Effect of Structural Changes in Proteins Derived from *GATA4* Nonsynonymous Single Nucleotide Polymorphisms in Congenital Heart Disease

D. S. MANJEGOWDA, P. KARUNAKAR[1] AND N. B. RAMACHANDRA[2]*

NUCSER, K S Hegde Medical Academy, NITTE University, Deralakatte, Mangalore-575 018, [1]Department of Biotechnology, PES Institute of Technology, BSK III Stage, Bengaluru-560 085, [2]Department of Studies in Zoology, Genomics Laboratory, University of Mysore, Manasagangotri, Mysore-570 006, India

Manjegowda, *et al.*: Role of nsSNPs of GATA4 in Manifestation of CHD

Congenital heart disease is the most common type of birth defect. The single nucleotide polymorphism in *GATA4* is associated with various congenital heart disease phenotypes. In the present study, we analysed the nonsynonymous single nucleotide polymorphism of *GATA4*, which are involved in congenital heart disease by predicting the changes in protein structures. Total of 49 nonsynonymous single nucleotide polymorphisms of *GATA4* was screened from congenital heart disease patients of Mysore and also globally reported nonsynonymous single nucleotide polymorphisms. To understand the role of nonsynonymous single nucleotide polymorphisms, we mutated the sequence and translated into amino acids. Further the mutated protein secondary structure is predicted and tertiary structure is predicted using homology modeling. The quantitative evaluation of protein structure quality was verified with Volume Area Dihedral Angle Reporter server. Results revealed the secondary, tertiary structural changes along with changes in free energy of folding, volume and accessible surface area. Thus, the structural changes in the mutated proteins impaired the normal function of *GATA4*.

Key words: Congenital heart disease, single nucleotide polymorphism, *GATA4*, homology modeling, free energy folding

Congenital heart disease (CHD), affects 1% of all live births and is the leading noninfectious cause of death in the 1st year of life[1]. With the progress in molecular genetics and developmental biology, genes associated with heart disease have been identified[2]. It is found that single nucleotide polymorphisms (SNPs) in a variety of single genes are associated with congenital heart defects and genetic syndromes[2]. Data from mouse and human demonstrated that the zinc finger containing transcription factor (GATA4) plays a critical role in heart disease development[3]. *GATA4* interacts directly with *NKX2.5* via the homeodomain of *NKX2.5* and zinc finger domain of *GATA4*. Further, regulates genes involved in embryogenesis and in myocardial differentiation and function[3]. Mutations in *GATA4* has been identified in multiplex families like septal defects, such as atrial septal defect (ASD), and ventricular septal defect (VSD)[4].

Molecular genetic testing of *GATA4* mutations is more widespread and applied clinically. These testing are the basis for improved diagnosis and therapy of human CHD. To discriminate between synonymous SNPs (sSNPs), which constitute the majority of genetic variation and nonsynonymous SNPs (nsSNPs) located in coding regions, an analysis of DNA and protein sequence would result in amino acid variation in the protein products of genes[5].

Earlier studies from our laboratory revealed the prevalence of CHDs ranging from 6.6 to 13.06 per

*Address for correspondence
E-mail: nallurbr@gmail.com

1000 live births in India[6]. Among the different CHDs and VSDs were the most common phenotype observed with high rate of consanguinity in the patient families[7]. Studies from our laboratory also reported numerical and structural chromosomal anomalies[8]. However, molecular studies on CHDs are limited. In light of this, we have screened SNPs of *GATA4* in 100 CHD patients and 50 controls in south Indian population and found 11 nsSNPs namely c.1129A>G (S377G), c.1180C>A (P394T), c.1081A>G (M361V), c.1138G>A (V380M), c.1273G>A (D425N), c.716A>G (N239S), c.278G>C (G93A), c.1207C>A (L403M), c.1232C>T (A411V), c.1295T>C (L432S) and novel SNP, c.1180C>G (P394A)[9].

Further, *in silico* analysis of these SNPs is very limited in exploring the molecular mechanisms of CHD. In view of this, an attempt to analyse the nsSNPs of *GATA4*, which are involved in CHD was made using *in silico* tools to predict structural changes in the protein as well as quantitative and qualitative evaluation of the protein.

