Pakistan
Journal of Biological Sciences

Structural and Functional Analysis of *KIT* Gene Encoding Receptor Tyrosine Kinase and its Interaction with Sunitinib and HDAC Inhibitors: An *in silico* Approach

R. Vanajothi, S. Rajamanikandan, A. Sudha and P. Srinivasan

Department of Bioinformatics, Alagappa University, Karaikudi-630 003, Tamil Nadu, India

Abstract: *KIT* is a growth factor receptor, important for normal germ cell migration and development. The malfunction of *KIT* gene results in constitutive activation of the tyrosine kinase activity of c-*KIT* which is believed to be the major oncogenic event in stomach, small intestine mastocytosis, acute leukemias, melanomas and colon tumors. The genetics of these diseases could be better understood by knowing the functional relevance of their SNP variation. In this study, a computational analysis to detect the most deleterious non-synonymous SNPs of *KIT* gene was performed and investigated its binding affinity to native and predicted mutant protein structure (D816V) with sunitinib and HDAC (Trichostatin A and Panobinostat) inhibitors was investigated. Out of 1,288 SNPs retrieved from dbSNP database against *KIT* gene, 11 non-synonymous SNPs were detected to be damaging and deleterious by SIFT, PolyPhen and I-Mutant2.0 servers. Further, we modeled the mutant protein based on the deleterious nsSNP (rs121913507) and showed that the mutation from Aspartic acid to Valine at 816 position exhibit greatest impact on stability. The RMSD values of mutant and native structures are found to be 0.40 and 1.9 Å, respectively. Furthermore, the binding affinity of sunitinib and HDAC inhibitors were compared with native and mutant protein. In this regard, it was found that trichostatin A has a high binding efficacy towards the mutant protein with a binding energy of -35.274 kcal mol^{-1}, as compared to the native structure which has a binding energy of -25.996 kcal mol^{-1}. Also, the FastSNP tool suggested that 3 SNPs found to affect protein splicing site and splicing regulation. From present results, it was clear that the non-synonymous SNP rs121913507 (D816V) could be the most deleterious SNP for *KIT* gene and HDAC inhibitors can serve as a best drug for the mutant protein.

Key words: *KIT* gene, SNP, SIFT, PolyPhen, I-mutant, HDAC inhibitors, sunitinib

INTRODUCTION

Gastrointestinal stromal tumors (GISTs) are usually occurred in the stomach (60%) or small intestine (30%) and sometimes it may takes place anywhere in the GI tract and also occurred as extra gastrointestinal stromal tumors (EGISTs) and commonly exhibit oncogenic activating mutations in the *KIT* tyrosine kinase (Rubin *et al.*, 2001; Lux *et al.*, 2000; Hornick and Fletcher, 2007). *KIT* gene encodes the Receptor Tyrosine Kinase (RTK) and belongs to type III receptor family which includes Platelet Derived Growth Factor (PDGF) α, β and macrophage colony stimulating factor (mCSF) and it regulates haematopoiesis, melanogenesis, gut and germ cell development (Lee, 1998; Ali and Ali, 2007). These type-III RTKs includes five extracellular IgG-like loops, encodes the ligand-binding and dimerization domains and also consist of an intracellular portion which is divided into juxtamembrane (JM) domain and a split kinase domain (Majumder *et al.*, 1998). These kinases gets activated by its ligand Stem Cell Factor (SCF) under certain physiological conditions and leads to receptor dimerization, auto-phosphorylation, phosphorylation of downstream signaling cascades like P13/AKT/mTOR, RAS/RAF/MAPK pathways and also JAK/STAT signaling (Nocka *et al.*, 1990; Mol *et al.*, 2004). However, all oncogenic *KIT* mutations cause a constitutive, ligand independent activation of the kinase thereby driving tumor genesis.

KIT mutations in GIST are detected in *KIT* exon 11 followed by mutations of exons 9, 13 and 17 (Duensing *et al.*, 2004). *KIT* expression plays a critical role in diverse range of human solid tumors like mast cell tumors, germ cell tumors, ovarian carcinomas, malignant melanomas, GISTs, Small Cell Lung Cancer (SCLC), neuroblastoma and breast carcinoma (Gonzalez *et al.*, 2004). The most common mutation of c-*KIT* is a substitution of the aspartic acid residue in the position of 816 to a valine (D816V) which leads to an essential stimulation of the receptor (Sun *et al.*, 2009). Different types of *KIT* mutations have been reported in the c-*KIT* JM domain in correlation to GISTs. These are intragenic

---

deletions (Hirota *et al.*, 1998; Lasota *et al.*, 1999); insertion mutations and missense mutations without affecting the reading frames. Myogenic and neurogenic tumors are also formed by the mutation of c-*KIT* JM domain (Yasuoka *et al.*, 2003).

