automatic tool that predicts the impacts of a missense SNP based on the sequence, biological process, and structural information characterizing the substitution.

PROVEAN is a tool similar to PolyPhen2, which predicts the impact of amino acid substitution.

SNPnexus tool uses the major gene annotation system to bring out essentially yields two results: tolerated and deleterious for SIFT prediction, and PolyPhen-2 prediction comes out as benign, perhaps damaging, and probably damaging (Oscanoa et al., 2020).

The SNAP2 server classifies the variations as having an effect or being neutral and provides a confidence score (Johnson et al., 2008).

PON-P2 is a pathogenic missense SNP identifier that groups the output into neutral, pathogenic, or unknown depending on the random forest probability score (Niroula et al., 2015).

In the case of PolyPhen-2, score 1 indicates the most damaging SNPs (http://genetics.bwh.harvard.edu/pph2/). For PROVEAN and SNPnexus, deleterious SNPs are indicated by scores lower than -2.5 and -0.5, respectively (http://provean.jcvi.org/index.php, https://www.nsNSSNP-nexus.org/v4/). The lower the score from the threshold, the more deleterious the SNPs are. In the case of SNAP2 (https://www.rostlab.org/servces/snap/) and PON-P2 (http://structure.bmc.lu.se/PON-P2/), SNPs are identified as deleterious if their scores are higher than 0 and 0.5 respectively. A higher score indicates more deleterious SNP. All the collected missense SNPs from dbSNP went through these five identifier tools to predict the most deleterious missense SNPs. SNPs, determined deleterious by at least four tools, were used for further investigation.

Results and discussions

Identification of disease-associated missense SNPs

SNPs&GO (http://snps-and-go.biocomp.unibo.it/snps-and-go/) and PMut (http://mmb.pcb.ub.es/PMut/) were applied to predict the association of SNPs with disease. The tool SNPs&GO is based on a supporting vector machine that classifies human disease correspondent single amino acid substitution with 82% accuracy, and results are classified into disease and neutral (Calabrese et al., 2009). Protein sequence, evolutionary information, and other information from gene ontology are used to predict (https://snps-and-go.biocomp.unibo.it/snps-and-go/). PMut is also a classifier that discovers possible disease-associated missense SNPs utilizing the manually curated variation database SwissVar. SNPs, sorted as a disease by both tools, were considered significant and selected as final short-listed SNPs for additional studies.

Protein evolutionary conservation analysis

A visual image was generated where "f" denoted a functional residue that is highly conserved and exposed, and "s" denoted a predicted structural residue that is highly conserved and buried. Color gradient showed the variable region to a highly conserved region. From this prediction, the SNPs and positions were found with conservation scores as color intensity.

Amino acid residues that are highly conserved are generally part of an important functional domain of a protein. ConSurf(https://consurf.tau.ac.il/), a Bayesian inference-based tool, was used for the conservation analysis of the DCDC2 protein sequence to detect highly conserved regions (Ashkenazy et al., 2010). It creates a phylogenetic tree based on the relationships between homologous sequences (Pupko et al., 2002). A conservation score of 1 to 4 was considered variable, a score of 5 to 6 was considered moderate, and a score of 7 to 9 was considered conserved (Jia et al., 2014).

Analysis of protein stability

The mutation causes an increase or decrease in protein stability, which causes an alteration of protein function. Changes in the stability of the protein due to point mutation were computed by two software, I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) and MUpro.

I-Mutant2.0 is a tool that predicts protein stability changes depending on support vector machine (SVM) utilizing data from ProTherm. It calculates ΔΔG value by subtracting the wild protein's unfolding free energy value from the mutant protein's folding Gibbs free energy (Bava and Gromiha, 2004; Capriotti et al., 2005). MUpro also utilizes ΔΔG value to predict the stability change resulting from missense SNPs (Baldi, 2005).