MATERIALS AND METHODS

The present study was conducted at Genomics Laboratory, Department of Studies in Zoology, University of Mysore, Karnataka, India between September 2007 and March 2013.

Secondary structure analysis:

Forty nine nsSNPs of *GATA4* cited in literature was used in this study, of which 11 nsSNPs reported from our laboratory is also included[4,10-21]. All 49 sequences were mutated *in silico* to generate nsSNPs. These sequences were translated to identify amino acid polymorphisms. The differences in protein sequences were compared through multiple sequence alignment. The secondary structures for all the 49 nsSNPs have been predicted by Chou Fasman secondary structure algorithm by using Accelrys gene software[22].

Comparative modeling of *GATA4* nsSNPs:

Mutational effect on protein structure was determined by using http://swissmodel.expasy.org/workspace[23]. Template structures N-terminal zinc finger of murine GATA-1 protein (PDB Id: 1GNF_A) between 213-251 residues and opposite GATA DNA binding protein (PDB Id: 3DFX_B) between 265-320 residues was used to generate 3-D structures[24]. The 19 nsSNPs of these regions were analysed. The root-mean-square

deviation (RMSD) calculation for all the modeled structure to the template found the difference from 0.01 to 0.015[23].

Evaluation and validation of three-dimensional structure:

Evaluation and validation of the 3-D structure was done using VADAR server (Alberta, Canada)[25]. The overall quality of the modeled protein was assessed by Ramachandran plot[25]. The structure superimposition of both control and mutant proteins was performed using Pymol[26].

RESULTS

Only 16 nsSNPs of *GATA4* showed positive in their secondary structure change (Table 1). These changes with reference to control are represented in fig. 1. The proteins derived from nsSNPs of *GATA4*, c.755G>C, c.1037C>T, c.1129A>G and c.1130G>A showed changes in their turn, c.155C>T, c.487C>T and c.1220C>A showed changes in their sheet and mutation c.278G>C showed changes only in helix. The nsSNP of *GATA4*, c.687G>T was responsible to change either into sheet or helix. The *GATA4* nsSNPs, c.779G>A and c.855T>C may change to helix or turn. The changes in helix or sheet was observed at c17C>T, c.82C>T, c.905A>G and c.1207C>A. Mutation at c.1295T>C showed changes in their helix, sheet or turn.

Two X-ray crystallographic structures showed homology to GATA4 protein. The N-terminal Zinc finger protein of GATA4 had homology with N-terminal zinc finger of murine GATA-1 protein (PDB Id: 1GNF_A) and with opposite GATA DNA binding protein (PDB Id: 3dfxB). For validation of modeling Ramachandran plot was applied using VADAR server (Alberta, Canada) (fig. 2).

Tertiary structural changes are showed in figs. 3 and 4. Since the tertiary structural are modeled using Homology modeling only side chain differences compared to control protein were observed (figs. 3 and 4). We observed differences in the side chain atoms such as, Delta carbon (CD), epsilon nitrogen (NE), zeta carbon (CZ), gamma carbon (CG), gamma sulphur (SG), gamma oxygen (OG1), delta sulphur (SD). The Quantitative and qualitative analysis of GATA4-mutated proteins showed, differences in free energy of folding, volume and

