PLOS ONE

RESEARCH ARTICLE

A comprehensive *in silico* analysis of the deleterious nonsynonymous SNPs of human FOXP2 protein

Mahmuda Akter[1], **Sumaiya Farah Khan**[1], **Abu Ashfaqur Sajib**[2], **Fahmida Sultana Rima**[3] *

1 Department of Genetic Engineering and Biotechnology, Jagannath University, Dhaka, Bangladesh,
2 Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh,
3 Department of Biochemistry and Biotechnology, University of Barishal, Barishal, Bangladesh

* fahmidarima7@gmail.com

Abstract

FOXP2 encodes the forkhead transcription factor that plays a significant role in language development. Single nucleotide polymorphisms in *FOXP2* have been linked to speech- language disorder, autism, cancer and schizophrenia. So, scrutinizing the functional SNPs to better understand their association in disease is an uphill task. The purpose of the current study was to identify the missense SNPs which have detrimental structural and functional effects on the FOXP2 protein. Multiple computational tools were employed to investigate the deleterious role of non-synonymous SNPs. Five variants as Y531H, L558P, R536G and R553C were found to be associated with diseases and located at the forkhead domain of the FOXP2 protein. Molecular docking analysis of FOXP2 DNA binding domain with its most common target sequence 5'-CAAATT-3' predicted that R553C and L558P mutant variants destabilize protein structure by changing protein-DNA interface interactions and disruption of hydrogen bonds that may reduce the specificity and affinity of the binding. Further experimental investigations may need to verify whether this kind of structural and functional variations dysregulate protein activities and induce formation of disease.

Introduction

One of the members of the winged-helix (FOX) family of transcription factors is FOXP2 that plays a crucial role in language processing and cognition [1]. During fetal development, the protein is expressed highly in the brain and several other parts of the body like the gut and lung [2]. The protein is characterized by a DNA-binding domain of forkhead box, large polyglutamine sequence, a conserved leucine zipper and zinc finger, acidic C-terminus and a dimerization domain [3]. The protein act as a transcription factor which is evolutionarily very conserved and may bind to near about 300–400 gene promoters in the human genome. Being a transcription element, FOXP2 can control a number of genes named *DISC1*, *CNTNAP2* and *SRPX2/μPAR*, that are likely to be considered as important players for speech and language development [4]. An experiment on mice suggested the role of FOXP2 in the process of

neuron formation, differentiation and migration [5]. The neural correlation of FOXP2 was shown in patients with verbal dyspraxia caused by disruption of the FOX domain [6].

All sorts of mutations have been described in *FOXP2* gene. For example, missense mutations [7], non-sense mutations [8], deletions [9,10] and indels [11] have been discovered in different cases and families [12]. Single nucleotide polymorphism in the enciphering area of the *FOXP2* gene leads to problems like speech-language disorder 1 (SPCH1) and developmental apraxia of speech (DAS) with impaired language communication and comprehension [13]. Polymorphisms of the gene, located at the chromosome 7q31, have also been related with frontotemporal lobar degeneration [14], schizophrenia [15], autism disorder [16] and development and progression of various types of cancer such as ovarian [17], breast [18] and prostate [19] cancer.

Single nucleotide polymorphisms (SNPs) in the human genome are most available genetic variations that are single base pair changes found in every 200–300 base pairs and act as genetic markers [20]. Approximately 0.5 million SNPs, that reside in the open reading frames of the human genome [21] have drawn much interest because they produce a large number of amino acid variations that lead to functionally diverse protein variants many of which eventually lead to diseases [22]. Several previous investigations have shown that more than fifty percent of the variations associated with hereditary genetic disorders are caused by substitution of amino acids known as non-synonymous SNPs (nsSNPs) [23]. These functional variations can exert deleterious or neutral effects on protein structure or function [24]. Detrimental effects of nsSNPs might cause destabilizing protein structure, alteration of gene regulation, changing ligand-binding site [25], changing protein hydrophobicity, geometry, charge [26], modify dynamics, stability, protein-protein interactions and alter translation, resulting to threatening the cellular structural integrity [27]. It has also been reported that the involvement of multiple nsSNPs stimulate the possibility of infections, autoimmune diseases and the expansion of inflammatory disorders [20].

Though over the past couple of years, various computational methods and techniques have been applied for sorting and detecting the effects of disease-associated SNPs in other subfamilies of the target protein, limited investigations have been performed on FOXP2 protein. Considering the significant role of the protein, the present research work has been designed to identify the deleterious nsSNPs and assess their pathogenic effects on the protein using various *in silico* algorithms. Since FOXP2 regulates the transcription of various target genes; molecular docking was performed to understand the effects of the variants on target DNA binding that may lead to altered gene regulation.

Materials and methods

Retrieval of missense SNPs

The list of all missense SNPs of *FOXP2* gene, relevant information (reference SNP ID, position, changed amino acid residue and protein accession number) of the gene and corresponding protein were retrieved from NCBI dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/) and UniProt (https://www.uniprot.org/) databases [28]. Only the nsSNPs were considered for further exploration.