Protein-protein interaction analysis

Based on the available experimental structures of N-DC domain and C-DC domain of DCX protein with tubulin subunits (6rev, 6RF2), two homology models have been generated using SWISS-MODEL (https://swissmodel.expasy.org/). Model No. 1 contains the N-DC domain of DCDC2 in complex with two monomers of tubulin alpha-1B chain and two monomers of tubulin beta-2B chain. Model no. 2 contains the C-DC domain of DCDC2 in a complex with identical tubulin monomers. Both models are larger than 500 amino acids, and no available *in silico* tool can refine and analyze the protein-protein interaction of such large models. So, each model was divided into two models: Model 1a containing N-DC

domain of DCDC2 and two monomers of tubulin alpha-1B chain, Model 1b containing N-DC domain of DCDC2 and two monomers of tubulin beta-2B chain, Model 2a containing C-DC domain of DCDC2 and two monomers of tubulin alpha-1B chain and Model 2b containing C-DC domain of DCDC2 and two monomers of tubulin beta-2B chain. These four models were refined with the Galaxy refine server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE). Later, the refined versions were analyzed with three different tools (MCSM-PPI2, SAMBE-3D, and Beat-Music) to determine each complex's binding free energy change (ΔΔG) due to missense SNPs. The final short-listed SNPs were considered for predicting binding energy change (ΔΔG). mCSM-PPI2 is a novel machine-learning computational tool that exploits graph-based structural signatures to model the effects of variations on the inter-residue interaction network for more accurately predicting the effects of missense mutations on protein-protein interaction binding affinity (Dijkmans et al., 2010).

SAAMBE-3D is another machine learning-based approach for fast and accurate protein-protein interaction predictions (Pahari et al., 2020).

BeAtMuSiC is a coarse-grained predictor of the changes in binding free energy induced by point mutations. It relies on a set of statistical potentials derived from known protein structures and combines the effect of the mutation on the strength of the interactions at the interface and on the overall stability of the complex (Dehouck et al., 2013).

Table 1. Short-listed 13 SNPs and scores of those SNPs in seven tools

No	SNP	PolyPhen2[a]	PROVEAN[b]	SNPnexus[c]	SNAP2[d]	PON-P2[e]	SNPs&GO[f]	PMut[g]
1	R23L	1	-5.33	0.03	83	0.758	0.619	0.83
2	G83R	1	-6.48	0.02	62	0.805	0.674	0.77
3	G60E	1	-4.19	0	84	0.054	0.897	0.83
4	I36N	0.99	-5.36	0	60	0.582	0.820	0.84
5	R186G	0.998	-4.50	0	81	0.484	0.589	0.80
6	L179R	0.997	-4.39	0	76	0.759	0.606	0.64
7	G25W	1	-7.12	0	90	0.656	0.803	0.89
8	G25R	0.733	-7.08	0.02	86	0.626	0.716	0.16
9	P68S	1	-6.06	0	48	0.679	0.625	0.87
10	V208E	1	-5.33	0	80	0.717	0.652	0.65
11	L20P	0.979	-3.44	0	63	0.576	0.580	0.88
12	D26E	0.999	-3.58	0.03	72	0.481	0.512	0.86
13	T174I	1	-4.30	0	59	0.677	0.619	0.83

[a]PolyPhen2 predicts damaging if score is less than or equal to 1, [b]PROVEAN predicts deleterious if score is less than -2.5, [c]SNPnexus predicts deleterious if score is between 0-0.05, [d]SNAP2 predicts effect if score is greater than 0, [e]PON-P2 predicts pathogenic if score is less than 0.5, [f]SNPs&GO predicts disease if score is higher than 0.5, [g]PMut predicts disease if score is less than 1.

Analysis of evolutionary conserved regions

Evolutionary conserved regions are essential in proteins as they are more involved in biological processes and mechanisms (Asthana et al., 2007). As a result, SNPs located in highly conserved regions are more prone to cause disease. ConSurf web server was used to examine the conserved regions of DCDC2 and the location of the 13 short-listed SNPs (Supplementary Fig. 1). All the thirteen short-listed SNPs were classified as conserved ones (Table. 2). It implies that these SNPs are part of an essential functional or structural domain of the protein and their presence may significantly affect the protein function and stability.

Analysis of protein stability

Stability changes for the mutation in amino acid sequence was predicted by I-Mutant and MUpro predicted stability changes for the mutation in the amino acid sequence. When protein stability increases, it works more efficiently, but a decrease in stability causes the hindrance of protein action by degradation, misfolding, and aggregation (Du et al., 2005; Schoorman et al., 2007; Platek and Singh, 2010). For both tools, ΔΔG value less than zero indicates a decrease in stability. All thirteen short-listed SNPs reduced the protein's stability (Table 2).