TABLE 1: NONSYNONYMOUS SNPS OF *GATA4* AND THEIR IMPACT ON SECONDARY STRUCTURE OF PROTEIN

SNP	Amino acid change	Secondary structure of the protein		
		Helix	Sheet	Turn
c.17C>T	A6V	+	+	–
c.82C>T	H28Y	+	+	–
c.155C>T	S52F	–	+	–
c.278G>C	G93A	+	–	–
c.454G>C	A152P	–	–	–
c.487C>T	P163S	–	+	–
c.622T>C	F208L	–	–	–
c.631T>C	F211L	–	–	–
c.640G>A	G214S	–	–	–
c.648C>G	E216D	–	–	–
c.668T>C	M223T	–	–	–
c.687G>T	R229S	–	+	+
c.700G>A	G234S	–	–	–
c.715A>G	N239D	–	–	–
c.716A>G	N239S	–	–	–
c.731A>G	Y244C	–	–	–
c.743A>G	N248S	–	–	–
c.755G>C	R252P	–	–	+
c.764T>C	I255T	–	–	–
c.779G>A	R260Q	+	–	+
c.782T>C	L261P	–	–	–
c.796C>T	R266X	–	–	–
c.818A>G	N273S	–	–	–
c.830C>T	T277I	–	–	–
c.848G>A	R283H	–	–	–
c.855T>C	N285K	+	–	+
c.874T>C	C292R	–	–	–
c.881C>T	A294V	–	–	–
c.886G>T	G296C	–	–	–
c.886G>A	G296S	–	–	–
c.905A>G	H302R	+	+	–
c.946C>G	Q316E	–	–	–
c.1037C>T	A346V	–	–	+
c.1075G>A	E359K	–	–	–
c.1081A>G	M361V	–	–	–
c.1129A>G	S377G	–	–	+
c.1130G>A	S377N	–	–	+
c.1138G>A	V380M	–	–	–
c.1180C>A	P394T	–	–	–
c.1207C>A	L403M	+	+	–
c.1220C>A	P407Q	–	+	–
c.1232C>T	A411V	–	–	–
c.1273G>A	D425N	–	–	–
c.1286G>C	S429T	–	–	–
c.1288C>G	L430V	–	–	–
c.1295T>C	L432S	+	+	+
c.1306C>T	H436Y	–	–	–
c.1324G>A	A442T	–	–	–
c.1325C>T	A442V	–	–	–

+: The secondary structure changes of the *GATA4* protein

accessible surface area (ASA) with control protein (Table 2).

DISCUSSION

Human SNPs represent the most frequent type of DNA variation. The main goals of SNP research is to understand the genetics of the human phenotype variation and especially the genetic basis of human complex diseases[5]. The nsSNPs comprise a group of SNPs that together with SNPs in regulatory regions are believed to have the highest impact on phenotype[27]. The nsSNPs also known as single amino acid polymorphism (SAPs) that causes amino acid changes in proteins, which have the potential to affect both protein structure and function[28]. Some of the mutations in SAP sites are not associated with any changes in phenotype and are considered functionally neutral, but others bringing deleterious effects to protein function and are responsible for many human genetic diseases[28]. By the analysis of the new incoming data on SNPs by mapping them at sequence and structural would address problems concerning population, medical and evolutionary genetics[29].

Reamon-Buettner and Borlak[11] reported that, zinc finger transcription factor, GATA4 is a master regulator of heart development. Zinc finger mutation identified in this gene affects a zinc coordinating cysteine in the C-terminal finger and is strongly associated with VSDs. *GATA4* zinc finger mutations are shown to affect DNA binding, contacts on zinc ion and protein secondary structure. The impaired *GATA4* interaction with the third helix of the homeodomain of *NKX2.5* results in septation defects of the human heart[11].

