International Journal of Theoretical & Applied Sciences, 12(1): 43-48(2020)

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

In Silico Analysis of the Functional and Structural Impact of SNPs in the *CD46* Gene

Chhaya Bawa
Zoology Department, DDE,
Annamalai University, Annamalai Nagar, Chidambaram, (Tamil Nadu), India.

(Corresponding author: Chhaya Bawa)
(Received 07 April, 2020, accepted 22 June, 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Innate immunity plays a significant role in natural immune response. Complement system is a critical player in innate immunity and tightly regulated by various regulatory components. The present study was undertaken to predict deleterious/damaging SNPs using *in silico* analysis of *CD 46* gene. Various tools like SIFT, PROVEAN, SNAP2, PredictSNP, I-mutant and Consurf were used for *in silico* analysis. CD46 study revealed 263 coding snSNPs. The SIFT sequence tool predicted a total of 85 variants, PROVEAN predicted 86 nsSNPs (out of 263) as deleterious and SNAP2 server predicted 115 variants to be effective in influencing the protein function. I mutant was used to calculate the free energy change value and reliability index to predict the functional deviations in proteins. The I-mutant revealed 37 deleterious mutations. Consurf results predicted conservation of residues in protein sites. While active site prediction analysis revealed 11 amino acid residues in chain B and 32 in D as active site residues in protein structure. Collectively these results might be useful in selection of target SNPs for genotyping in disease association studies and could pave the way for future experimental studies.

Keywords: CD46, Complement system, In silico Analysis, nsSNPs

I. INTRODUCTION

The complement system is among the most important components of the innate immune system. In vertebrate the complement system involves a sequence of various proteins which acts in a sequence and work mainly through three major routes [1]. These proteins and their products are involved significantly in activation of inflammatory and pathogen killing responses. The complement mediated responses are tightly regulated by the complement regulatory protein [2]. Over activation of complement system or uncontrolled regulation of complement system can lead to the various pathological conditions which can be life threatening. One of the major reasons for uncontrolled response is pathological mutations in the complement regulatory proteins [3]. One of the major complement regulatory proteins is CD46 [4]. It is a complement regulator protein and is a complement inhibitor receptor. Its gene is located on chromosome 1q32 with other genes responsible for structural components of the complement system [5]. CD46 has been documented for inactivation of complement part C3b and C4b by serum factor 1 [6]. Serum factor one is known to protect the host cell from damage from complement. The structure of CD46 has not been fully described yet but CD46 has been documented to play a role in various clinical pathological conditions [7].

Single nucleotide polymorphisms (SNPs) are a very critical feature in the genome, and are proposed to play a significant role towards vulnerability to many diseases [8]. The most of genetic variations may vary from neutral to pathogenic. Sometimes a single base alteration in and around a gene can alter expression and influence the role of its protein products [9]. A non-synonymous SNP is a single base pair modification in a coding region that alter an amino acid in its corresponding protein. If nsSNP alters protein function, it can result in pathophysiological state of the disease. Nonsense variants, associated with premature termination, were most likely to be related with diseases with 2.77% probability. Interestingly, 1.46% of nsSNPs, 1.38% of SNPs within the 5′-UTR region, and 1.26% of sSNPs (1.26%) have also been known to be associated with human disease [10]. The activity of proteins and susceptibilities of various diseases can be influenced by the genotype and genetic variations. It has been shown that CD46 protein plays a role and acts as a receptor for various pathogenic viruses like measles virus [11]. Therefore, in the current study efforts were made *in silico* analysis of various variations in CD 46.

II. MATERIAL AND METHODS

Dataset Collection. The SNP dataset for CD46 gene was obtained from the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/snp) for further computational analysis. The protein sequence was retrieved from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein).

Prediction of Deleterious nsSNPs. Functional aspects of nsSNPs were predicted using the *in-silico*

algorithms. SIFT (Sorting Intolerant From Tolerant) was used (http://sift.jcvi.org/) to predict the impact of amino acid replacement on protein functions. It predicts the effects of an amino acid substitution on the function of a protein based on the sequence resemblance and physical behaviour of amino acids. SIFT predicts harmful SNPs based on the extent of conserved amino acid. SIFT links residues in matched sequences to closely related sequences found using PSI-BLAST with a score of 0 to 1. Score ≤0.05 are predicted as damaging.