Table 2. Analysis of and stability alteration due to short-listed 13 SNPs and conserved status of respective amino acid positions

SNP	Stability alteration		Conservation analysis
	ΔΔG[#] (I-Mutant)	**ΔΔG[#]** (MUpro)	**Score of each residue position[*]** (ConSurf)
R23L	-1.26	-0.65	9
G83R	-0.91	-0.58	9
G60E	-0.14	-0.17	9
I36N	-1.29	-1.65	8
R186G	-1.06	-1.32	9
L179R	-2.49	-2.57	8
G25W	-0.49	-0.94	9
G25R	-0.82	-1.01	
P68S	-2.01	-1.51	9
V208A	-1.56	-1.92	9
L20P	-1.72	-1.37	6
D26E	-2.24	-1.26	9
T174I	-0.57	-0.17	9

[#]ΔΔG<0 indicates decreased stability

[*]ConSurf scores are given on a scale from 1-9, where 1 means variable and 9 means highly conserved.

Analysis of protein-protein interaction

The DCDC2 protein (UniProt ID: Q9UHG0) contains two DC domains: one is N-terminal (17-100 aa), and the other is C-terminal (139-222aa). All the thirteen short-listed SNPs are part of either of those domains. It has been mentioned previously that DC domains of other proteins bind 13-protofilaments of microtubules at grooves surrounded by four tubulin monomers (Manka and Moores, 2020). Homology modeling was used to prepare Models 1 and 2 (Fig. 1).

Each model was divided into two and refined, resulting in four 3D models of DC domains and two tubulin monomers. These models were used to analyze the effect of thirteen short-listed SNPs on DCDC2 and tubulin interaction. Model 1, a and 1, b were used for SNPs of the N-DC domain (17-100aa), whereas Model 2,a and 2,b were used for SNPs of the C-DC domain (139-222aa) (Fig. 2). ΔΔG values calculated by three different tools of the considered SNPs are listed in Table 3 and 4. A positive value of ΔΔG indicates that the SNP lowers the binding energy of the DC domain with respective tubulin subunits. Among the considered six SNPs of the C-DC domain, four showed positive ΔΔG change in all three tools and for models 2a and 2b. These SNPs have been considered significant in case of destabilizing the interaction of the C-DC domain of DCDC2 with tubulin alpha-1B dimer and beta-2B dimer (Table. 5). Similarly, among the seven SNPs of N-DC domains, three were considered as significant (Table. 6).

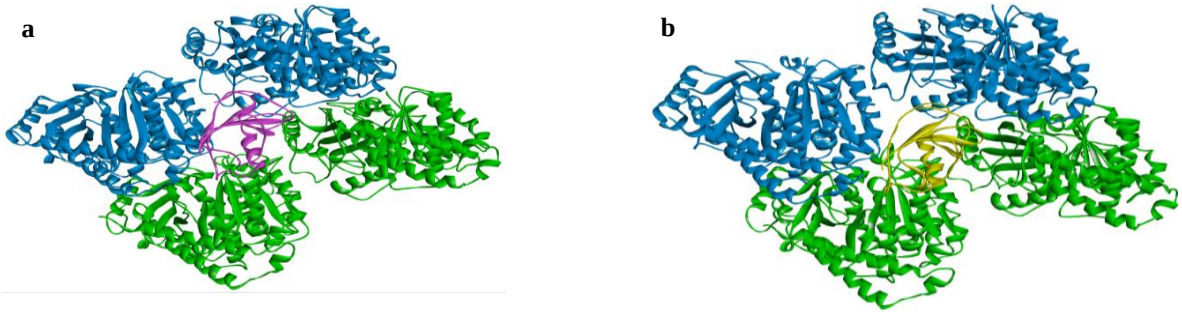

Fig. 1. a) Model 1- Complex of N-DC domain of DCDC2 with four tubulin subunits, b) Model 2- Complex of C-DC domain of DCDC2 with four tubulin subunits. N-DC domain, C-DC domain, tubulin alpha-1B subunits, and tubulin beta-2B subunits are indicated by purple, yellow, green, and blue color, respectively.