The structural changes in the analysed GATA4 mutated protein might be responsible for non functional GATA4 proteins. However, we found only side chain differences of the amino acids at the tertiary structure level of protein. Theoretical accuracy of prediction of the tertiary structure of a protein from a sequence is 90%[30]. Local conformation of a protein varies under the native conditions. These limitations are also imposed by secondary structure prediction's inability to account for tertiary structure. The only one amino acid difference occurs by particular nsSNP
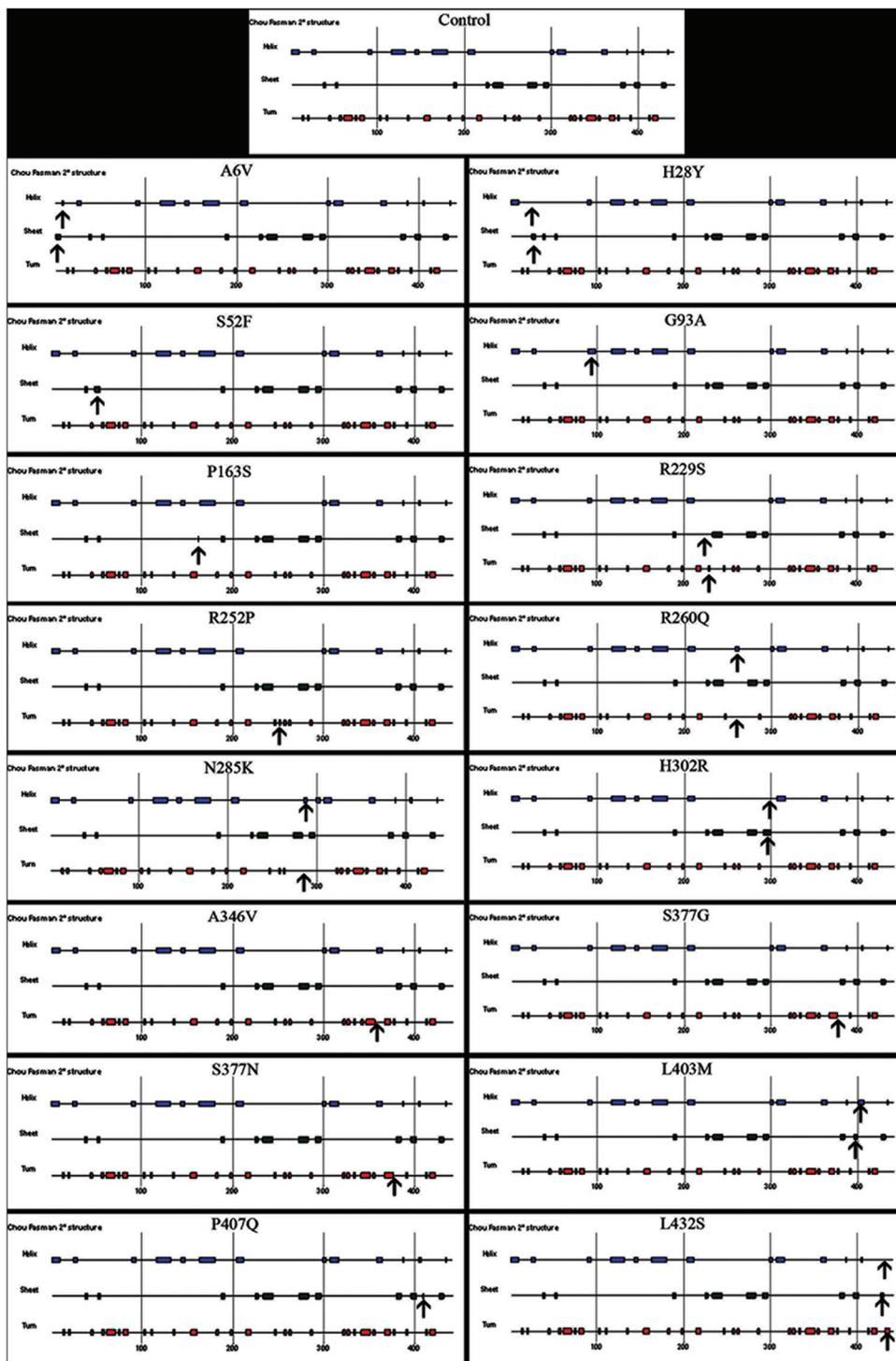
Fig. 1: Secondary structure of GATA4 mutated proteins.
Secondary structure of GATA4 mutated proteins derived from 16 nsSNPs shows structural changes in their helix (blue), sheet (green) and turn (red) compare to normal protein of GATA4.

and modeling tool considers the same template for all nsSNPs. Hence, all models have same structure except side chain difference for the mutated.