Identifying the most detrimental nsSNPs

The substitution of amino acid at a particular position often greatly affects the function of the protein. Five various computational tools-PROVEAN [29], PolyPhen-2 [30], SNPnexus [31], SNAP2 [32] and PON-P2 were employed to anticipate the effect of nsSNPs on FOXP2 function [33]. Based on position-specific independent count (PSIC) scores, PolyPhen-2 estimates

the impact of mutation on protein at both structural and functional levels. The difference between the scores clarifies the nsSNPs into benign, possibly damaging and probably damaging [34]. An open accessible web server PROVEAN considers the list of mutant variants and generates a score after homology searching. The threshold value ≥-2.5 labelled the SNPs into neutral or deleterious [29]. SNPnexus is comprised of with the combination of SIFT and Poly-Phen tools [31]. SIFT defines the variants as deleterious when the tolerance index score is ≤0.05. Accessible properties of the wild and mutant protein were compared by the neural network based *in silico* tool SNAP2. The tool generates a heat map and differentiates the variants as effect or neutral providing a confidence score [32]. PON-P2, a machine learning-based server, computes random forest probability score to categorize the harmful variants into unknown, neutral or pathogenic [33].

Prediction of disease related SNPs

Both SNPs&GO (DOI: 10.1002/humu.21047) and PhD-SNP (DOI: 10.1093/bioinformatics/btl423) online tools were applied to predict whether SNPs are disease associated or not. Support vector machine-based method SNPs&GO predicts the relationship of SNPs with disease at 81% accuracy. A probability score ≥0.5 points out that SNPs are related with clinical complications [35]. An online tool, PhD-SNP was also used to assess the connection of SNPs with diseases and classify them into neutral or disease associated on a 0–9 grade reliability index score [36].

Alteration of structural and functional characteristics

MutPred 2 (http://mutpred.mutdb.org/) was used to predict the impact of disease associated SNPs based on the changes in 14 different biophysical properties. The tool evaluated the possibilities of addition or loss of some of the features providing a p-values. The p-values <0.05 and <0.01 were denote as significant and very significant output respectively [37].

Protein stability prediction

A pair of software was used to check whether alterations of amino acid affect the protein stability. I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) utilizes support vector machine to analyze any change in the stable state of the protein. The tool was derived from ProTherm which is a broad dataset of experimental data on protein mutations [38]. Conditions for all selected inputs were fixed at temperature 25^0 C and pH 7.0. The outcomes provide a free energy change (DDG) and unfolding free energy value of the mutated and wild proteins [36]. MUpro (https://www.ics.uci.edu/~baldig/mutation.html) identifies the impact of single-site mutations in protein stability. A confidence score <0 dictates the reduction in stability of the protein while the score >0 interprets the opposite effect [39].

Evolutionary conservation analysis

ConSurf (https://consurf.tau.ac.il/) is an effective tool that addresses the evolutionary pattern of conservation at each amino acid site [40]. By using the Bayesian calculation method, the server analyzes close sequence homologues and estimates the phylogenetic relationship [40]. For each residue of the candidate protein, the degree of conservation is estimated on a scale of 1 to 9 and classifies these as variable (1–4), intermediate (5–6) and conserved (7–9) [41].

Prediction of post translational modifications sites (PTMs)

Diversified cellular functions like signaling cascade and protein-protein interactions are regulated by post-translational modifications (PTMs) of protein [42]. Since mutated residues can

induce allosteric or orthosteric effects that lead to shifts in energy conformations and stabilization, a deep understanding of the structure encompassing PTM sites helps to clarify the impact of PTMs on protein folding [43]. Phosphorylation, methylation, ubiquitination, acetylation, N-linked glycosylation and palmitoylation are some of the remarkable PTM that play crucial role in the study of diseases [44]. ModPred (http://www.modpred.org/), a sequence-based predictor, was applied to predict the overall influence of highly risky nsSNPs on the PTM of the target protein. This database evaluates the tendency of a specific amino acid to be modified [43].

Molecular docking between FOXP2 variants and the consensus DNA motif

RaptorX, a deep-learning based structure prediction tool was applied to build the three dimensional structures of the wild type as well as the variants (L558P, R536G, R553C and Y531H) of FOXP2 DNA binding domain (DBD) [45]. Quality of the structures were assessed using ProSA-web (PMID: 17517781) and Ramachandran plot through PDBsum (PMID: 9433130) at the EMBL-EBI site. 3D structure (in Protein Data Bank or pdb format) of the most common target site of FOXP2 (5'-CAAATT-3') in B-form was generated using the web-based DNA Sequence to Structure conversion tool [46] at the Supercomputing Facility for Bioinformatics and Computational Biology, IIT, Delhi. Binding of FOXP2 DBD 3D models to CAAATT was predicted with HDock (PMID: 28521030). DNA docked structures of FOXP2 were further analyzed using Discovery Studio (v20.1.0.19295) tool (Discovery Studio Visualizer v20.1.0.19295, San Diego, USA) to identify the interacting residues of FOXP2 with its consensus DNA motif.

Results and discussion

Retrieval of nsSNPs dataset

The dbSNP database of NCBI provides a total of 144643 SNPs data for the FOXP2 protein. Out of 144643 SNPs, 141823 were reported to be present in intron region (98.05%), 393 were missense SNPs (0.27%), 240 were synonymous (0.17%) and 2353 were non-coding transcript variant (1.63%). Since the non-synonymous SNPs often alter the encoded amino acid, the present study only considered these SNPs for further analysis. The detail information about these SNPs was given in S1 Table. The proportions of each of the SNPs category are represented in graphical form (Fig 1).

Identification of high-risk SNPs

Five computational algorithms like PolyPhen2, PROVEAN, SNPnexus, SNAP-2, PON-P2 were applied in the present study to identify detrimental nsSNPs in human FOXP2 protein. The identification outcomes with these tools are presented in Fig 2. Among the tools, Poly-Phen-2 anticipated 257 missense SNPs as probably or possibly damaging while PROVEAN determined 136 nsSNPs have deleterious effect on the proteins function. From the SIFT data of SNPnexus, 123 nsSNPs were found to be damaging. SNAP2 identified nonneutral 190 nsSNPs that might cause change in protein's activity while PON-P2 estimated 109 nsSNPs as pathogenic. The list of these SNPs is provided in S2 Table. Out of the 393 missense mutations, potent 43 nsSNPs that are found to be highly risky in all of the prediction tools were prioritized for further analysis and listed in Table 1.