SNPs exhibit considerable interest for tracing the genetics of complex human diseases. SNPs are mostly bi-allelic in practice and mostly occur outside the coding regions of the genome. SNPs found within a coding region are of particular interest of scientist as they play major role in the alteration of biological functions of the proteins (Javed and Mukesh, 2010). The coding regions of the human genome include about 500,000 SNPs (Collins *et al.*, 1998). The occurrence of SNP in the genome is at random and they are reported to occur in a given population consistently. Hence, these are used as ideal biomarkers (Khan and Jamil, 2008). Among them, the non-synonymous SNPs (nsSPNs) were broadly divided into two different categories on the basis of their phenotypic effects. The nsSNPs that cause deleterious effect on protein function was believed to be disease associated and others are considered as functionally neutral (Fredman *et al.*, 2002). Hence, they are the important source of inter individual human variation, including disease susceptibility and drug sensitivity. The protein function gets altered by many nsSNPs which in turn causes change in the stability of native structure and/or its binding properties. Due to this reason, several studies have been tried over to discover the functional impact of such nsSNPs by mapping onto protein structures and to distinguish them into disease-associated and neutral nsSNPs (Stitziel *et al.*, 2003; Wang and Moult, 2001). Protein stability and the efficiency of protein interactions get altered by many nsSNPs which results in pathological condition. Yet, there is no biological evidence for such nsSNPs involvement in various conditions (Nuchnoi *et al.*, 2011). Therefore, a close investigation of the polymorphism is needed prior to the conduction of phenotype/genotype association study by *in silico* analysis. Moreover, the biological explanation of the diseases associated gene could be provided using computational biology.

MATERIALS AND METHODS

Datasets: National Centre for Biotechnology Information (NCBI) has established the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) to serve as central repository for molecular variation (Sherry *et al.*, 1999). The SNPs associated with the human *KIT* gene were collected from this database and were used for an *in silico* analysis.

Analysis of functional consequences of coding nsSNPs by SIFT method: Identifying substitutions that affect protein function is the major interest of those studying proteins and their implication in disease. Disease-causing mutations are found in polymorphism sites which are likely to be functionally and structurally important. We used the program SIFT available at (http://blocks.fhcrc.org/SIFT. html) (Ng and Henikoff, 2003). SIFT is a sequence homology based tool that was used to detect the deleterious nsSNPs in *KIT* gene. Predicting substitutions as deleterious or neutral may help to identify disease associated alleles. It also predicts the conservation level of a particular position in a protein. SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT works in a multistep procedure which are listed as fellows: (a) searches for similar sequences, (b) chooses closely related sequences that may share similar function, (c) obtains the multiple alignment of these chosen sequences and (d) calculated normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious; those greater than or equal to the cutoff values are predicted to be tolerated. The cutoff values ranges from 0.00 to 0.05 was predicted to be intolerance/deleterious. The positions with normalized probabilities <0.05 are predicted to be deleterious; those ≥ 0.05 are predicted to be tolerated.

Simulation for functional change in coding nsSNPs by PolyPhen: Analyzing the damaged coding non-synonymous SNPs at the structural level is considered to be very important to understand the functional activity of the protein (Rajasekaran *et al.*, 2008a). We used the program PolyPhen available at http://cott.embl.de/PolyPhen. Polymorphism Phenotyping (PolyPhen) was used to predict and analysis the effect on protein structure and function. PolyPhen server gets input either as amino acid sequence of a protein or the SWALL database ID or accession number together with sequence position and two amino acid variants characterizing the polymorphism. Sequence based homologous sequences and mapping of substitution site to a known protein's 3-dimensional structure are the parameters taken into account by the PolyPhen server. It calculates Position-Specific Independent Counts (PSIC) scores for each of the two variants and then computes the difference between the PSIC scores. The higher the PSIC score difference will have higher the functional impact for the particular amino acid substitution (Ramensky *et al.*, 2002).