Side chain of an amino acid is specific to each amino acid of a protein. The side chain can make

an amino acid as a weak acid or a weak base, and a hydrophile if the side chain is polar or a hydrophobe if it is nonpolar[31]. The distribution of hydrophilic and hydrophobic amino acids determines the tertiary structure of the protein. There physical location of proteins influences their quaternary structure[31]. These
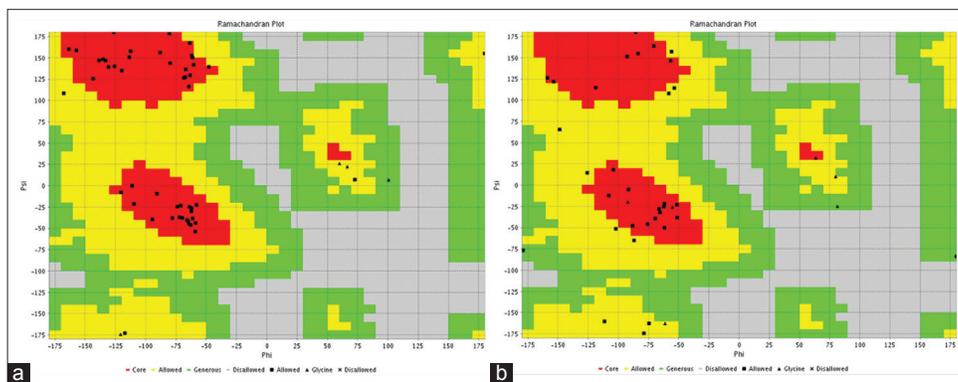
Fig. 2: Ramachandran plot.
Ramachandran plot was analysed using Volume Area Dihedral Angle Reporter server (Alberta, Canada) for the evaluation and validation of 3-D reference structures: (a) n-terminal zinc finger of murine GATA-1 protein (PDB Id: 1gnfA): fully allowed region (41 residues, 78.85%), additionally allowed region (10 residues, 19.23%), outside region (1 residue GLY24, 1.92%) (b) opposite GATA DNA binding protein (PDB Id: 3dfxB): Fully allowed region (20 residues, 4.05%), additionally allowed region (14 residues, 37.84%), generously allowed region (2 residues GLY2, GLY37, 5.41%), outside region (1 residue ASN27, 2.70%).
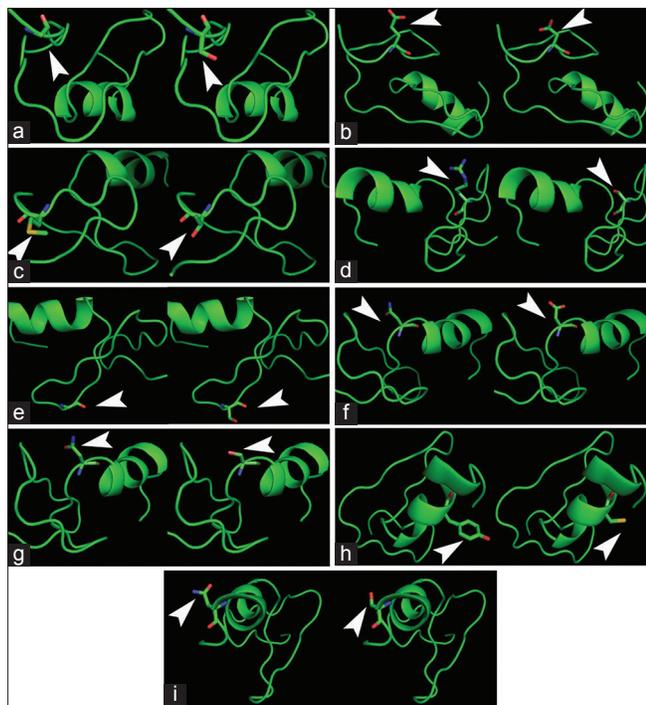
Fig. 3: 3-D models of mutated proteins derived from 9 nsSNPs of zinc finger region of GATA4.
Each model was superimposed by control (left side of each image) and mutated protein (right side of each image). Models were done using solution structure of the n-terminal zinc finger of murine GATA-1 (PDB Id: 1gnfA) and also for mutant. Ball and stick representation in both the models showed the side chain structural differences between control and mutated proteins. (a) E216D: In control- Delta Carbon (CD), Epsilon Oxygen (OE1, OE2) and in mutant OD1, OD2. (b) G214S: In control-no changes and in mutant Gamma Carbon (CG), Gamma Oxygen (OG). (c) M223T: In control- CB, CG, Delta Sulphur (SD), CE and in mutant CG2, OG1. (d) R229S: In control- CG, CD, Epsilon Nitrogen (NE), CE2, Zeta Carbon (CZ). NH1, NH2 and in mutant OG. (e) G234S: In control- no changes and in mutant CB, OG. (f) N239D: In control-ND2, OD1 and in mutant OD1, OD2. (g) N239S: In control- CG, ND2, OD1 and in mutant OG. (h) Y244C: In control- CG, CD1, CD2, CE1, CZ, OH and in mutant Gamma Sulphur (SG). (i) N248S: In control- CG, ND2, OD1and in mutant OG.
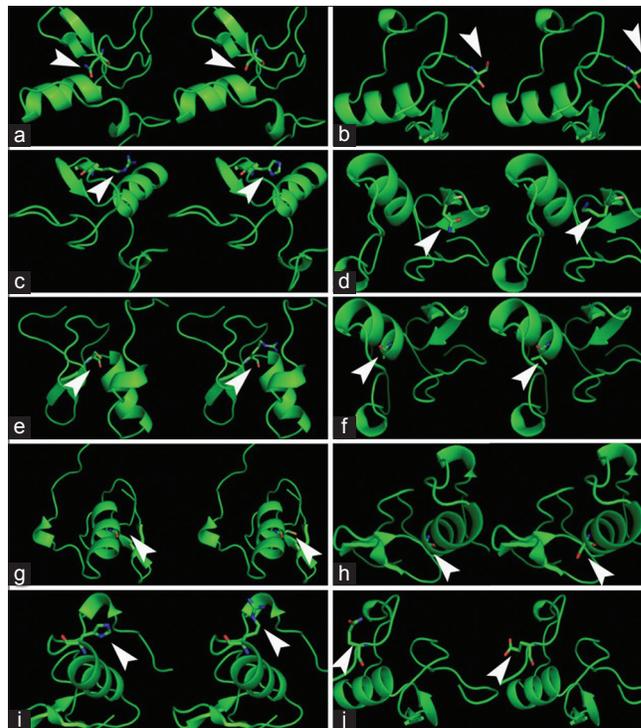
Fig. 4: 3-D models of mutated proteins derived from 10 nsNPs of GATA4.
Each model was superimposed by control (left side of each image) and mutated protein (right side of each image). Models were done using solution structure of the opposite GATA DNA binding protein (PDB Id: 3dfxB) and also for mutant. Ball and stick representation in both the models showed the side chain structural differences between control and mutated proteins. (a) T277I: In control-CG2, OG1and in mutant CG1, CG2, CD1. (b) R283H: In control- CD, NE, CZ, NH1, NH2 and in mutant CD2, ND1, CE1, NE2. (c) Q316E: In control-NE2, OE1 and in mutant OE1, OE2. (d) N285K: In control-ND2, OD1 and in mutant CD, CE, NZ. (e) N273S: In control-CG, ND2, OD1 and in mutant OG. (f) C292R: In control-SG and in mutant CG, CD, NE, CZ, NH1, NH2. (g) A294V: In control- no changes and in mutant CG1, CG2. (h) G296C: In control-no changes and in mutant CB, SG (i) G296S: In control-no changes and in mutant CB, OG. (j) H302R: In control-CD2, ND1, CE1, NE2 and in mutant CD, NE, CZ, NH1, NH2.