PROVEAN (http://provean.jcvi.org/index.php) tool was used to predict whether an amino acid substitution has an effect on protein biological functions. PROVEAN facilitates the creation of precomputed predictions by assisting us in obtaining pairwise sequence alignment scores. Without modifying the default settings, the protein sequence and amino acid mutations were introduced. Based on their alignment score of greater or less than 2.5, mutations are projected to have neutral or detrimental effects.

Sequence Conservation. Phylogenetic analysis of SNPs found in the conserved region of CD46 gene was determined using ConSurf web-server (consurf.tau.ac.il/) [12]. ConSurf determined the evolutionarily conserved regions by aligning the FASTA sequence of CD46 to homologs. An empirical Bayesian approach was used to calculate position-specific scores. With the help of a coloring scheme, the preserved regions were projected in terms of conservation scores and then separated into nine different scales. The amino acids with a score of 7 to 9 were considered evolutionary conservative.

Prediction of Change in Protein Stability. The alteration in protein stability associated with nsSNPs was predicted using I-Mutant 2.0 (http://folding.biofold.org/cgi-bin/i-mutant 2.0), a web-based tool. This web server is meant for automatic prediction of changes in protein stability and output obtained in terms of free energy change value (DDG) along with prediction as increase or decrease. DDG values are resulted from the difference in unfolding Gibbs free energy value of the mutated protein and wild type in kcal/mol [13].

III. RESULTS AND DISCUSSION

The CD46 (P15529) consists of 1179 bp and 392 amino acids. The CD46 investigated in this study had a total of 263 coding nsSNPs. These non-synonymous coding SNPs were used for further investigation since non synonymous mutations might modify protein sequences, affecting the structure and function of the protein.

Prediction of deleterious coding nsSNPs. A total of 85 variants with an influence on protein function were predicted using the SIFT sequence algorithm while 178 variants had no effect on CD46. Overall, 178 nsSNPs with a score of less than 0.05 were found to be tolerated, while 6 nsSNPs with a score of 0.01, 13 nsSNPs with a score of 0.02 and also 13 more nsSNPs with a score of 0.03, and 3 nsSNPs with a score of 0.04 were identified as deleterious. Furthermore, remaining 48 nsSNPs had a highly deleterious predictions with a PROVEAN score below -2.5, 86 nsSNPs (out of 263) were predicted to be harmful, while the remaining nsSNPs (177) with values above the cutoff were classified as neutral. The PROVEAN tool utilizes a cut-off score of -2.5 for all predictions.

The query protein's amino acid sequence, mutation sites, and intended mutations were entered into the PredictSNP input page in FASTA format. In CD46, PredictSNP predicted 53 mutations to be harmful, whereas the remaining 210 were confirmed to be neutral. The results from the SNAP2 server predicted that 115 variations would be effective, whereas the other 148nsSNPs would be neutral. By combining the observations of four prediction tools (SIFT, PROVEAN, PredictSNP and SNAP2), 37nsSNPs (R6L, L20P, L28Q, C35Y, Y54H, G57A, Y62C, G67V, Y68H, C94S, P105H, G130V, Y131H, G135D, P155A, P165L, P165S, I168M, N170K, N170S, Y189D, C191Y, D192Y, L202F, C210Y, W216G, S217G, P221S, P231R, G236A, S240P, F246C, Y248H, G259R, L262P, W276S and P329R) have been anticipated to influence protein function by all software tools (Table 1). These nsSNPs were used for further analyses. Previously also many variations inCD46 gene has been co related with nephritis [14]. Hemolytic uremic syndrome has been also corelated to the pathogenic variation in CD46 previously [15].