Predicting stability change on mutation by I-Mutant2.0:
Knowledge on protein structure-function relationships can aid in the prediction of whether an nsSNP will alter protein function and therefore influence disease (Cmarik, 2008). However, most of the deleterious non-synonymous SNP associated proteins show less stability. We used the program, I-Mutant2.0 available at http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant 2.0. cgi and it is a Support Vector Machine (SVM) based tool for the automatic prediction of protein stability changes upon single point mutations. I-Mutant2.0 predictions are performed starting either from the protein structure or, more importantly, from the protein sequence (Capriotti *et al.*, 2005). The method was trained and tested on a data set derived from ProTherm (Bava *et al.*, 2004), which is presently the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability upon mutation under different conditions. Four different outputs can be retrieved, depending on the selected mode. In either case, the prediction of the value of free energy change or only its sign can be obtained. Positive DDG value indicates that the mutated protein is expected to have high stability whereas negative values indicate lower stability.

Functional significance of noncoding SNPs in regulatory untranslated regions: Recent studies shown that SNPs may have functional effects on protein structures, by changing single amino acids, transcriptional regulation and on alternative splicing regulations (Prokunina *et al.*, 2002; Prokunina and Alarcon-Riquelme, 2004). We used the program available at http://fastsnp.ibms.sinica.edu.tw. The function analysis and selection tool (FastSNP) was used to predict the risk level of SNPs according to its phenotypic and putative functional effects. The impact of the nsSNPs, 3' UTR regions and intronic SNPs on the regulation of the *KIT* gene was analyzed. FastSNP server follows the decision tree principle with external web service access to TF search, to obtain the predicted transcription factor-binding sites of the gene. Thus the decision tree will assign risk rankings for SNP prioritization with a ranking of 0, 1, 2, 3, 4 and 5. This signifies the levels of no, very low, medium, high and very high effects, respectively (Yuan *et al.*, 2006).

Modeling of nsSNPs on protein structures and calculation of their RMSD difference: Structural bioinformatics is concerned with computational approaches were used to predict and analyze the spatial structure of proteins. In this view, homology modeling or comparative modeling is based on the fact that a structure of a protein can be reliably modeled when its sequence is

sufficiently similar to a protein sequence with known 3D structure (Prajapat *et al.*, 2010). The available structure for the *KIT* gene has a Protein Database Bank id (PDBid) of 1 Pkg (Mol *et al.*, 2003). We confirmed the mutation position and mutation residues in 1 Pkg by using I-Mutant2.0, SIFT and PolyPhen servers. The mutation and the energy minimization of the mutant structure were performed using the SWISS PDB viewer. Divergence of mutant structure from native structure is due to mutation, deletions and insertions and the deviation between the two structures is evaluated by their RMSD values which could affect stability and functional activity (Rajasekaran *et al.*, 2008b). The superimposition of native and mutant structures was evaluated using Superpose server (Maiti *et al.*, 2004).

Molecular interaction and binding affinity with sunitinib and HDAC inhibitors: Molecular interaction is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. In order to calculate the binding affinity and interaction between native and mutant protein with sunitinib and Histone deacetylase (HDAC) inhibitors, we used Glide 5.5 module of Schrodinger suite (Friesner *et al.*, 2004). The native protein (1 Pkg) and mutant structure (D816V) were selected for analysis and the structures with favorable shape similarity measures were prepared for docking simulations using the Protein Preparation Wizard Script within Maestro. This protein preparation procedure involves removal of waters, optimization of contacts by changing hydroxyl group orientations, flipping of Asn and Gln side chains and selecting His tautomeric states, followed by constrained energy refinement using the OPLS-AA force field. The chemical structures of the drug sunitinib (CID: 5329102) and HDAC inhibitors namely Trichostatin A, Panobinostat (CID: 444732, 6918837), were retrieved from Pubchem database. Ligands were then prepared and geometry optimized by using the optimized potentials for liquid simulation-2005 (OPLS) force field using the LigPrep process (Schrodinger, 2009). The LigPrep is utility in Schrodinger software suite that combines tools for generating 3D structures from 1D (smiles) and 2D (SDF) representation, searching for tautomers and steric isomers and geometry minimization of ligands. The prepared protein and the ligands were employed to build energy grids using the default value of protein atom scaling (1.0 Å) within a cubic box, centered on the centroid of the X-ray ligand pose. After Grid generation, the ligands were docked with the protein by using Glide 5.5 module in extra precision mode (XP) which uses MCSA (Monte Carlo Based Simulated Algorithm) based minimization. The docked results obtained from

Glide were analyzed. Finally, the best orientation for each docked compound was rescored according to its binding free energy, ΔG_{bind} which was calculated using the Prime MM-GBSA module in Maestro (Guimaraes and Cardozo, 2008).