TABLE 2: SUMMARY OF FREE ENERGY OF FOLDING AND VOLUME AND ACCESSIBLE SURFACE AREA

SAP	ASA in Angs 2 Control (213-251)= 66.6	Volume of the protein (Angs 3) Control (213-251)= 4618.1	Free energy of folding (kcal/mol) Control (213-251)= −28.00
E216D	69.2	4598.5	−27.71
G214S	49.6	4644.4	−27.93
G234S	63.3	4644.9	−27.82
M223T	51.4	4577.4	−27.02
N239D	90.2	4619.6	−27.93
N239S	59.9	4587.6	−27.47
N248S	33.8	4602.6	−27.81
R229S	70.3	4551.6	−27.81
Y244C	42.8	4504.0	−28.00
	ASA in Angs 2 Control (265-320)= 77.5	Volume of the protein (Angs 3) Control (265-320)= 6448.4	Free energy of folding (kcal/mol) Control (265-320)= −32.15
T277I	79.6	6459.3	−32.64
R283H	48.0	6418.2	−33.19
Q316E	117.4	6455.0	−31.32
N285K	57.5	6513.6	−33.31
N273S	23.2	6430.3	−32.09
H302R	113.4	6469.1	−32.35
G296S	17.8	6472.3	−32.32
C292R	27.4	6985.4	−29.98
G296C	23.1	6485.1	−32.82
A294V	67.4 (42.9)	7008.4	−32.32