Prediction of mutation impacts on the stability of proteins. The selected variants were uploaded to the I-Mutant 2.0 web server in predicting the free energy change value (DDG) and reliability index (RI) during mutation. The outcomes of amino acid substitutions indicated either a gain or a loss in free energy, as per I-Mutant 2.0. Except for N170S, all of the changed nsSNPs after mutation resulted in a decrease in protein stability, with a reliability index ranging from 0 to 10. Table 2 shows the I-Mutant 2.0 prediction of changes in stability for the 37 harmful nsSNPs. This finding suggests that these CD46 mutations may disrupt amino acid interactions directly or indirectly, resulting in protein functional deviations. Changes in CD46 protein have been also shown to impact CD 46 function in previously [16].

Conservation of amino acids. The ConSurf tool produces a structural representation of the protein together with a colorimetric conservation score. ConSurf projected 17 amino acids with conservation score 9, 8 amino acids with conservation score 8, and 4 amino acids with conservation score 7 among the most damaging SNPs. Positions P105 and D192 were predicted with conservation score 5 (Table 3). Except for P105 and D192, all other locations were expected to be in highly conserved regions, indicating that there are more opportunities to change the protein structure. Residues that are highly conserved are frequently required for biological function.

Active site identification

From the active site prediction, a pocket was identified with an area (SA) of 1498.892 and a volume (SA) of 3553.701 (Figure 1). A total of 11 amino acid residues in chain B and 32 in D protein structure were predicted

to be the active sites for the CD46 protein structure (Table 4). There is no mutation in these active sites.

Table 1: nsSNPs predicted to be deleterious by at least two programs (SIFT, PROVEAN, SNAP2 and PredictSNP) in CD46.

Amino acid change	SIFT	PROVEAN	SNAP2	PredictSNP
R6L	0.01	-2.923	27	51
L20P	0.02	-4.291	86	55
L28Q	0.03	-2.902	45	61
C35Y	0.0	-10.172	74	87
Y54H	0.0	-4.667	80	61
G57A	0.0	-5.754	20	87
Y62C	0.0	-8.211	75	87
G67V	0.0	-8.325	76	87
Y68H	0.0	-4.273	80	72
C94S	0.0	-9.33	58	76
P105H	0.01	-8.142	30	87
G130V	0.0	-8.982	68	87
Y131H	0.0	-4.561	72	61
G135D	0.0	-6.989	80	87
P155A	0.0	-7.986	41	61
P165L	0.0	-9.909	48	72
P165S	0.0	-7.927	59	76
I168M	0.0	-2.961	42	61
N170K	0.0	-5.806	66	76
N170S	0.0	-4.845	39	51
Y189D	0.0	-9.776	90	87
C191Y	0.0	-10.811	79	87
D192Y	0.03	-5.626	61	61
L202F	0.02	-3.934	75	76
C210Y	0.0	-10.799	87	87
W216G	0.0	-12.276	88	87
S217G	0.0	-3.934	73	55
P221S	0.0	-7.867	83	87
P231R	0.0	-8.393	53	87
G236A	0.0	-5.6	41	51
S240P	0.02	-4.045	65	61
F246C	0.0	-6.166	44	87
Y248H	0.0	-4.576	77	51
G259R	0.0	-7.23	80	87
L262P	0.01	-5.26	81	87
W276S	0.0	-13.049	87	87
P329R	0.04	-2.64	50	55

SIFT - Sorting Intolerant from Tolerant, PROVEAN - Protein Variation Effect Analyzer, SNAP2 - Screening for Non-acceptable Polymorphisms, PredictSNP - Consensus classifiers for prediction of disease-related mutations.

Table 2: Prediction of protein stability by I-Mutant due to mutations in CD46.

Position	WT	NEW	Stability	RI
6	R	L	Decrease	6
20	L	P	Decrease	2
28	L	Q	Decrease	7
35	C	Y	Decrease	2
54	Y	H	Decrease	8
57	G	A	Decrease	7
62	Y	C	Decrease	4
67	G	V	Decrease	4
68	Y	H	Decrease	7
94	C	S	Decrease	7
105	P	H	Decrease	9
130	G	V	Decrease	1
131	Y	H	Decrease	7
135	G	D	Decrease	8