Identification of disease associated deleterious SNPs

The deleterious polymorphisms were then subjected to SNPs&GO and PhD-SNP servers to predict the variants that are potentially associated with diseases or not. Out of the 43 high
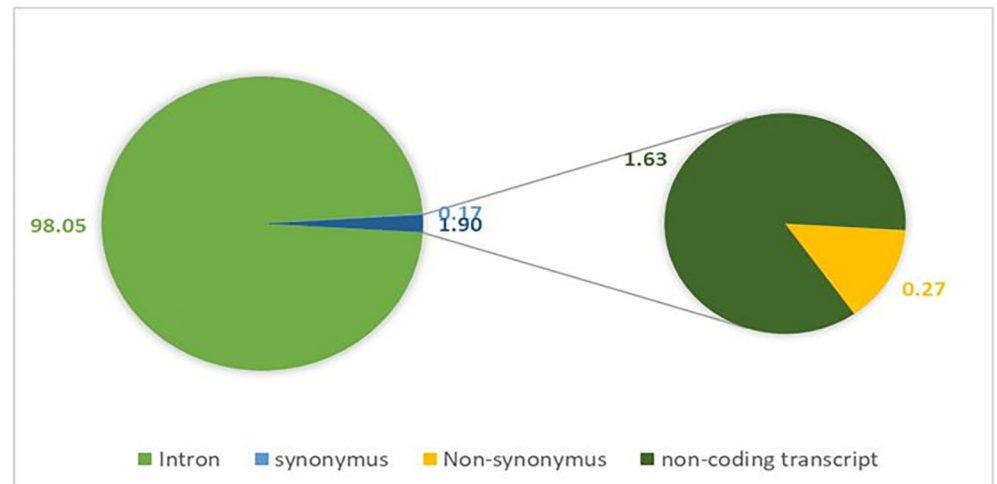
Fig 1. Percentages of various kinds of SNPs in human FOXP2 protein. Intronic SNPs: 98.05%; nsSNPs: 0.28%; synonymous SNPs: 0.17%; non-coding transcript: 1.63%.

https://doi.org/10.1371/journal.pone.0272625.g001

confidence nsSNPs, SNPs&GO and PhD-SNP server designated 11 and 14 nsSNPs to have potential role in diseases. Among these, 5 nsSNPs were common in both tools and selected for further investigations (Table 2).

Functional and structural modification prediction

The selected disease related nsSNPs were analyzed by MutPred 2 server to understand their molecular mechanism of pathogenicity (Table 3). The predicted modifications of structural and functional properties include alteration of disordered interface, transmembrane protein,

Fig 2. Number of predicted damaging nsSNPs in human FOXP2. The highest number of damaging nsSNPs were screened by Polyphen 2 and PON-P2 predicted the lowest number.

https://doi.org/10.1371/journal.pone.0272625.g002

Table 1. High-risk non-synonymous SNPs predicted by five computational tools.

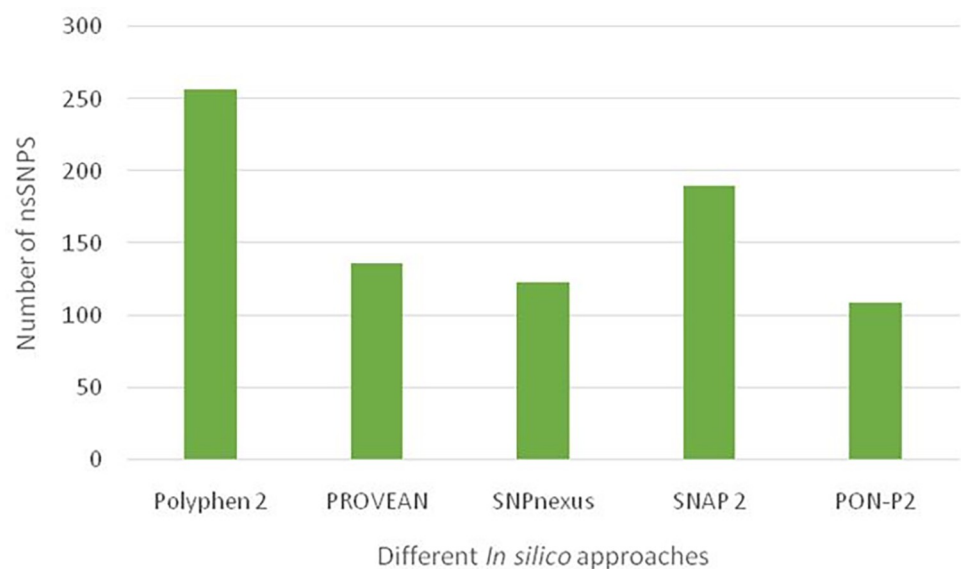| rsID	Position	Provean	SIFT (SNPnexus)	Polyphen-2	SNAP2	PON P2
rs121908377	R553H	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs797045587	M94I	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs879253772	Y531H	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs112732214	L558P	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs199776572	T451M	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs377420314	Y604C	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs566961630	R553C	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs745342916	D375G	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs751931499	V690G	Deleterious	Deleterious	Probably damaging	Effect	pathogenic
rs755297474	E334V	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs758427088	D644H	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs758513311	S75F	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs762564041	R536G	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs765157455	G319E	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs765157455	Q117H	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs772694863	W270R	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs779754644	S305F	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs779921362	E676K	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs889341368	G473R	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs948249504	S255W	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1158865993	P697L	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1175210435	P115S	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1183578823	R376L	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs1184039801	W270C	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1191637371	N567I	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1219589831	N694Y	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1224336230	V563A	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1250616703	L269F	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1276471970	N424K	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1290398957	H340Y	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs1295112601	L236F	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs1302127838	M406K	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs1330529378	P677S	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1343377230	F507L	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1362466494	L291P	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1375575897	Y604N	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1383948441	S315F	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1394757420	V112F	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1428334171	P486H	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1436939063	E334K	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs1445779721	K365M	Deleterious	Deleterious	Probably damaging	Effect	Pathogenic
rs1447805795	E334K	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic
rs1459605752	H651R	Deleterious	Deleterious	Possibly damaging	Effect	Pathogenic