RESULTS AND DISCUSSION

Datasets: SNPs have become marker of choice for many applications in genome analysis because SNPs are abundance, stable, ubiquity and interspersed in nuclear genome (Tchin *et al.*, 2011). Single nucleotide polymorphism data of human *KIT* gene was retrieved from the dbSNP database. It consists of a total of 1,288 SNPs, of which 35 were non-synonymous; we took these non-synonymous SNPs for our computational analysis.

Deleterious nsSNPs by SIFT program: The substitutions which may affect the protein functions were identified by using sequence homology based tool. The protein sequence accession number along with the two alleles for each of the 35 nsSNPs in *KIT* gene were submitted independently to the SIFT server to check their corresponding deleterious/tolerance index. Among the 35 nsSNPs, 11 were found to be deleterious with a tolerance index score of less than 0.05 and the results are shown in Table 1. Of 11 nsSNPs, 5 showed a high tolerance index of 0.00, 2 had a tolerance index of 0.01 and 4 nsSNPs had a tolerance index of 0.02 respectively. We found that five nsSNPs, namely rs121913507, rs28933371, rs121913508, rs121913509, rs121913513 which were showed a less stability by I-Mutant2.0 server, were also seen to be deleterious by SIFT server. Low tolerance index will have higher functional impact of particular amino acid. This low index value of the nsSNP (rs121913507) indicates its involvement in changing protein stability and its functionality (Ng and Henikoff, 2003).

Damaged nsSNP found by PolyPhen server: The structural level of the alteration in protein was analyzed using the PolyPhen server. The protein sequences of 35 nsSNPs examined in this work were submitted as input to the PolyPhen server and the results are shown in Table 1. The position specific independent count score was predicted for all 35 nsSNP. Among these, we found 11 nsSNPs were found to be damaging and only one nsSNPs (rs121913507) exhibited a highest PSIC score of 3.185 (Table 1) and it was supposed to affect protein structure and function. A PSIC score difference of 1.1 and above is considered to be damaging (Ramensky *et al.*, 2002). The five nsSNPs namely rs121913507, rs2893337, rs121913508, rs121913509 and rs121913513 which were

Table 1: Deleterious SNPS found by SIFT, PolyPhen and I-Mutant2.0 servers

SNP ID	Nucleotide change	Amino acid change	Tolerance	PSIC score	DDG
rs121913523	T/C	V556G	0.01	2.209	0.54
rs121913522	G/T	V650A	0.01	2.281	-0.51
rs121913518	A/T	V555D	0.02	2.686	-0.21
rs121913514	A/T	N818K	0.02	2.672	-0.18
rs121913513	C/T	L576P	0.00	2.806	-0.21
rs121913509	A/G	E839K	0.00	2.256	-2.03
rs121913508	G/T	D816Y	0.00	2.995	-0.46
rs121913507	A/T	D816V	0.00	3.185	-0.67
rs121913506	C/G	D816H	0.02	1.632	-0.72
rs56225530	C/T	T488M	0.02	2.221	-1.38
rs28933371	G/T	F584C	0.00	2.880	-1.91

predicted to be deleterious by SIFT server, were also found to be damaging by PolyPhen server. Meanwhile, other studies had also reported on the association of SNPs with various applications such as the search for SNPs in PS1 in AD patients (Al-Khedhairy *et al.*, 2005), identification of Polymorphic Sites (1236 and 3435) in Multi Drug Resistance Gene 1 Influencing Drug Response in Breast Cancer Patients (Khan *et al.*, 2007), genetic polymorphisms in the ESR genes and the risk of breast cancer among Iranian women (Abbasi *et al.*, 2009a), genome wide SNP analysis in *Mycobacterium* sp. (Srivastava *et al.*, 2009).

