ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 10, Issue 8, 2017

Online - 2455-3891
Print - 0974-2441

Research Article

IN SILICO CHARACTERIZATION AND MOLECULAR MODELLING OF SODIUM-DEPENDENT SEROTONIN TRANSPORTER PROTEIN FROM *HOMO SAPIENS*

HINA BANSAL*, NEETU JABALIA

Department of Biotechnology, Amity University, Sector 125, Noida, Uttar Pradesh, India. Email: hbansal@amity.edu

Received: 06 April 2017, Revised and Accepted: 15 May 2017

ABSTRACT

Objective: The objective of our investigation is to apply computational tools for a protein sodium-dependent serotonin transporter (SERT). It plays a role in sudden infant death syndrome, aggressive behavior in Alzheimer disease, and depression-susceptibility. Although various conventional and experimental therapies have been directed for the treatment, still it needs attention for more effective treatments. Toward this pursuit, we performed *in silico* analysis of the protein using computational tools and servers.

Methods: Homology modeling approach has been used to define the tertiary structure of the protein using SWISS-MODEL workspace. Modal validation was done to verify the generated modal. Furthermore, primary and secondary structural and functional analysis was performed to provide more perceptions into the selected protein. The protein disorder analysis was performed using PrDOS server.

Results: The results of the primary structure analyses suggested that SERT is an acidic and hydrophobic protein in nature. It is structurally stable. The secondary structural analysis results revealed that random coils dominated among secondary structure elements. The homology modeling showed that the QMEAN score of the model was −5.17, and the sequence identity was 52%. Validation protein models using Rampage revealed that more that 95.9% residues were in favored regions. The protein disorder detected by PrDOS showed the total disorder amino acid residues were 89 (14.1%).

Conclusion: The study provides valuable clues for initiation of experimental characterization of this protein and throws light on some novel insights into the structural features of sodium-dependent SERT protein from *Homo sapiens*. This will also helpful in conducting docking studies for the receptor protein against various drug molecules.

Keywords: Homology modeling, *In silico* analysis, Protein disorder analysis, Random coils.

© 2017 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4. 0/) DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10i8.18954

INTRODUCTION

Solute carrier family 6 (neurotransmitter transporter), member 4 (SLC6A4) is a protein-coding gene. A repeat length polymorphism in the promoter of this gene has been shown to affect the rate of serotonin uptake and may play a role in sudden infant death syndrome, aggressive behavior in Alzheimer disease patients, and depression-susceptibility in people experiencing emotional trauma [1,2]. All brain regions express various serotonin receptors and individual neurons may express multiple serotonin receptors [3]. In this new era, there are a huge number of computational tools have been established for making consistent predictions about the identification and structural analysis of protein [4]. The amino acid sequence can provide most of the information required for insight into protein physicochemical properties. Our investigation applied computational tools for a special class of proteins –SLC6A4. Diseases associated with SLC6A4 include - alcoholic psychosis, obsessive-compulsive disorder, Parkinson's disease, and Alzheimer disease. The serotonin transporter (SERT or 5-HTT) is a kind of monoamine transporter protein that transports serotonin from the synaptic cleft to the presynaptic neuron. The SERT plays a primary function in the central nervous system. It involves the regulation of serotonergic signaling via transport of serotonin molecules from the synaptic cleft back into the pre-synaptic terminal for re-utilization. It plays an important role in mediating regulation of the accessibility of serotonin to other receptors of serotonergic systems. Terminates the action of serotonin and recycles it in a sodium-dependent manner [5]. Our investigation applied computational tools for a special protein sodium-dependent SERT protein which is encoded by the gene SLC6A4. It plays a role in sudden infant death syndrome, aggressive behavior in Alzheimer disease, and depression-susceptibility. Here, we have modeled and characterized sodium-dependent SERT protein *in silico* using various computational tools and servers [6,7]. Since there is no experimental structure was available for the protein in the all available database. Homology modeling approach has been used to define the tertiary structure of the protein receptor using SWISS-MODEL workspace. Modal validation was done to verify the generated modal. The solvent accessibility analysis was performed using ASA – view. The primary and secondary structure analysis was performed to provide more perceptions into the selected receptor. The results revealed that random coils dominated among secondary structure elements. Further, motifs and transmembrane regions were also identified in the sequence. Phylogenetic analysis were performed for 10 orthologous sequences of sodium-dependent SERT of human using molecular evolutionary genetics analysis version 7 (MEGA7). The binding site prediction and protein disorder analysis were done using BSpred and ProDOS server.