metal binding, coiled coil, stability, gain of loop, B factor, intrinsic disorder, amidation, loss of helix as well as loss and gain of glycosylation. L291P variant significantly gain GPI-anchor amidation and N-linked glycosylation. L558P was predicted to induce loss of helix and N-linked

Table 2. Disease related non-synonymous SNPs identified by SNPs&GO and PhD-SNP.

rsID	Variant	Mutation	SNP&GO	PhD-SNP
rs879253772	T/C	Y531H	Disease	Disease
rs112732214	T/C	L558P	Disease	Disease
rs566961630	C/T	R553C	Disease	Disease
rs762564041	C/T	R536G	Disease	Disease
rs1362466494	T/C	L291P	Disease	Disease

RI = Reliability Index, P = Probability.

https://doi.org/10.1371/journal.pone.0272625.t002

glycosylation with p value 0.04 and 0.05 respectively. R536G variants also predicted to have loss of N-linked glycosylation.

Effect on protein stability

Both I-Mutant 2.0 and MUpro servers were used to interpret whether the protein will be in stable or denatured form due to point mutation. The result expressed either in the free energy change value ($\Delta\Delta G$) or in the sign of DDG where positive and negative DDG value indicates increased or decreased stability, respectively. All of the variants represented in Table 4 have negative DDG value indicating that they reduce the stability of the variants.

Table 3. Impacts of nsSNPs on structural & functional properties of FOXP2.

Mutation	Probability of deleterious mutation	Structural & functional properties
Y531H	0.897	Altered coil coil (P = 0.53)
		Altered ordered interface (P = 0.37)
		Altered disordered interface (P = 0.31)
		Altered transmembrane protein (P = 0.25)
L291P	0.573	Gain of intrinsic disorder (P = 0.40)
		Gain of loop (P = 0.27)
		Gain of B factor (P = 0.1797)
		Gain of GPI-anchor amidation at N-295 (**P = 0.05**)
		Gain of N linked glycosylation at N294 (**P = 0.02**)
L558P	0.923	Loss of helix (P = 0.04)
		Altered disordered interface (P = 0.24)
		Altered transmembrane protein (P = 0.20)
		Altered Stability (P = 0.12)
		Loss of N linked glycosylation at N555 (**P = 0.05**)
R553C	0.9	Altered Metal binding (P = 0.39)
		Altered Disordered interface (P = 0.38)
		Loss of Helix (P = 0.29)
		Altered Ordered interface (P = 0.27)
		Altered Coiled coil (P = 0.25)
R536G	0.914	Altered Transmembrane protein (P = 0.18)
		Loss of N-linked glycosylation at N555 (**P = 0.04**)
		Altered Coiled coil (P = 0.48)
		Altered Disordered interface (P = 0.42)
		Altered Transmembrane protein (P = 0.31)
		Altered Ordered interface (P = 0.30)
		Loss of Helix (P = 0.29)

https://doi.org/10.1371/journal.pone.0272625.t003

Table 4. Impact of single nucleotide polymorphisms on the stability of protein.

dbSNP	Amino acid changes	Mu-pro		I-Mutant	
		Prediction	DDG value	Prediction	DDG value
rs879253772	Y531H	Decrease	-1.244777	Decrease	-1.34
rs112732214	L558P	Decrease	-2.0758766	Decrease	-1.77
rs566961630	R553C	Decrease	-0.71979678	Decrease	-0.57
rs762564041	R536G	Decrease	-1.2125418	Decrease	-1.63
rs1362466494	L291P	Decrease	-1.4678293	Decrease	-1.24

DDG = Free energy change value, RI = Reliability index.

https://doi.org/10.1371/journal.pone.0272625.t004

Prediction of post translational modifications

Out of 5 nsSNPs, the Modpred server reported only 2 nsSNPs (Y531H and R536G) to be present at PTM sites and 1 variant (R553C) at amidation sites. Table 4 depicts the post translational modification (PTM) of altered amino acids.

Sequence conservation study of nsSNPs

The variants Y531H, L558P, R536G and L291P were reported by consurf to be highly conserved as structural and buried and R553C was predicted to be conserved being functional and exposed. The server estimated the confidence level by giving a conservation score (1–9) for the conservation of sequence. The functional and structural consequences of 5 high-risk nsSNPs with their phylogenetic conservation scores are given in S1 Fig and Table 5.

Structural validation of wild and mutant models

The four mutants Y531H, L558P, R553C, R536G are present in the winged-helix DNA binding domain (DBD) of the FOXP2 protein. So, the native DBD structure of the protein and its four detrimental, disease associated variants were using RaptorX. Model quality validation output file of Ramachandran plot and ProSA Z- score were described in Table 6. The Ramachandran plot depicted that 96.1% residues were in most favored region whereas 0% in disallowed region among the 82 (503–584) residues in native DBD of the FOXP2 protein. Over 94% residues also occupy most favorable region in the predicted DBD structure of the 4 variants. ProSA Z- score and Ramachandran plot indicate the overall good quality of the 3D modelling.