Predicting stability change on mutation by I-Mutant2.0: Out of 11 non-synonymous SNPs, the server predicts 10 SNPs that were found to be less stable. Among these SNPs, 3 nsSNPs namely rs56225530, rs28933371 and rs121913509 showed a DDG value of >-1.0. The remaining 7 nsSNPs namely rs121913506, rs121913507, rs121913522, rs121913508, rs121913513, rs121913518 and rs121913514 showed a DDG value <-1.0 (Table 1). The high negative DDG values of nsSNPs predicted by this server indicate, less stable the given point mutation is likely to have (Capriotti *et al.*, 2005). Hence, based on the results predicted from all the three programs we concluded that only one nsSNPs namely rs121913507 was seen to be less stable, deleterious and found to be damaging. So we took this nsSNP for mutational and molecular interaction analysis.

Functional SNPs in UTR found by the FastSNP: We used FastSNP server to analyze and predict the functional significance of SNPs in *KIT* gene. According to this server, 21 coding SNPs were predicted to be functionally significant (Table 2). Among these, three SNPs (rs28933968, rs56094246 and rs28933371) were predicted with a risk rank of medium to high in splicing site and splicing regulation. It is clear from the result that these SNPs breaks the exonic splicing enhancer/silencer binding site in coding sequence, leading to abolish the protein domain (Yuan *et al.*, 2006).
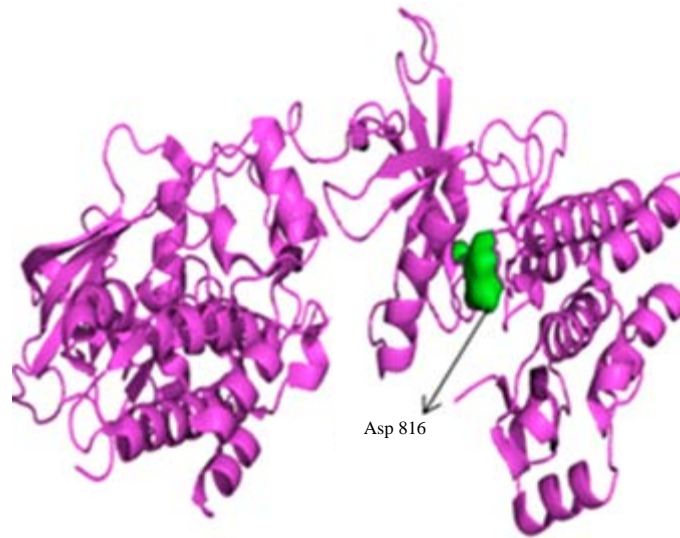
Fig. 1: Native structure (pink) of *KIT* gene encoding receptor tyrosine kinase (1 Pkg)

Table 2: List of SNPs found to be functionally significant by FastSNP server

dbSNP ID	Nucleotide change	Region	Level of risk	Possible functional effect
rs28933968	A/T	coding	Medium-High(3-4)	Splicing site
rs28933371	G/T	coding	Medium-High(3-4)	Splicing regulation
rs56094246	A/G	coding	Medium-High(3-4)	Splicing site
rs55755457	C/T	coding	Low-Medium(2-3)	Splicing regulation
rs72549300	A/T	coding	Low-Medium(2-3)	Splicing regulation
rs72549301	A/C	coding	Low-Medium (2-3)	Splicing regulation
rs56411694	G/T	coding	Low-Medium (2-3)	Splicing regulation
rs72549293	C/T	coding	Low-Medium(2-3)	Splicing regulation
rs73137716	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs72549294	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs55966164	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs56225530	C/T	coding	Low-Medium(2-3)	Splicing regulation
rs72550822	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs55792975	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs3822214	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs55986963	A/G	coding	Low-medium(2-3)	Splicing regulation
rs35200131	C/G	coding	Low-Medium(2-3)	Splicing regulation
rs55717477	C/T	coding	Low-Medium(2-3)	Splicing regulation
rs28933969	C/G	coding	Low-Medium(2-3)	Splicing regulation
rs55817813	A/G	coding	Low-Medium(2-3)	Splicing regulation
rs72549297	A/G	coding	Low-Medium(2-3)	Splicing regulation