METHODS

Sequence retrieval

The amino acid sequence of sodium-dependent SERT of human was retrieved in the FASTA format from National Center for Biotechnology Information (NCBI) having accession no. NP_001036.1. The BLAST search was also performed to find the sodium-dependent SERT protein in other organism and to see its evolutionary history.

Physicochemical characterization

The physicochemical characterization of the protein was computed using the Expasy's ProtParam server [8]. The physicochemical parameters include theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, halflife, aliphatic index, instability index, and grand average hydropathicity (GRAVY).

Functional characterization

Motifs present in the protein sequence of sodium-dependent SERT of human were scanned using Motif Search. For the identification of transmembrane regions and hydropathicity, SOSUI server was used [9]. Phylogenetic tree was constructed based on neighbor-joining method [10] using MEGA7 [11].

Secondary and tertiary structural characterization

The secondary structure prediction was done using self-optimized prediction method from alignment (SOPMA) [12] and Garnier-Osguthorpe-Robson (GOR) IV method [13]. For tertiary structural characterization, homology modeling approach was applied. Homology modeling generally begins through searching the PDB for similar protein structure using target sequence as the query [14]. The modeling of 3D structure of the sequences was executed by SWISS-MODEL workspace [15].

Assessment of 3D structure prediction

Rampage server was used for evaluating and assessing the accuracy of the model [16,17]. The structure was visualized using SWISS-PDB VIEWER to understand the insight of molecular structure [18]. Solvent accessibility of the amino acid residues in the modelled protein was determined by ASA – view [19].

Protein disorder

The disordered regions of a protein chain from its amino acid sequence were predicted by PrDOS. It returns disorder probability of each residue as prediction results [20,21].

RESULT AND DISCUSSION

The 630 amino acid long sequence of sodium-dependent SERT of human was retrieved in the FASTA format from NCBI having accession no. NP_001036.1. The BLAST search was performed against this protein and 10 orthologous sequences were selected having sequence identity >90% (Table 1).

Physiochemical characterization

The physicochemical properties (Table 2) and amino acid composition (Table 3) have been computed using the Expasy's ProtParam. Amino acid composition and physicochemical properties determine the fundamental properties of the protein. The results of primary structure analyses suggest that all of the SERT are hydrophobic in nature due to the presence of high content of nonpolar residues (Val, Ala, Leu, etc.) (Table 3). The presence of cysteine residues indicates the presence of disulfide bridges (SS bonds) in this SERT. Moreover, the primary structure analysis suggests that the SERT have high percentage of aliphatic residues such as Leu (9.7%), Ile (8.9%), Gly (7.9%), Ala (7.3%), Thr (7.3%), and Val (7.1%) and absence of Phl (0.0%), Sec (0.0%) (Table 3). The computed isoelectric point of SERT is 5.89 (Table 2). It suggested that SERT protein is acidic in nature because their calculated pI value was less than 7. Aliphatic index of SERT protein is 100.44. The aliphatic index refers to the relative volume of a protein that is occupied by aliphatic side chains. Higher aliphatic index of protein indicated their structural stability. An increase in the aliphatic index increases the thermostability of enzyme. The grand average of hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. A calculated GRAVY value of the protein is 0.422. A positive GRAVY value for protein designates it to be hydrophobic in nature.

Functional characterization

Two motifs were identified in the protein sequence of sodium-dependent SERT of human using MotifScan (Table 4). The first motif was predicted from amino acid residues from 103 to 117 and the second motif is from 186 to 206. The transmembrane regions were identified using SOSUI server (Table 5). 13 transmembrane helices of equal length were identified from the membrane protein sodium-dependent SERT of human. The average hydrophobicity calculated was 0.292996. The phylogenetic tree has been constructed for 10 orthologous sequences of sodium-dependent SERT of human by NJ method using MEGA7 (Fig. 1). The tree divides all the sequences into three major groups. The phylogenetic analysis results revealed that the closest organisms from human are *Macaca fascicularis* and *Macaca mulatta*. Their common names are cynomolgus monkey and rhesus monkeys, respectively.