155	P	A	Decrease	8
165	P	L	Decrease	6
165	P	S	Decrease	7
168	I	M	Decrease	8
170	N	K	Decrease	0
170	N	S	Increase	1
189	Y	D	Decrease	7
191	C	Y	Decrease	3
192	D	Y	Decrease	1
202	L	F	Decrease	9
210	C	Y	Decrease	5
216	W	G	Decrease	10
217	S	G	Decrease	8
221	P	S	Decrease	9
231	P	R	Decrease	8
236	G	A	Decrease	7
240	S	P	Decrease	1
246	F	C	Decrease	8
248	Y	H	Decrease	6
259	G	R	Decrease	8
262	L	P	Decrease	7
276	W	S	Decrease	9
329	P	R	Decrease	8

Where, "WT" is the amino acid in native protein, "NEW" is mutant amino acid, "RI" is the reliability index.

Table 3: Evolutionary stability of amino acid positions in CD46.

Position	SEQ	SCORE (normalized)	COLOR
35	C	-1.37	9
54	Y	-1.158	8
57	G	-1.01	8
62	Y	-1.123	8
67	G	-1.219	9
68	Y	-1.015	8
94	C	-1.217	9
105	P	-0.118	5
130	G	-1.348	9
131	Y	-0.554	7
135	G	-1.3	9
155	P	-1.348	9
165	P	-1.494	9
168	I	-1.45	9
170	N	-1.288	9
189	Y	-1.514	9
191	C	-1.515	9
192	D	0.054	5
202	L	-1.353	9
210	C	-1.515	9
216	W	-1.335	9
217	S	-1.212	9
221	P	-1.495	9
231	P	-1.159	8
236	G	-1.158	8
240	S	-0.611	7
246	F	-0.776	7
248	Y	-0.974	8
259	G	-1.052	8
262	L	-0.664	7
276	W	-1.458	9

Conservation score is 1–4 for variable, 5–6 for intermediate and 7–9 for conserved. Except positions P105 and D192 other positions were predicted in highly conserved regions, hence showing more chances to alter the protein structure.

Table 4: Active site prediction in CD46.

PocID	Chain	SeqID	AA
1	B	25	ARG
1	B	36	TYR
1	B	37	ILE
1	B	38	PRO
1	B	39	PRO
1	B	40	LEU
1	B	41	ALA
1	B	43	HIS
1	B	58	ASP
1	B	59	ALA
1	B	61	TYR
1	D	75	GLN
1	D	91	PHE
1	D	92	ILE
1	D	93	CYS
1	D	94	ASN
1	D	95	GLU
1	D	95	GLU
1	D	97	TYR
1	D	98	TYR
1	D	99	LEU
1	D	103	GLU
1	D	126	VAL
1	D	149	TYR
1	D	150	LEU
1	D	152	ALA
1	D	168	LEU
1	D	170	GLY
1	D	171	GLU
1	D	172	SER
1	D	173	THR
1	D	174	ILE
1	D	175	TYR
1	D	176	CYS
1	D	177	GLY
1	D	178	ASP
1	D	179	ASN
1	D	181	VAL
1	D	183	SER
1	D	184	ARG
1	D	214	TYR
1	D	215	LYS
1	D	217	THR

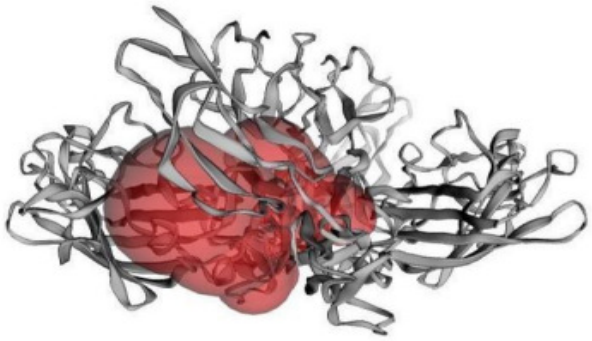

Fig. 1. The surface of the binding pocket of the protein as computed using CASTp.

CONCLUSION

As CD46 is linked to various disorders such as cancer, autoimmune, and immunodeficiencies, the current work used in silico methods to explore the influence of functional SNPs associated with the CD46 gene. 263 missense SNPs were discovered among the 10,546 SNPs in the CD46 gene. Moreover, various software methods identified 37 SNPs as harmful or detrimental. Our findings, we feel, will serve as a foundation for future experimental and computational research.