Differences in free energy of folding, volume and ASA with control protein. ASA of control is mentioned in parenthesis. ASA: Accessible surface area, SAP: single amino acid polymorphism

properties are important in protein structure and protein–protein interactions[31].

Accessible surface area (ASA) is the exposed surface area of the protein (or residue) that a water molecule could access or touch. ASA of the side chain values are also calculated for polar (N, O, S) atoms, charged atoms (N, O) and for nonpolar atoms (C) to permit the calculation of polar, charged and nonpolar surface area. These ASA values can be quite useful in structure assessment and thermodynamic calculations. ASA is highly dependent on the choice of atomic or van der Waals radii. Protein structures are stabilized by hydrophobic and van der Waals forces, and by hydrogen bonds. Hydrophobic energy is gained by the reduction of surface in contact with water[32]. In the present study, we observed differences in free energy of folding and volume of the protein. There is a linear relationship between the solvation free energy of folding and the protein size and misfolded structures showed higher solvation free energies[33].

Usually nsSNP do not inactivate protein functionality completely, instead, nsSNPs change the protein activity at some level, either directly or indirectly through interactions with other proteins in the pathway. Such information has to be considered mutually[34]. Therefore, side chain differences of the amino acids in the 3-D structure of GATA4 mutated protein may be responsible for nonfunctioning of GATA4 proteins. This study will facilitate to further study structural changes and distinguish the CHD-causing nsSNPs from neutral SNPs. This information will explore further for probing and utilization in pharmacogenetics study and also in biomedical applications.

Conflicts of interest:
There are no conflicts of interest.

REFERENCES

1. Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol 2002;39:1890-900.
2. Huang JB, Liu YL, Sun PW, Lv XD, Du M, Fan XM. Molecular mechanisms of congenital heart disease. Cardiovasc Pathol 2010;19:e183-93.
3. Sander TL, Klinkner DB, Tomita-Mitchell A, Mitchell ME. Molecular and cellular basis of congenital heart disease. Pediatr Clin North Am 2006;53:989-1009.
4. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 2003;424:443-7.
5. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: Server and survey. Nucleic Acids Res 2002;30:3894-900.
6. Smitha R, Karat SC, Narayanappa D, Krishnamurthy B, Prasanth SN, Ramachandra NB. Prevalence of congenital heart diseases in Mysore. Int J Hum Genet 2006;12:12-7.
7. Ramegowda S, Ramachandra NB. Parental consanguinity increases congenital heart diseases in South India. Ann Hum Biol 2006;33:519-28.

8. Smitha R, Harshavardhan MG, Hyderi A, Savitha MR, Krishnamurthy B, Karat SC, *et al.* Chromosomal anomalies and congenital heart disease in Mysore, South India. Int J Hum Genet 2010;10:131-9.

9. Dinesh SM, Lingaiah K, Savitha MR, Krishnamurthy B, Narayanappa D, Ramachandra NB. GATA4 specific nonsynonymous single-nucleotide polymorphisms in congenital heart disease patients of Mysore, India. Genet Test Mol Biomarkers 2011;15:715-20.

10. Hirayama-Yamada K, Kamisago M, Akimoto K, Aotsuka H, Nakamura Y, Tomita H, *et al.* Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. Am J Med Genet A 2005;135:47-52.

11. Reamon-Buettner SM, Borlak J. GATA4 zinc finger mutations as a molecular rationale for septation defects of the human heart. J Med Genet 2005;42:e32.