Structural analysis

The secondary structure predication of the SERT protein is predicted by two online tools SOPMA and GOR IV showing that the protein had a significant percentage of random coil and extended sheets followed by moderate content of alpha helices and beta turns (Table 6). The content of extend strand predicted by the two tools matches significantly, while there was less agreement between the two tools in terms of alpha helices, beta turn and random coil (Fig. 2).

Protein disorder analysis

The protein disorder spread over the nine regions as observed in Fig. 3. The longest disordered region was spread from Meth1 to ARG79. The

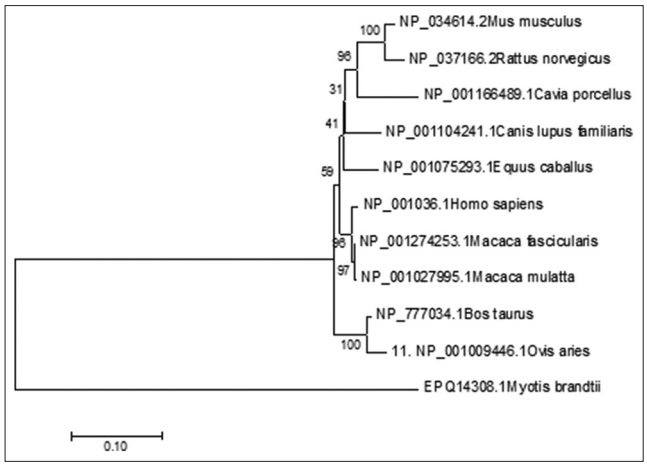

Fig. 1: Phylogenetic tree of 10 orthologous sequences of sodium-dependent serotonin transporter of human using molecular evolutionary genetics analysis software version 7

Table 1: Blast search results against the protein sequence of sodium-dependent serotonin transporter of human

S.No.	Organism name	Max. score	Query cover	E value	Percent identity	Accession no.
1.	*Homo sapiens*	1290	100	0.0	100	NP_001036.1
2.	*Macaca fascicularis*	1278	100	0.0	99	NP_001274253.1
3.	*Macaca mulatta*	1278	100	0.0	99	NP_001027995.1
4.	*Equus caballus*	1233	100	0.0	93	NP_001075293.1
5.	*Canis lupus familiaris*	1211	100	0.0	93	NP_001104241.1
6.	*Bos taurus*	1209	100	0.0	93	NP_777034.1
7.	*Mus musculus*	1204	100	0.0	93	NP_034614.2
8.	*Myotis brandtii*	1197	100	0.0	86	EPQ14308.1
9.	*Rattus norvegicus*	1194	100	0.0	92	NP_037166.2
10.	*Cavia porcellus*	1184	100	0.0	90	NP_001166489.1
11.	*Ovis aries*	1182	98	0.0	92	NP_001009446.1

protein disorder detected by PrDOS showed the total disorder amino acid residues were 89 (14.1%).

Table 2: Physicochemical properties of sodium-dependent serotonin transporter of human using Expasy's ProtParam tool

Accession number	NP_001036.1
Molecular weight	70324.8
Theoretical pI	5.89
Aliphatic index	100.44
Grand average of hydropathcity (GRAVY)	0.422
Total number of atom	9917
Extinction coefficient	151815
Total number of negatively charged residues (Asp+Glu)	45
Total number of positively charged residues (Arg+Lys)	40

Table 3: Amino acid composition (in %) of sodium-dependent serotonin transporter of human

Amino acids residue	N (%)
Ala (A)	46 (7.3)
Arg (R)	19 (3.0)
Asn (N)	22 (3.5)
Asp (D)	18 (2.9)
Cys (C)	18 (2.9)
Gln (Q)	18 (2.9)
Glu (E)	27 (4.3)
Gly (G)	50 (7.9)
His (H)	7 (1.1)
Ile (I)	56 (8.9)
Leu (L)	61 (9.7)
Lys (K)	21 (3.3)
Met (M)	12 (1.9)
Phe (F)	43 (6.8)
Pro (P)	30 (4.8)
Ser (S)	41 (6.5)
Thr (T)	46 (7.3)
Trp (W)	19 (3.0)
Tyr (Y)	31 (4.9)
Val (V)	45 (7.1)
Pyl (O)	0 (0.0)
Sec (U)	0 (0.0)