    Docking analysis. The docking analysis of FOXP2 DBD with its most common target sequence reported that the mutant variants bind to the DNA sequence in a slightly diverged orientation than wild type variants of DBD. Upon binding, the native protein shows interaction with Arg 504, Tyr 509, His 554, Leu 558, Pro 505, Trp 548 and Ala 551residues. Positively

Table 5. Conservation profile of nsSNPs in FOXP2 protein and their post translational modification.

rsID	Residue & position	Conservation score	Prediction	PTMs
rs879253772	Y531H	9	Highly conserved and buried (s)	Proteolytic cleavage
rs112732214	L558P	9	Highly conserved and buried (s)	-
rs566961630	R553C	9	Highly conserved and exposed (f)	Amidation
rs762564041	R536G	9	Highly conserved and buried (s)	Proteolytic cleavage
rs1362466494	L291P	9	Highly conserved and buried (s)	

PTMs = Post translational modification.

https://doi.org/10.1371/journal.pone.0272625.t005

Table 6. The results of three-dimensional model validation.

PDBSum	Most favoured region (%)	Additional allowed region (%)	Generously allowed region (%)	Disallowed region (%)	ProSA web (Z-score)
FOXP2 DBD (WT)	96.1	2.6	1.3	0	-5.81
FOXP2 DBD (L558P)	94.7	5.3	0	0	-6.5
FOXP2 DBD (R536G)	97.4	1.3	1.3	0	-6.22
FOXP2 DBD (R553C)	96.1	2.6	0	1.3	-6.2
FOXP2 DBD (Y531H)	98.7	0	1.3	0	-6.18

https://doi.org/10.1371/journal.pone.0272625.t006

charged Arg 504 shows electrostatic interaction with Thymine10′ and Tyr 509 forms hydrogen bond with Thymine5. All of the variants form this type of bond with Arg 504 and Tyr 509 on average 0.05 or marginal deviation of distance respectively. Y531H missense variant interacts with the same amino acid residues except Ala 551. In native protein, Ala 551 form pi-alkyl type hydrophobic bond with Adenine8′ but Adenine8′ forms carbon hydrogen bond to Thr 547 in Y531H variants. Thr 547 makes contact with Adenine8′ and due to substitution of Leu to Pro, Thymine6 binds with Pro 558 and His 559 in L558P mutants. Similar pattern of interactions was also observed in R536G variants. Due to alteration of arginine to cysteine, Phe forms Pi-Amino bond with Thymine10′ at position 553. The details of the interactions have been described in Fig 3 and S3 Table.

The winged helix DNA binding domain of FOXP2 is crucial for analysis as majority of the disease-causing mutations are found in that region [47]. The mutation of arginine to histidine at 553th position in human FOXP2 has been associated with severe congenital speech impairement disorder which plays a vital role in the FOXP2-DNA interaction [48]. Most of the interactions between deoxyribonucleic acid and transcription factors are between the negatively charged phosphodiester bond of the DNA and positively charged amino acid residues of the protein [49]. Banerjee-Basu and Baxervanis in 2004 observed that the R553H mutation causes a net reduction of the electrostatic potential in the DNA binding surface of the protein. Due to the disturbance to the electrostatics of the protein the interactions between the backbone of the DNA and protein might be hampered. Several studies also suggest that electrostatic interactions also dominate the protein-DNA binding specificity [50]. In case of R553C mutation, a new electrostatic bond is formed between Phe 541 and Thy 10′ which may change the electrostatic potential that likely led to imbalance in the organization of protein -DNA interface.

From various mutational investigations it was known that His554 plays a significant role in hydrogen bond formation between FOXP2 FHD and DNA [51]. This interaction was also conserved in other FOX DBDs like FOXO1, FOXK1 and FoxA1 [52]. The substitution of leucine to proline at 558th position missed the interaction with His554, which may lead to the reduction of specificity of the protein with its cognate sequence. Since hydrogen bond stabilizes the protein-DNA interaction [53], the missing of most conserved H-bond might destabilize the interaction. The more interactions between the DNA and protein may predict the higher affinity of binding. Upon binding, the native protein shows interaction with Arg 504, Tyr 509, His 554, Leu 558, Pro 505, Trp 548, Ala 551 residues. L558P mutation reduces the interactions with other residues compared to native protein which may lead to lower affinity of protein-DNA binding.