REFERENCES

[1]. Sahu, A. and Lambris, J.D. (2001). Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunological reviews*, **180**(1), 35-48.

[2]. Jarva, H., Jokiranta, T.S., Hellwage, J., Zipfel, P.F. and Meri, S. (1999). Regulation of complement activation by C-reactive protein: targeting the complement inhibitory activity of factor H by an interaction with short consensus repeat domains 7 and 8–11. *The Journal of Immunology*, **163**(7), 3957-3962.

[3]. Ruiz-Argüelles, A. and Llorente, L. (2007). The role of complement regulatory proteins (CD55 and CD59) in the pathogenesis of autoimmune hemocytopenias. *Autoimmunity reviews*, **6**(3), 155-161.

[4]. Murray, K.P., Mathure, S., Kaul, R., Khan, S., Carson, L.F., Twiggs, L.B., Martens, M.G., Kaul, A. (2000). Expression of complement regulatory proteins—CD 35, CD 46, CD 55, and CD 59—in benign and malignant endometrial tissue. Gynecologic oncology. 1; **76**(2): 176-82.

[5]. Liszewski, M.K., Post, T.W. and Atkinson, J.P. (1991). Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annual review of immunology*, **9**(1), 431-455.

[6]. Barilla-LaBarca, M.L., Liszewski, M.K., Lambris, J.D., Hourcade, D. and Atkinson, J.P. (2002). Role of membrane cofactor protein (CD46) in regulation of C4b and C3b deposited on cells. *The Journal of Immunology*, **168**(12), 6298-6304.

[7]. Naniche, D., Varior-Krishnan, G.A.Y.A.T.H.R.I., Cervoni, F., Wild, T.F., Rossi, B., Rabourdin-Combe, C.A. and Gerlier, D. (1993). Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. *Journal of virology*, **67**(10), 6025-6032.

[8]. Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R. and Amos, C.I. (2008). Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *The American Journal of Human Genetics*, *82*(1), 100-112.

[9]. Brule, C.E. and Grayhack, E.J. (2017). Synonymous codons: choose wisely for expression. *Trends in Genetics*, **33**(4), 283-297.

[10]. Gleason, K.J., Yang, F., Pierce, B.L., He, X. and Chen, L.S. (2020). Primo: integration of multiple GWAS and omics QTL summary statistics for elucidation of molecular mechanisms of trait-associated SNPs and detection of pleiotropy in complex traits. *Genome biology*, **21**(1), 1-24.

[11]. Bhattacharjee, S. and Yadava, P.K. (2018). Measles virus: background and oncolytic virotherapy. *Biochemistry and biophysics reports*, **13**, 58-62.

[12]. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research*, **38**, W529-W533.

[13]. Capriotti, E., Fariselli, P. and Casadio, R. (2005). I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, **33**, W306-W310.

[14]. Jönsen, A., Nilsson, S.C., Ahlqvist, E., Svenungsson, E., Gunnarsson, I., Eriksson, K.G., Bengtsson, A., Zickert, A., Eloranta, M.L., Truedsson, L. and Rönnblom, L. (2011). Mutations in genes encoding complement inhibitors CD46 and CFH affect the age at nephritis onset in patients with systemic lupus erythematosus. Arthritis research & therapy, **13**(6), 1-9.

[15]. Fremeaux-Bacchi, V., Moulton, E.A., Kavanagh, D., Dragon-Durey, M.A., Blouin, J., Caudy, A., Arzouk, N., Cleper, R., Francois, M., Guest, G. and Pourrat, J. (2006). Genetic and functional analyses of membrane cofactor protein (CD46) mutations in atypical hemolytic uremic syndrome. *Journal of the American Society of Nephrology*, *17*(7), 2017-2025.

[16]. Liszewski, M.K., Leung, M., Cui, W., Subramanian, V.B., Parkinson, J., Barlow, P.N., Manchester, M. and Atkinson, J.P. (2000). Dissecting sites important for complement regulatory activity in membrane cofactor protein (MCP; CD46). *Journal of Biological Chemistry*, **275**(48), 37692-37701.