Homology modeling

Homology (or comparative) modeling methods use experimental protein structures as templates to generate models for protein or target sequences which share evolutionarily relationship [22]. Many online tools/server are accessible for comparative modeling of proteins and its has been analyzed from previous studies that a sequence similarity higher than 25% among two proteins is considerable for analogs 3D-structures [23]. Since there is no experimental structural data available for the protein sodium-dependent SERT of human; the homology modeling was carried out for predicting the 3D-structure

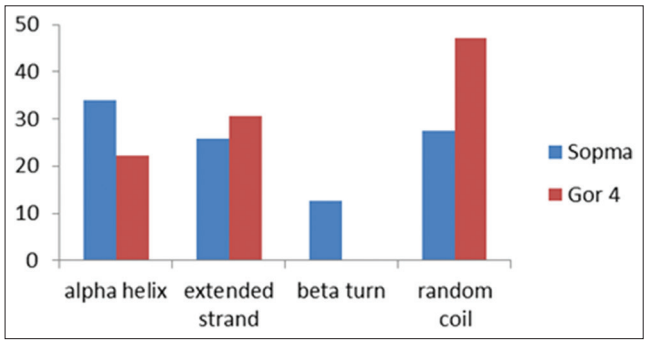

Fig. 2: Comparison of the secondary structural features of sodium-dependent serotonin transporter of human using self-optimized prediction method from alignment and Garnier-Osguthorpe-Robson 4

Fig. 3: Protein disorder analysis of sodium-dependent serotonin transporter of human using PrDOS

Table 4: Motifs identified in the protein sequence of sodium-dependent serotonin transporter of human using MotifScan

Motif information	No. of sites	Amino acid residue
NA_NEUROTRAN_SYMP_1 Sodium: neurotransmitter symport family signature 1	1	103-117
NA_NEUROTRAN_SYMP_2 Sodium: neurotransmitter symport family signature 2	1	186-206

Table 5: Transmembrane regions identified using SOSUI server

No.	N terminal	Transmembrane region	C terminal	Type	Length
1.	127	VDFLLSVIGYAVDLGNVWRFPYI	149	Secondary	23
2.	157	AFLLPYTIMAIFGGIPLFYMELA	179	Primary	23
3.	195	ICPIFKGIGYAICIIAFYIASYY	217	Primary	23
4.	219	TIMAWALYYLISSFTDQLPWTSC	241	Secondary	23
5.	289	LGGISWQLALCIMLIFTVIYFSI	311	Primary	23
6.	319	GKVVWVTATFPYIILSVLLVRGA	341	Primary	23
7.	366	VWIDAAAQIFFSLGPGFGVLLAF	388	Secondary	23
8.	404	VTSVVNCMTSFVSGFVIFTVLGY	426	Secondary	23
9.	458	MPASTFFAIIFFLMLITLGLDST	480	Primary	23
10.	508	LAVVITCFFGSLVTLTFGGAYVV	530	Primary	23
11.	537	ATGPAVLTVALIEAVAVSWFYGI	559	Primary	23
12.	576	WFWRICWVAISPLFLLFIICSFL	598	Primary	23
13.	618	LGYCIGTSSFICIPTYIAYRLII	640	Secondary	23

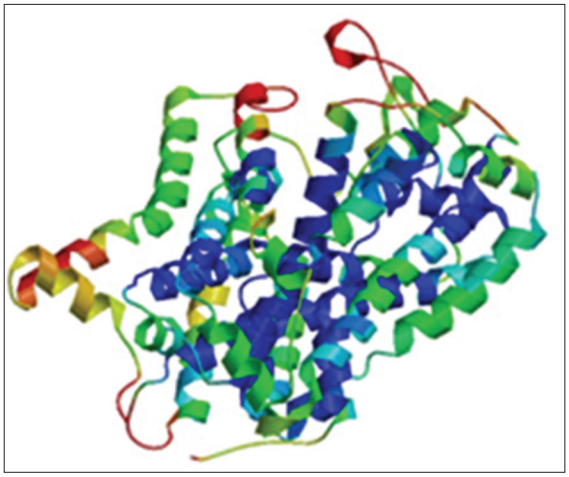

Fig. 4: 3D-model of sodium-dependent serotonin transporter of human using SWISS-MODEL workspace having a QMEAN score is −5.17 and the sequence identity is 52%