Conclusion

With the combination of various *in silico* tools, the study determined the deleterious nsSNPs in FOXP2 protein. The outcomes predicted R553C and L558P mutants as the most damaging
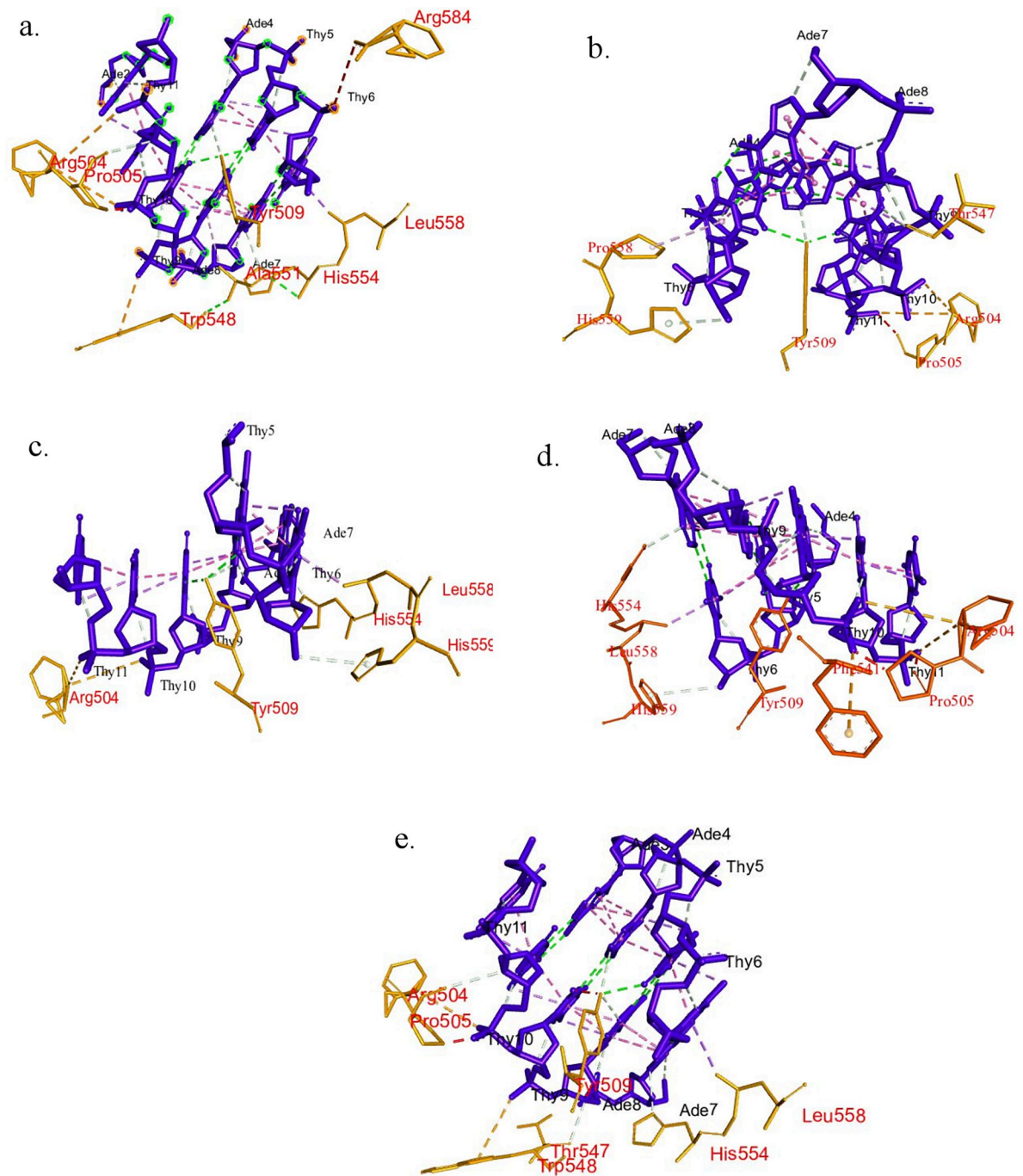
Fig 3. Molecular docking analysis of native and mutant DBD domain of FOXP2 with novel sequence. Interactions of DNA binding domain with (a) Native FOXP2 (b) Mutant L558P (c) R536G (d) R553C (e) Y531H.

nsSNPs that may alter protein-DNA interface interactions, reduce the specificity and affinity of the binding. This kind of structural and functional alterations is likely to interrupt regulation of protein and initiate formation of disease. In-depth experimental verification is needed to determine the role of mutant variants more precisely.

Supporting information

S1 Fig. Prediction of ConSurf about the conservation status of amino acids in human FOXP2 protein.
(TIFF)

S1 Table. The list of reference sequence ID, alteration of allele, protein accession no, position of amino acid and residual change of all 393 non-synonymous SNPs.
(DOCX)

S2 Table. The results of five *in silico* tools that are used to analyze all of the 393 non-synonymous SNPs.
(DOCX)

S3 Table. Interactions of amino acid residues with native and mutant FoxP2 protein.
(DOCX)

Author Contributions

Conceptualization: Sumaiya Farah Khan, Fahmida Sultana Rima.

Data curation: Mahmuda Akter, Sumaiya Farah Khan, Abu Ashfaqur Sajib, Fahmida Sultana Rima.

Formal analysis: Mahmuda Akter, Abu Ashfaqur Sajib.

Methodology: Sumaiya Farah Khan, Fahmida Sultana Rima.

Supervision: Sumaiya Farah Khan, Abu Ashfaqur Sajib.

Validation: Fahmida Sultana Rima.

Writing – original draft: Mahmuda Akter.

Writing – review & editing: Sumaiya Farah Khan, Abu Ashfaqur Sajib.

References

1. Reuter MS, Riess A, Moog U, Briggs TA, Chandler KE, Rauch A, et al. FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. J Med Genet. 2017; 54(1): 64–72. https://doi.org/10.1136/jmedgenet-2016-104094 PMID: 27572252

2. Atkinson EG, Audesse AJ, Palacios JA, Bobo DM, Webb AE, Ramachandran S, et al. No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. Cell [Internet]. 2018/08/02. 2018 Sep 6; 174(6):1424–1435.e15. Available from: https://pubmed.ncbi.nlm.nih.gov/30078708. https://doi.org/10.1016/j.cell.2018.06.048 PMID: 30078708

3. Thulo M, Rabie MA, Pahad N, Donald HL, Blane AA, Perumal CM, et al. The influence of various regions of the FOXP2 sequence on its structure and DNA-binding function. Biosci Rep. 2021; 41(1): 1–13. https://doi.org/10.1042/BSR20202128 PMID: 33319247