Modeling of mutant structure and calculation of their RMSD difference: Information about mapping the deleterious nsSNPs in the protein structure was obtained from dbSNP database. The native protein structure associated with *KIT* gene was retrieved from dbSNP database with a PDB id: 1 Pkg (Fig. 1). Based on the result obtained from SIFT, PolyPhen server and I-Mutant2.0, we took the nsSNP rs121913507 for structural analysis. The amino acid variant and the mutational position were identified to be D/V at the residue position 816 (i.e., aspartic acid was replaced by valine) and the structure of

mutant protein is shown in Fig. 2. This result also coincides well with the experimental observation (Ma *et al.*, 2002). The mutation and energy minimization analysis were performed using SWISSPDB viewer. The total energy for the native and mutant structures were found to be -22,483.596 and -22,395.652 kcal mol^{-1}, respectively. The higher the total energy, the less stable the protein structure is. The total energy comparison showed that the mutant protein is little higher in its energy value than native type (Rajasekaran *et al.*, 2008c). The superimposition of mutant and native structure was performed in SuperPose server and the RMSD value difference was found to be 1.41 Å. The observed RMSD value denotes that some deviation exists between the mutant and native structures which in turn may alters their normal and efficient activity. The superimposed structure of native (PDB ID: 1 Pkg) and mutant type protein (rs121913507) is shown in Fig. 3. Mutation and polymorphism of cancer-associated genes have been found to predict tumor formation and prognosis. It is also considered as an effective risk factor with positive effects and negative effects in different studies (Abbasi *et al.*, 2009b). Based on these results, the mutant allele at the nsSNP (rs121913507) was found to be less stable, deleterious and could be believed to affect the structure of the protein and the mutation occurring with this particular nsSNPs would be of key importance in the identification of GISTs caused by the *KIT* gene. Single amino acid mutation can significantly change the stability of a protein structure. Thus, biologists and protein
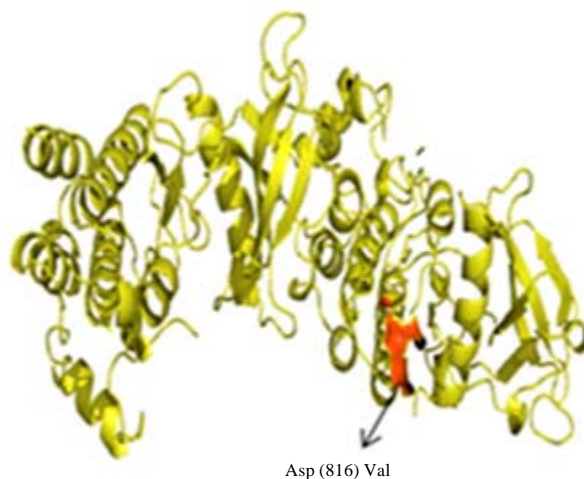
Asp (816) Val

Fig. 2: Mutant structure (yellow) of *KIT* gene encoding receptor tyrosine kinase (1 Pkg)
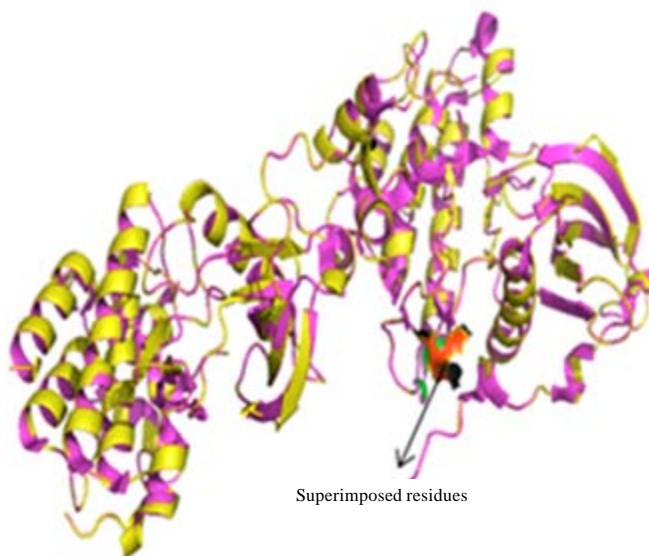

Superimposed residues

Fig. 3: Superimposed structure of native protein 1 Pkg (pink) with mutant structure (yellow) D816V

designers need to find out accurate predictions of how such mutations will affect the stability of a protein structure (De Alencar and Lopes, 2010).