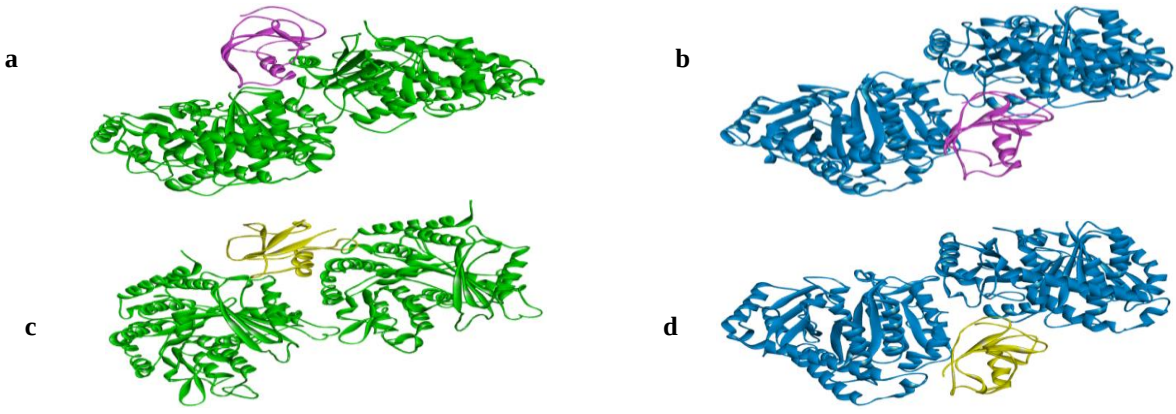

Fig. 2. a) Model 1a- Complex of N-DC domain with two tubulin alpha-1B subunits, b) Model 1b- Complex of N-DC domain with two tubulin beta-2B subunits, c) Model 2a- Complex of C-DC domain with two tubulin alpha-1B subunits, d) Model 2b- Complex of C-DC domain with two tubulin beta-2B subunits. The n-DC domain, C-DC domain, tubulin alpha-1B subunits, and tubulin beta-2B subunits are indicated by purple, yellow, green, and blue colors, respectively.

Table 3. Alteration of interaction of C-DC domain with α-1B pair and β-2B pair of tubulin 13-protofilament

Wild type	Residue No[b]	Mutant	Model 2a C_DCDC2_α-1B tubulin pair ΔΔG[a]			Model 2b C_DCDC2_β-2B tubulin pair ΔΔG[a]		
			MCSM-PPI2	SAAMBE-3D	Beat-Music	MCSM-PPI2	SAAMBE-3D	Beat-Music
P	152	A	0.106	0.22	*-0.06*	0.483	0.61	1.48
T	174	I	0.462	0.16	0.56	0.056	0.05	0.13
L	179	R	0.809	0.61	0.48	0.728	0.89	0.47
S	181	R	0.305	0.68	0.32	*-0.182*	0.14	0.34
R	186	G	1.339	0.92	1.41	0.354	0.80	0.65
V	208	E	0.453	0.09	1.33	0.42	0.09	1.61

[a]Positive ddG value indicates decreasing binding energy
[b]SNPs that create positive ddG value in all three tools are written in bold.

Table 4. Alteration of interaction of N-DC domain with α-1B pair and β-2B pair of tubulin13-protofilament

Wild	Residue No[b]	Mutant	Model 1a N_DCDC2_ α-1B tubulin pair ΔΔG[a]			Model 1b N_DCDC2_ β-2B tubulin pair ΔΔG[a]		
			MCSM-PPI2	SAAMBE-3D	Beat-Music	MCSM-PPI2	SAAMBE-3D	Beat-Music
L	**20**	**P**	**0.254**	**1.25**	**0.92**	**0.328**	**1.25**	**1.25**
R	23	L	0.253	0.53	*-0.04*	0.803	0.81	*-0.08*
G	25	w	0.419	-0.20	0.47	1.569	0.13	*-0.38*
G	25	R	0.199	-0.20	0.48	1.101	0.69	0.85
D	**26**	**E**	**0.140**	**0.9**	**0.44**	**0.549**	**1.00**	**0.99**
G	60	E	1.407	0.44	2.43	-0.169	0.24	1.14
G	**83**	**R**	**0.245**	**0.20**	**0.48**	**0.278**	**0.29**	**0.53**

[a] positive ddG value indicates decreasing binding energy
[b]SNPs that create positive ddG value in all three tools are written in bold.

The DCDC2 protein contains a doublecortin (DC) domain. This domain was found in the doublecortin (DCX) protein that binds microtubules. doublecortin (DCX) has been shown to stabilize microtubules and cause bundling in both *in vivo* and *in vitro* assays. Doublecortin, a basic protein, has an isoelectric point of 10, typical of microtubule-binding proteins.