4. Meredith LJ, Wang CM, Nascimento L, Liu R, Wang L, Yang WH. The Key Regulator for Language and Speech Development, FOXP2, is a Novel Substrate for SUMOylation. J Cell Biochem. 2016; 117(2): 426–38. https://doi.org/10.1002/jcb.25288 PMID: 26212494

5. Bowers JM, Perez-Pouchoulen M, Roby CR, Ryan TE, McCarthy MM. Androgen modulation of Foxp1 and Foxp2 in the developing rat brain: impact on sex specific vocalization. Endocrinology [Internet]. 2014/09/23. 2014 Dec; 155(12):4881–94. Available from: https://pubmed.ncbi.nlm.nih.gov/25247470. https://doi.org/10.1210/en.2014-1486 PMID: 25247470

6. Chabout J, Sarkar A, Patel SR, Radden T, Dunson DB, Fisher SE, et al. A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. Front Behav Neurosci. 2016; 10(OCT): 1–18. https://doi.org/10.3389/fnbeh.2016.00197 PMID: 27812326

7. Laffin JJS, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. Genet Med. 2012; 14(11): 928–36. https://doi.org/10.1038/gim.2012.72 PMID: 22766611

8. MacDermot KD, Bonora E, Sykes N, Coupe A-M, Lai CSL, Vernes SC, et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet. 2005; 76(6): 1074–80.

9. Palka C, Alfonsi M, Mohn A, Cerbo R, Franchi PG, Fantasia D, et al. Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. Pediatrics. 2012; 129(1): e183–8. https://doi.org/10.1542/peds.2010-2094 PMID: 22144704

10. Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama-Kurtycz CM, Hartley SL, et al. Phenotype of FOXP2 haploinsufficiency in a mother and son. Am J Med Genet Part A. 2012; 158(1): 174–81.

11. Turner SJ, Hildebrand MS, Block S, Damiano J, Fahey M, Reilly S, et al. Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. Am J Med Genet Part A. 2013; 161(9): 2321–6. https://doi.org/10.1002/ajmg.a.36055 PMID: 23918746

12. Hoogman M, Guadalupe T, Zwiers MP, Klarenbeek P, Francks C, Fisher SE. Assessing the effects of common variation in the FOXP2 gene on human brain structure. Front Hum Neurosci. 2014; 8(JULY): 1–9. https://doi.org/10.3389/fnhum.2014.00473 PMID: 25013396

13. Oswald F, Klöble P, Ruland A, Rosenkranz D, Hinz B, Butter F, et al. The FOXP2-Driven network in developmental disorders and neurodegeneration. Front Cell Neurosci. 2017; 11(July): 1–24.

14. Torres-Ruiz R, Benítez-Burraco A, Martínez-Lage M, Rodríguez-Perales S, García-Bellido P. Functional characterization of two enhancers located downstream FOXP2. BMC Med Genet. 2019; 20(1): 1–12.

15. Li T, Zeng Z, Zhao Q, Wang T, Huang K, Li J, et al. FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han Population. World J Biol Psychiatry. 2013; 14(2): 146–50. https://doi.org/10.3109/15622975.2011.615860 PMID: 22404659

16. Bowers JM, Konopka G. The role of the FOXP family of transcription factors in ASD. Dis Markers. 2012; 33(5): 251–60. https://doi.org/10.3233/DMA-2012-0919 PMID: 22960337

17. Ying H, Lyu J, Ying T, Li J, Jin S, Shao J, et al. RETRACTED ARTICLE: Risk miRNA screening of ovarian cancer based on miRNA functional synergistic network. J Ovarian Res. 2014; 7(1): 1–9.

18. Cuiffo BG, Campagne A, Bell GW, Lembo A, Orso F, Lien EC, et al. MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis. Cell Stem Cell. 2014; 15(6): 762–74. https://doi.org/10.1016/j.stem.2014.10.001 PMID: 25515522

19. Stumm L, Burkhardt L, Steurer S, Simon R, Adam M, Becker A, et al. Strong expression of the neuronal transcription factor FOXP2 is linked to an increased risk of early PSA recurrence in ERG fusion-negative cancers. J Clin Pathol. 2013; 66(7): 563–8. https://doi.org/10.1136/jclinpath-2012-201335 PMID: 23559350

20. Islam MJ, Khan AM, Parves MR, Hossain MN, Halim MA. Prediction of Deleterious Non-synonymous SNPs of Human STK11 Gene by Combining Algorithms, Molecular Docking, and Molecular Dynamics Simulation. Sci Rep [Internet]. 2019; 9(1):1–16. Available from: https://doi.org/10.1038/s41598-019-52308-0

21. Alshatwi AA, Hasan TN, Syed NA, Shafi G, Grace BL. Identification of Functional SNPs in BARD1 Gene and In Silico Analysis of Damaging SNPs: Based on Data Procured from dbSNP Database. PLoS One. 2012; 7(10): 1–10. https://doi.org/10.1371/journal.pone.0043939 PMID: 23056176

22. Ul Hasnain MJ, Shoaib M, Qadri S, Afzal B, Anwar T, Abbas SH, et al. Computational analysis of functional single nucleotide polymorphisms associated with SLC26A4 gene. PLoS One. 2020; 15(1): 1–20.