Analysis of binding affinity for native and mutant protein:
We used Glide 5.5 module of Schrodinger suite to find out the molecular interaction between the receptor and the ligand. The X-ray structure of c-*KIT* protein tyrosine kinase (PDB ID: 1 Pkg) complex with ADP, magnesium ions was selected as the reference structure for molecular docking analyses. In this study, we used the tyrosine

kinase inhibitor sunitinib and HDAC inhibitors (Trichostatin A and panobinostat) for docking studies with both native and mutant proteins. Sunitinib has recently been approved for patients with advanced GISTs who have progressed on imatinib (Demetri *et al.*, 2006). HDAC inhibitors have recently shown promising results in preclinical models for GIST (Muhlenberg *et al.*, 2009) and are currently in clinical trials. The docking results of sunitinib and HDAC inhibitors are reported in Table 3. The method of evaluating the accuracy of a docking procedure is to determine how closely the lowest energy

Table 3: Glide extra-precision (XP) results for native and mutant protein

| Structure	Ligands	Docking score	Glide energy	No. of hydrogen bond interactions	Interaction residues
Native (1 Pkg)	Sunitinib	-5.165	-35.612	3	Leu 595, Asp 810, Arg 796
	Trichostatin A	-3.910	-32.052	3	Asp 816, Asp 572, Val 824
	Panobinostat	-4.447	-43.754	4	Asp 572, Asp 816, Glu 635, Gly 812
Mutant(D816 V)	Sunitinib	-5.353	-31.389	3	Asn 797, Asn 792, Arg 796
	Trichostatin A	-7.582	-51.639	3	Asp 677, Asn 797, Glu 640
	Panobinostat	-6.423	-42.155	2	Asp 810, Asp 792
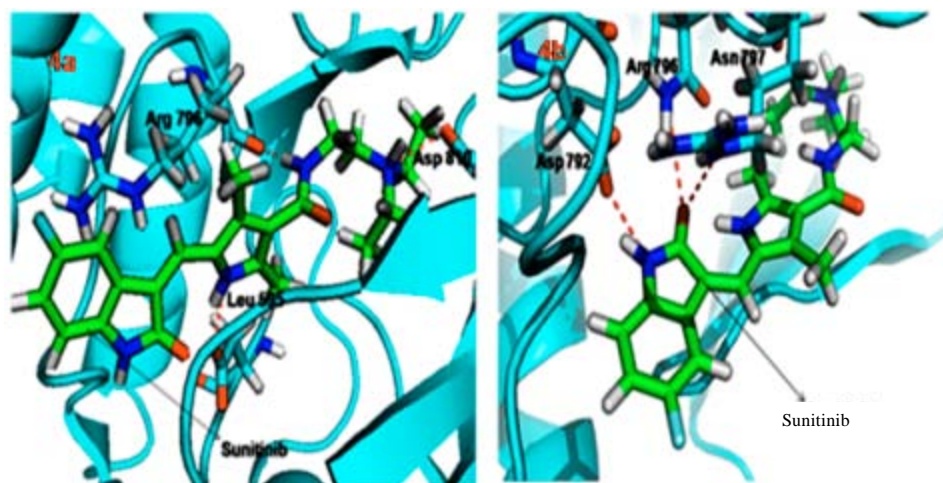
Fig. 4(a-b): (a) Sunitinib docked with native and (b) Sunitinib docked with mutant

pose (binding conformation) predicted by the object scoring function, Glide score (G score). We found three hydrogen bonding interactions that exist between native type-sunitinib and also between mutant type-sunitinib with difference in two amino acid residues. The residue Arg 796 is common in both types. The hydrogen bonding interactions was also studied comparatively with HDAC inhibitors (Trichostatin A, panobinostat). Trichostatin A forms three hydrogen bonds with native and mutant proteins. On the other hand, Panobinostat forms four hydrogen bonds with native type and only two hydrogen bonds with mutant type. From the docking results, it was found that the HDAC inhibitors: Trichostatin A (-7.582 kcal mol^{-1}) and Panobinostat (-6.423 kcal mol^{-1}) possess the maximum Glide docking score and glide energy with mutant structure when compared to sunitinib (-5.353 kcal mol^{-1}) respectively. The amino acid residue Asp 816 is common in forming interaction between HDAC inhibitors with native protein and this indicates that Asp 816 plays crucial role in maintaining closed or open conformation of c-*KIT* protein tyrosine kinase domain (Paulson *et al.*, 1996). Both HDAC inhibitors and sunitinib forms low glide docking score with native protein than mutant structure.