In this *in silico* study, thirteen SNPs were short-listed after analysis with seven tools (L20P, R23L, G25W, G25R, D26E, I36N, G60E, P68S, G83R, T174I, L179R, R186G, V208E). These SNPs have not been reported to be associated with neurological or other physiological disorders before. All thirteen SNPs were either in N-terminal (17-100 aa) or C-terminal (139-222aa) DC domains. This corresponds to the finding that they are classified to be highly conserved in the ConSurf server. All thirteen SNPs have been found to lower the stability of the protein too. It indicates that these particular SNPs can compromise the structural integrity of the protein. As regions except N-DC and C-DC domains are disordered and no significant SNPs were found in the disordered regions, protein interaction analysis was focused only on N-DC and C-DC domains. Protein-protein interaction analysis revealed that four SNPs of the N-DC domain and three SNPs of the C-DC domains should hamper the interaction between the DC domains of DCDC2 and the microtubule. Inside the cell, microtubule-bound DC domains have been reported to be at the vertex of four tubulin subunits, two of which are alpha-1B monomers and the other two are beta-2B monomers (6rev, 6RF2) (Manka and Moores, 2020). The simultaneous interaction of the DC domain with four different subunits makes it very complicated to perform molecular docking analysis. Recently, the SWISS-MODEL server has extended its functionality to the modeling of homo- and heteromeric complexes (Waterhouse et al., 2018). It is based on the fact that, like homology modeling of monomeric proteins, the information of a protein's quaternary structure can be transferred to another model of homologous protein complex through homology (Szilagyi and Zhang, 2014). So, to avoid the problem with molecular docking, homology modeling via SWISS-MODEL was utilized to construct the structure of a single DC domain bound to two monomers of tubulin alpha chain and two monomers of tubulin beta chain simultaneously. Only one homology domain was generated for each of the DC domains. To improve the quality, the structures needed refinement, but the 3D structure of the complex is enormous and unsuitable for computational analysis. So, each of the complexes was divided into two structures; one has a DC domain with two monomers of tubulin alpha chain, and another has a DC domain with two monomers of tubulin beta chain. As homology modeling outputs are devoid of water molecules and contain H-atoms, these models were directly submitted to interaction analysis software.

ΔΔG values of model 2,a indicate the binding energy change of C-DC with alpha-1B dimer, and model 2,b indicates that of beta-2B dimer. So, for each SNP, the summation of the average ΔΔG values of both models can suggest the total binging energy change of the C-DC domain with four tubulin subunits. Table 5 shows that SNP R186G can be considered most destabilizing for interaction as it has the highest positive total ΔΔG value. Similarly, the summation of the average ΔΔG values of models 1,a and 1,b from Table 5 indicates the total binding energy change of the N-DC domain with four tubulin subunits. Among the three SNPs considered significant in altering the N-DC domain and tubulin interaction, L20P SNP can be considered as having the most destabilizing effect on the interaction of the N-DC domain, with four tubulin subunits for having the largest total positive ΔΔG value (Table 6). Positions of these SNPs in the tertiary structure of C-DC domain and N-DC domain are shown in Fig. 3.

Table 5. Order of missense SNPs of C-DC domain based on ΔΔG value

SNP	Average ΔΔG of C-DC with α-1B pair[#]	Average ΔΔG of C-DC with β-2B pair[*]	Sum of ΔΔG of C-DC domain[$]	Significance
R186G	1.223	0.601	1.824	
V208E	0.624	0.706	1.33	
L179R	0.633	0.696	1.33	
T174I	0.394	0.078	0.472	

[#]ΔΔG values indicating interaction alteration of C-DC domain with α-1B sub-unit pair
[*]ΔΔG values indicating interaction alteration of C-DC domain with β-2B sub-unit pair
[$]Total ΔΔG values indicating interaction alteration of C-DC domain with both α-1B pair and β-2B pair

Table 6. Order of missense SNPs of N-DC domain based on ΔΔG value

SNP	Average ΔΔG of N-DC with α-1B pair[#]	Average ΔΔG of N-DC with β-2B pair*	Sum of ΔΔG of N-DC domain[$]	Significance
L20P	0.808	0.942	1.75	
D26E	0.493	0.846	1.34	
G83R	0.308	0.366	0.674	

[#]ΔΔG values indicating interaction alteration of N-DC domain with α-1B subunit pair
*ΔΔG values indicating interaction alteration of N-DC domain with β-2B subunit pair
[$]Total ΔΔG values indicating interaction alteration of N-DC domain with both α-1B pair and β-2B pair