23. Mitra S, Das R, Hamza A, Dash R. In silico analysis of nonsynonymous single-nucleotide polymorphisms (nsSNPs) of the SMPX gene. 2019;(April):1–18. https://doi.org/10.1111/ahg.12350 PMID: 31583691

24. Capriotti E, Altman RB. Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics. 2011; 12(SUPPL. 4). https://doi.org/10.1186/1471-2105-12-S4-S3 PMID: 21992054

25. Khan FI, Aamir M, Wei D-Q, Ahmad F, Hassan MI. Molecular mechanism of Ras-related protein Rab-5A and effect of mutations in the catalytically active phosphate-binding loop. J Biomol Struct Dyn. 2017; 35(1): 105–18. https://doi.org/10.1080/07391102.2015.1134346 PMID: 26727234

26. Petukh M, Kucukkal TG, Alexov E. On human disease-causing amino acid variants: Statistical study of sequence and structural patterns. Hum Mutat. 2015; 36(5): 524–34. https://doi.org/10.1002/humu.22770 PMID: 25689729

27. Arshad M, Bhatti A, John P. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. PLoS One. 2018; 13(1): 1–13. https://doi.org/10.1371/journal.pone.0188143 PMID: 29329296

28. Bhagwat M. Searching NCBI's dbSNP database. Curr Protoc Bioinforma. 2010; 32(1): 1–19.

29. Choi Y, Chan AP. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015; 31(16): 2745–7. https://doi.org/10.1093/bioinformatics/btv195 PMID: 25851949

30. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4): 248–9. https://doi.org/10.1038/nmeth0410-248 PMID: 20354512

31. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: A web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012; 40(W1): 65–70. https://doi.org/10.1093/nar/gks364 PMID: 22544707

32. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, De Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24(24): 2938–9. https://doi.org/10.1093/bioinformatics/btn564 PMID: 18974171

33. Niroula A, Urolagin S, Vihinen M. PON-P2: Prediction method for fast and reliable identification of harmful variants. PLoS One. 2015; 10(2): 1–17.

34. Mustafa HA, Albkrye AMS, AbdAlla BM, Khair MAM, Abdelwahid N, Elnasri HA. Computational determination of human PPARG gene: SNPs and prediction of their effect on protein functions of diabetic patients. Clin Transl Med [Internet]. 2020; 9(1): 1–10. Available from: https://doi.org/10.1186/s40169-020-0258-1

35. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics [Internet]. 2013; 14 Suppl 3(Suppl 3):S6. Available from: http://www.biomedcentral.com/1471-2164/14/S3/S6.

36. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics. 2006; 22(22): 2729–34. https://doi.org/10.1093/bioinformatics/btl423 PMID: 16895930

37. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 2009; 25(21): 2744–50. https://doi.org/10.1093/bioinformatics/btp528 PMID: 19734154

38. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005; 33(SUPPL. 2): 306–10.

39. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins Struct Funct Bioinforma. 2006; 62(4): 1125–32. https://doi.org/10.1002/prot.20810 PMID: 16372356

40. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016; 44(W1): W344–50. https://doi.org/10.1093/nar/gkw408 PMID: 27166375

41. Jia M, Yang B, Li Z, Shen H, Song X, Gu W. Computational analysis of functional single nucleotide polymorphisms associated with the CYP11B2 gene. PLoS One. 2014; 9(8): e104311. https://doi.org/10.1371/journal.pone.0104311 PMID: 25102047

42. Nussinov R, Tsai C-J, Xin F, Radivojac P. Allosteric post-translational modification codes. Trends Biochem Sci. 2012; 37(10): 447–55. https://doi.org/10.1016/j.tibs.2012.07.001 PMID: 22884395

43. Pejaver V, Hsu W, Xin F, Dunker AK, Uversky VN, Radivojac P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci. 2014; 23(8): 1077–93. https://doi.org/10.1002/pro.2494 PMID: 24888500

44. Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, et al. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res. 2020 Jul; 48 (W1): W140–6. https://doi.org/10.1093/nar/gkaa275 PMID: 32324217

45. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc. 2012; 7(8): 1511–22. https://doi.org/10.1038/nprot.2012.085 PMID: 22814390

46. Fasman GD. Handbook of biochemistry and molecular biology. Cleveland, Ohio: CRC Press; 1976.

47. Perumal K, Dirr HW, Fanucchi S. A Single Amino Acid in the Hinge Loop Region of the FOXP Forkhead Domain is Significant for Dimerisation. Protein J. 2015; 34(2): 111–21. https://doi.org/10.1007/s10930-015-9603-4 PMID: 25724387

48. Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, et al. Structure of the forkhead domain of FOXP2 bound to DNA. Structure. 2006; 14(1): 159–66. https://doi.org/10.1016/j.str.2005.10.005 PMID: 16407075

49. Luscombe NM, Austin SE, Berman HM, Thornton JM. An overview of the structures of protein-DNA complexes. Genome Biol. 2000; 1(1): 1–37. https://doi.org/10.1186/gb-2000-1-1-reviews001 PMID: 11104519

50. Ficici E, Andricioaei I. On the Possibility of Facilitated Diffusion of Dendrimers Along DNA. J Phys Chem B. 2015; 119(23): 6894–904. https://doi.org/10.1021/acs.jpcb.5b02090 PMID: 25989627

51. Blane A, Fanucchi S. Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain. Biochemistry. 2015 Jun 9; 54. https://doi.org/10.1021/acs.biochem.5b00155 PMID: 26055196

52. Tsai K-L, Huang C-Y, Chang C-H, Sun Y-J, Chuang W-J, Hsiao C-D. Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. J Biol Chem. 2006; 281(25): 17400–9. https://doi.org/10.1074/jbc.M600478200 PMID: 16624804

53. Coulocheri SA, Pigis DG, Papavassiliou KA, Papavassiliou AG. Hydrogen bonds in protein–DNA complexes: where geometry meets plasticity. Biochimie. 2007; 89(11): 1291–303. https://doi.org/10.1016/j.biochi.2007.07.020 PMID: 17825469