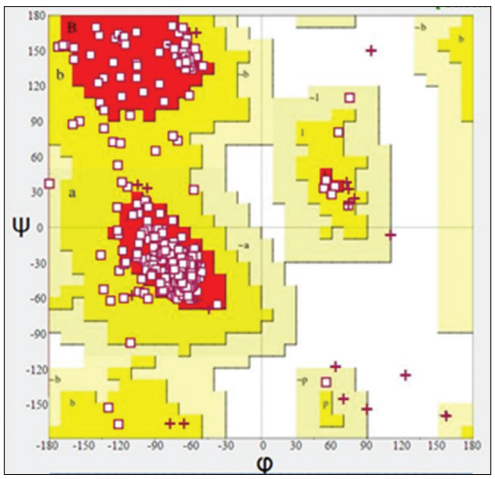

Fig. 5: Ramachandran plot of sodium-dependent serotonin transporter of human showing 98% of the residues are present in the favored region and 2% of the residues present in the allowed region

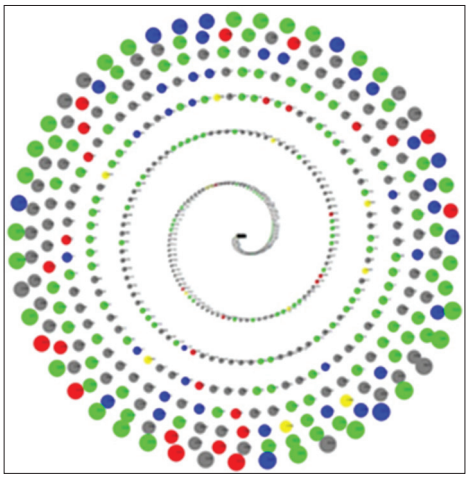

Fig. 6: Solvent accessibility plot of sodium-dependent serotonin transporter of human using ASA – View

of the protein using SWISS-MODEL Workspace (Fig. 4). The protein sequence of sodium-dependent SERT of human was retrieved from the NCBI database, and then it was used for model building. The QMEAN score of the model is −5.17, and the sequence identity is 52%.

Table 6: Secondary structure features of sodium-dependent serotonin transporter of human using SOPMA and GOR IV

Tool	N (%)			
	Alpha helix	Extended strand	Beta turn	Random coil
SOPMA	214 (33.97)	163 (25.87)	80 (12.70)	173 (27.46)
GOR IV	140 (22.22)	193 (30.63)	0 (0)	297 (47.14)

SOPMA: Self-optimized prediction method from alignment, GOR: Garnier-Osguthorpe-Robson

Table 7: 3D-structure validation using Rampage

Accession number	Rampage percentage of residues		
	Favored region	Allowed region	Outlier region
NP_001036.1	95.9	3.1	0.9

Model validation and evaluation

Quality of generated models was evaluated by Ramachandran plot analysis using RAMPAGE program. The results obtained are shown in Fig. 5. Validation of the three dimensional protein models using Rampage (Assessment of the Ramachandran Plot) in web-based server revealed that more that 95.9% residues were in favored regions (Table 7).

Solvent accessibility

The solvent accessibility of a residue was predicted by ASA – view (absolute surface area) in a protein measures the extent of burial or exposure of that residue in the 3D-structure and used to describe a protein's biophysical or evolutionary properties (Fig. 6). The blue color represents positive charged residues, red color represents negative charged residues, green color represents polar charged residue, and yellow and gray color represents other hydrophobic residues. The ASA plot pointed out that the majority of green polar charged residues were present on the outermost surface and hydrophobic residues were confined to the inner rings of spiral. However, few of red negative charged residues are also present on the outer rings of the spiral.