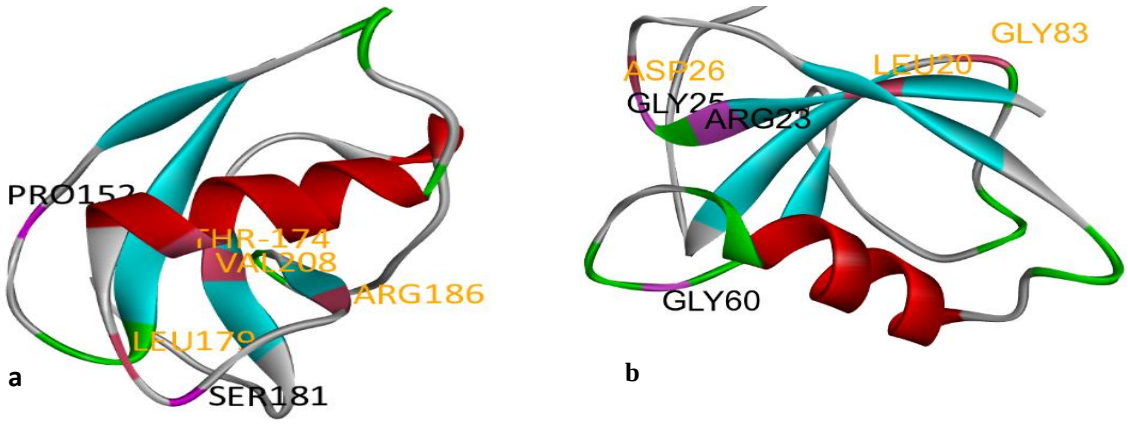

Fig. 3. a) C-DC domain structure with significant SNPs' amino positions. b) N-DC domain with amino positions of significant SNPs. Helices are red, beta sheets are cyan, turns are green, and coils are white. Residues written in yellow are significant for microtubule interaction.

Conclusion

Being implicated in several neurological disorders makes DCDC2 an important protein for analysis. This study utilized multiple tools to identify missense SNPs that are deleterious to the protein itself and destabilizing to the interaction of this protein with tubulin subunits. Thirteen SNPs were found to have deleterious effects on the protein, which are L20P, R23L, G25W, G25R, D26E, I36N, G60E, P68S, G83R, T174I, L179R, R186G, V208E. All are part of either the N-DC domain or C-DC domain. No SNP of the disordered region was found to be deleterious. Among these, four SNPs, T174I, L179R, R186G, and V208E, were suggested to be significantly destabilizing for the interaction of the C-DC domain with microtubule, and three SNPs- L20P, D26E, and G83R were suggested to be significantly destabilizing for the interaction of N-DC domain with microtubule. Based on the total ΔΔG value, SNP R186G and L20P seem most destabilizing for the interaction of the C-DC and N-DC domains. These SNPs have been identified for negatively affecting the protein in several types of analysis performed via tools of different algorithms. As *in-vitro* studies are yet to be performed on these SNPs, this study proposes a few potential deleterious SNPs worth looking into. Genetic association and protein-protein interaction studies focused on these SNPs can reveal new findings about dyslexia or other neurodevelopmental disorders.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS and Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat. Methods.* 2010; 7(4): 248-249.

Ashkenazy H, Erez E, Martz E, Pupko T and Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 2010; 38: 529-533.

Asthana S, Roytberg M, Stamatoyannopoulos J and Sunyaev S. Analysis of sequence conservation at nucleotide resolution. *PLoS Comput. Biol*. 2007; 3(12): 1-10.

Baldi JCARP. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins.* 2005; 62(4):1125-1132.

Bava KA and Gromiha MM. ProTherm, version 4.0: thermodynamic database for proteins and mutants, *Nucleic Acids Res.* 2004; 32(90001): D120-D121.

Bechstedt S and Brouhard GJ. Doublecortin recognizes the 13-protofilament microtubule cooperatively and tracks microtubule ends. *Dev. Cell.* 2012;23(1):181-192.

Bechstedt S, Lu K and Brouhard GJ. Doublecortin recognizes the longitudinal curvature of the microtubule end and lattice. *Curr. Biol.* 2014;24(20): 2366-2375.