12. Sarkozy A, Conti E, Neri C, D'Agostino R, Digilio MC, Esposito G, *et al.* Spectrum of atrial septal defects associated with mutations of NKX2.5 and GATA4 transcription factors. J Med Genet 2005;42:e16.

13. Nemer G, Fadlalah F, Usta J, Nemer M, Dbaibo G, Obeid M, *et al.* A novel mutation in the GATA4 gene in patients with tetralogy of fallot. Hum Mutat 2006;27:293-4.

14. Zhang L, Tümer Z, Jacobsen JR, Andersen PS, Tommerup N, Larsen LA. Screening of 99 Danish patients with congenital heart disease for GATA4 mutations. Genet Test 2006;10:277-80.

15. Rajagopal SK, Ma Q, Obler D, Shen J, Manichaikul A, Tomita-Mitchell A, *et al.* Spectrum of heart disease associated with murine and human GATA4 mutation. J Mol Cell Cardiol 2007;43:677-85.

16. Reamon-Buettner SM, Cho SH, Borlak J. Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). BMC Med Genet 2007;8:38.

17. Schluterman MK, Krysiak AE, Kathiriya IS, Abate N, Chandalia M, Srivastava D, *et al.* Screening and biochemical analysis of GATA4 sequence variations identified in patients with congenital heart disease. Am J Med Genet A 2007;143A:817-23.

18. Posch MG, Perrot A, Schmitt K, Mittelhaus S, Esenwein EM, Stiller B, *et al.* Mutations in GATA4, NKX2.5, CRELD1, and BMP4 are infrequently found in patients with congenital cardiac septal defects. Am J Med Genet A 2008;146A:251-3.

19. Zhang W, Li X, Shen A, Jiao W, Guan X, Li Z. GATA4 mutations in 486 Chinese patients with congenital heart disease. Eur J Med Genet 2008;51:527-35.

20. Hamanoue H, Rahayuningsih SE, Hirahara Y, Itoh J, Yokoyama U, Mizuguchi T, *et al.* Genetic screening of 104 patients with congenitally malformed hearts revealed a fresh mutation of GATA4 in those with

21. Zhang WM, Li XF, Ma ZY, Zhang J, Zhou SH, Li T, *et al.* GATA4 and NKX2.5 gene analysis in Chinese Uygur patients with congenital heart disease. Chin Med J (Engl) 2009;122:416-9.

22. Al-Ali H, Khachfe HM. The N-terminal domain of apolipoprotein B-100: Structural characterization by homology modeling. BMC Biochem 2007;8:12.

23. Kopp J, Schwede T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. Nucleic Acids Res 2004;32:D230-4.

24. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 1997;18:2714-23.

25. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, *et al.* VADAR: A web server for quantitative evaluation of protein structure quality. Nucleic Acids Res 2003;31:3316-9.

26. Bramucci E, Paiardini A, Bossa F, Pascarella S. PyMod: Sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. BMC Bioinformatics 2012;13 Suppl 4:S2.

27. Våge J, Lingaas F. Single nucleotide polymorphisms (SNPs) in coding regions of canine dopamine- and serotonin-related genes. BMC Genet 2008;9:10.

28. Hu J, Yan C. Identification of deleterious non-synonymous single nucleotide polymorphisms using sequence-derived information. BMC Bioinformatics 2008;9:297.

29. Sunyaev S, Lathe W 3rd, Bork P. Integration of genome data and protein structures: Prediction of protein folds, protein interactions and "molecular phenotypes" of single nucleotide polymorphisms. Curr Opin Struct Biol 2001;11:125-30.

30. Saunders R, Deane CM. Protein structure prediction begins well but ends badly. Proteins 2010;78:1282-90.

31. Creighton TE. Proteins: Structures and Molecular Properties. Ch. 1. San Francisco: WH Freeman; 1993.

32. Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. Nature 1987;328:834-6.

33. Chiche L, Gregoret LM, Cohen FE, Kollman PA. Protein model structure evaluation using the solvation free energy of folding. Proc Natl Acad Sci U S A 1990;87:3240-3.

34. Uzun A, Leslin CM, Abyzov A, Ilyin V. Structure SNP (StSNP): A web server for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways. Nucleic Acids Res 2007;35:W384-92.