CONCLUSION

In this study, an attempt has been made to perform analysis of sequence and structural features of sodium-dependent SERT of human. The analyses included their physicochemical characterization, secondary structure prediction, and homology modeling. The sequence for SERT was retrieved from NCBI in fasta format. The study suggested that sodium-dependent SERT protein is acidic in nature because its calculated pI value was <7. A positive GRAVY value of the protein and high content of polar residues indicates its hydrophobic nature, which might be useful for its membrane association. The secondary structure prediction shows that the protein had a significant percentage of random coil and extended sheets followed by moderate content of alpha helices and beta turns. The tertiary structure predicted by SWISS-MODEL workspace. It is showing a QMEANE score of −5.17 and 52% sequence identity. The Ramachandran plot was generated by RAMPAGE program which is showing that 98% of the residues are present in the favored region and 2% of the residues present in the allowed region. These findings may provide valuable clues for initiating the experimental characterization of this protein.

REFERENCES

1. Masson J, Sagné C, Hamon M, El Mestikawy S. Neurotransmitter transporters in the central nervous system. Pharmacol Rev 1999;51(3):439-64.

2. Hoyer D, Hannon JP, Martin GR. Molecular, pharmacological and functional diversity of 5-HT receptors. Pharmacol Biochem Behav 2002;71(4):533-54.
3. Roth BL. Irving Page Lecture: 5-HT(2A) serotonin receptor biology: Interacting proteins, kinases and paradoxical regulation. Neuropharmacology 2011;61(3):348-54.
4. Bansal H, Narang D, Jabalia N. Computational characterization of antifreeze proteins of *Typhula ishikariensis* - Gray snow mould. J Proteins Proteomics 2014;5(4):169-76.
5. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: Integrating information about genes, proteins and diseases. Trends Genet 1997;13(4):163.
6. Jabalia N, Bansal H, Mishra PC, Chaudhary N. *In-silico* investigation of cysteine proteases from *Zingiber officinale*. J Proteins Proteomics 2015;6(3):245-53.
7. Bansal H, Kriti, Kumar M, Narad P. homology modeling and docking studies on SEMA3A as a receptor for targeting multiple myeloma. Int J Pharm Sci Rev Res 2016;36(1):50-3.
8. Thomas S, Balaji PV. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*. Protein Sci 2005;14(3):582-92.
9. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: Classification and secondary structure prediction system for membrane proteins. Bioinformatics 1998;14(4):378-9.
10. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4(4):406-25.
11. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24(8):1596-9.
12. Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995;11(6):681-4.
13. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods in enzymol 1996;266:540-53.
14. Dignya D, Vaibhav M, Manali D. *In silico* analysis of acral peeling skin syndrome: A proteomic approach Dinya Desai. Asian J Pharm Clin Res 2016;9(4):316-9.
15. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modeling. Bioinformatics 2006;22(2):195-201.
16. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, *et al.* Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. Proteins 2003;50(3):437-50.
17. Mukherjee S, Zhang Y. Protein-protein complex structure prediction by multimeric threading and template recombination. Structure 2011;19(7):955-66.
18. Dhanalakshmi R, Manavalan R. *In silico* docking approach for antiatherosclerotic activity of phytoconstituents of *Corchoris aestuans* and ADMET prediction. Asian J Pharm Clin Res 2015;8(2):350-3.
19. Ahmad S, Gromiha M, Fawareh H, Sarai A. ASAView: Database and tool for solvent accessibility representation in proteins. BMC Bioinformatics 2004;5(1):51.
20. Ishida T, Kinoshita K. PrDOS: Prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res 2007;35:W460-4.
21. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum allowed solvent accessibilites of residues in proteins. PLoS One 2013;8(11):e80635.
22. Vemulapati BM, Meghana C, Suharitha. *In silico* prediction of deleterious and non-deleterious nsSNPs in CFTR gene variants. Int J Pharm Pharm Sci 2016;8(12):303-6.
23. Medha D, Daga A, Rakesh R. Structural and functional analysis of AF9-MLL oncogeneic fusion protein using homology modeling and simulation based approach. Int J Pharm Pharm Sci 2015;7(12):155-61.