Burger D, Stihle M, Sharma A, Lello PD, Benz J, D'Arcy B, Debulpaep M, Fry D, Huber W, Kremer T, Laeremans T, Matile T, Ross A, Rufer AC, Schoch G, Steinmetz MO, Steyaert J, Rudolph MG, Thoma R and Ruf A. Crystal structures of the human doublecortin C- and N-terminal domains in complex with specific antibodies. *J. Biol. Chem.* 2016; 291(31):16292-16306.

Calabrese R, Capriotti E, Fariselli P, Martelli PL and Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* 2009; 30(8): 1237-1244.

Capriotti E, Fariselli P and Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 2005; 33: 306-310.

Chaaban S and Brouhard GJ. A microtubule bestiary: structural diversity in tubulin polymers. *Mol. Biol. Cell*. 2017; 28(22): 2924-2931.

Choi Y and Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid

substitutions and indels. *Bioinformatics.* 2015; 31(16): 2745-2747.

Conde C and Caceres A. Microtubule assembly, organization and dynamics in axons and dendrites. *Nat. Rev. Neurosci.* 2009;10: 319-332.

Dehouck Y, Kwasigroch JM, Rooman M and Gilis D. BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res.* 2013; 41(W1): 333-339.

Dijkmans T, Hooijdonk LWA, Fitzsimons CP and Vreugdenhil E. The doublecortin gene family and disorders of neuronal structure. *Cent. Nerv. Syst. Agents Med. Chem.* 2010; 10(1): 32-46.

Du K, Sharma M and Lukacs GL. The ΔF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat. Struct. Mol. Biol.* 2005; 12(1): 17-25.

Fourniol FJ, Sindelar CV, Amigues B, Clare DK, Thomas G, Perderiset M, Francis F, Houdusse A and Moores CA. Template-free 13-protofilament microtubule-MAP assembly visualized at 8 A resolution. *J. Cell Biol.* 2010; 191(3): 463-470.

Gabel LA, Marin I, LoTurco JJ, Che A, Murphy C,Manglani M and Kass S. Mutation of the dyslexia-associated gene Dcdc2 impairs LTM and visuo-spatial performance in mice. *Genes, Brain Behav.* 2011; 10(8): 868-875.

Girard M. DCDC2 mutations cause neonatal sclerosing cholangitis. *Hum. Mutat.* 2016; 37(10): 1025-1029.

Gleeson JG, Allen KM, Fox JW, Lamperti ED, Berkovic S, Scheffer I, Cooper EC, Dobyns WB, Minnerath SR, Ross ME and Walsh CA. Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein. *Cell*. 1998; 92(1): 63-72.

Gleeson JG, Lin PT, Flanagan LA and Walsh CA. Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons. *Neuron*. 1999; 23(2): 257-271.

Grati M, Chakchouk I, Ma Q, Bensaid M, Desmidt A, Turki N and Yan D. A missense mutation in DCDC2 causes human recessive deafness DFNB66, likely by interfering with sensory hair cell and supporting cell cilia length regulation. *Hum.Mol. Genet.* 2015; 24(9): 2482-2491.

Horesh D, Sapir T, Francis F, Wolf SG, Caspi M, Elbaum M, Chelly J and Reiner O. Doublecortin, a stabilizer of microtubules. *Hum. Mol. Genet*. 1999; 8(9): 1599-1610.

Jia M, Yang B, Li Z, Shen H, Song X and Gu W. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 gene. *PLoS ONE.* 2014; 9(8): e104311.

Johnson AD, Handsaker RE, Pulit SL, Nizzari SM, O'Donnell CJ and Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using Hap Map. *Bioinformatics*. 2008; 24(24): 2938-2939.

Kim MH, Cierpicki T, Derewenda U, Krowarsch D, Feng Y, Devedjiev Y and Dauter Z. The DCX-domain tandems of doublecortin and doublecortin-like kinase. *Nat. Struct. Mol. Biol.* 2003; 10(5): 324-333.

Lind PA, Luciano M, Wright MJ, Montgomery JW, Martin NG and Bates TC. Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *Eur. J. Hum. Genet*. 2010; 18(6): 668-673.

Liu JS, Schubert CR, Fu X, Stultz CM, Moores CA and Walsh CA. Molecular basis for specific regulation of neuronal kinesin-3 motors by doublecortin family proteins. *Mol. Cell*. 2012; 47(5): 707-721.

Manka SW and Moores CA. Pseudo-repeats in doublecortin make distinct mechanistic contributions to microtubule regulation. *EMBO Rep.* 2020; 21(12): e51534.

Meng H, Powers NR, Tang L, Cope NA, Zhang PX and Fuleihan R. A dyslexia-associated variant in DCDC2 changes gene expression. *Behav. Genet.* 2011; 41(1): 58-66.

Moores CA, Perderiset M, Francis F, Chelly J, Houdusse A and Milligan RA, mechanism of microtubule stabilization by doublecortin. *Mol. Cell.* 2004;14(6): 833-839.

Moores CA, Perderiset M, Kappeler C, Kain S, Drummond D and Perkins SJ. Distinct roles of doublecortin modulating the microtubule cytoskeleton. *EMBO J.* 2006; 25(19): 4448-4457.

Niroula A, Urolagin S and Vihinen M. PON-P2: Prediction method for fast and reliable identification of harmful variants. *PLoS ONE*. 2015; 10(2): e0117380

Oscanoa J, Sivapalan L, Gadaleta E, Ullah AZD, Lemoine NR and Chelal C. SNPnexus: a web server for functional annotation of human genome sequence variation. *Nucleic Acids Res*. 2020; 48(W1): W185-W192.

Pahari S, Li G, Murthy AK, Liang S, Fragoza R,Yu H, and Alexov E. SAAMBE-3D: Predicting effect of mutations on protein-protein interactions. *Int. J. Mol. Sci.*2020; 21(7): 2563.

Platek SM and Singh D. Optimal waist-to-hip ratios in women activate neural reward centers in men. *PLoS ONE*. 2010; 5(2):1-5.

Pupko T, Bell RE, Mayrose I, Glaser F and Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*. 2002; 18(Suppl 1): S71-S77.

Reiner O, Coquelle FM, Peter B, Levy T, Kaplan A, Sapir T and Orr I. The evolving doublecortin (DCX) superfamily. *BMC Genom*. 2006; 7(1): 188.

Schoorman FD, Mayer RC and Davis JH. An integrative model of organizational trust: past, present, and future. *Acad. Manage. Rev*. 2007; 32(2):344-354.

Schueler M, Braun DA, Chandrasekar G, Gee HY, Klasson TD, Halbritter J, Bieder A, Porath JD, Airik R, Zhou W, LoTurco JJ, Che A and Otto EA. DCDC2 mutations cause a renal-hepatic ciliopathy by disrupting wnt signaling. *Am. J. Hum. Genet*. 2015; 96(1): 81-92.

Shastry BS. *Single Nucleotide Polymorphisms*. 2nd ed. Springer; 2009;p 3-22.

Srivastava S, Molinari E, Raman S and Sayer JA. Many genes—one disease? Genetics of Nephronophthisis (NPHP) and NPHP-associated disorders. *Front. Pediatr*.2018;5:287.

Syryn H, Hoorens A, Grammatikopoulos T, Deheragoda M, Symoens S and Velde SV. Two cases of related neonatal sclerosing cholangitis with developmental delay and literature review. *Clin. Genet.* 2021; 100(4): 447-452.

Szilagyi A and Zhang Y. Template-based structure modeling of protein-protein interactions. *Curr. Opin. Struct. Biol.* 2014; 24: 10-23.

Taylor KR, Holzer AK, Bazan JF, Walsh CA and Gleeson JG. Patient mutations in doublecortin define a repeated tubulin-binding domain. *J. Biol. Chem.* 2000; 275(44): 34442-34450.

Tilney LG, Bryan J, Bush DJ, Fujiwara K, Mooseker MS, Murphy DB and Snyder DH. Microtubules: evidence for 13 protofilaments. *J. Cell Biol.*1973; 59(2): 267-275.

Vignal A, Milan D, SanCristobal M and Egge A. A review on SNP and other types of molecular markers and their use in animal genetics.*Genet. Sel. Evol.* 2002; 34(3): 275-305.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R andHeer FT. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.*2018; 46(W1): W296-W303.

Zhang Y, Li J,Song S,Tardif T,Burmeister M,Villafuerte SM,Su M,McBride C and Shu H. Association of DCDC2 polymorphisms with normal variations in reading abilities in a chinese population. *PLoS ONE*. 2016; 11(4): e0153603.

Zhong R, Yang B, Tang H, Zou L, Song R, Zhu LQ and Miao X. Meta-analysis of the association between DCDC2 polymorphisms and risk of dyslexia. *Mol. Neurobiol.*2013; 47(1): 435-442.