The complex of sunitinib with native and mutant type is shown in Fig. 4a and b. The complex of trichostatin A and Panobinostat with native and mutant structures are shown in Fig. 5a, b and 6a, b. These results imply that HDAC inhibitors can be serve as best drug candidates for experimentally proven mutant allele (D816V) in GIST than sunitinib.

We further investigated the binding free energy for the native and mutant *KIT* gene encoding receptor tyrosine kinase and ligands. For this purpose, we used Prime MM-GBSA module for the nsSNP analysis in accordance with binding energy calculation. The binding energy scores can be used for a relative ranking of our test ligands. The binding free energy for sunitinib with native protein was -15.78 kcal mol^{-1}, whereas with mutant protein it was -17.52 kcal mol^{-1}. Moreover through this exploration, it was found that the HDAC inhibitors exhibit higher binding free energy with the mutant structure compared to native protein (Table 4). Although some interactions are similar as HDAC inhibitors, the docking score and calculated Prime MM-GBSA ΔG is worse than those calculated ones of HDAC inhibitors, indicating sunitinib may be have a worse binding affinity than HDAC inhibitors to mutant structure (D816V). In the past
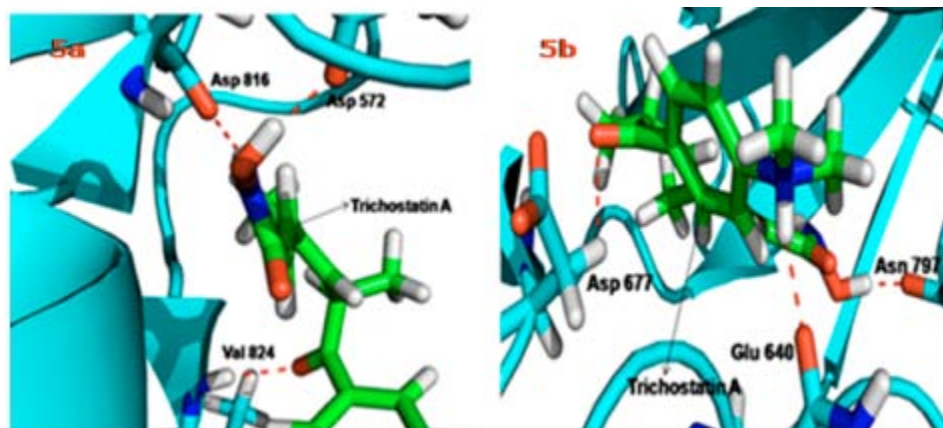
Fig. 5(a-b): (a) Trichostatin A docked with native and (b) Trichostatin A docked with mutant
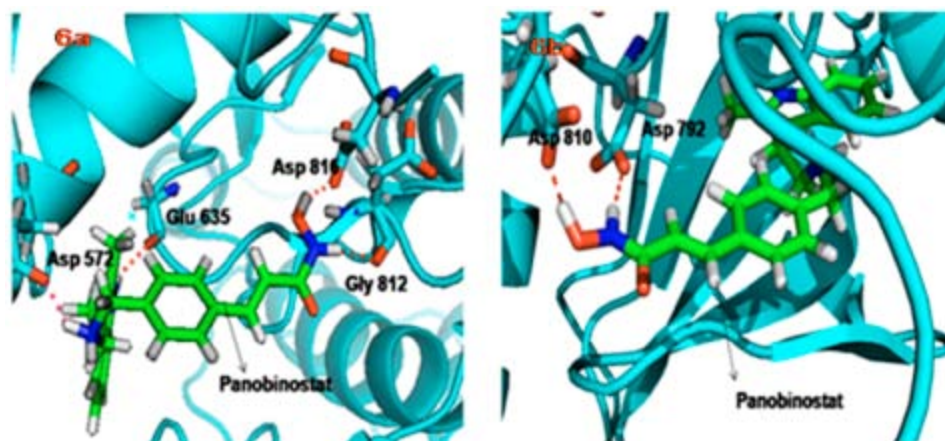

Fig. 6(a-b): (a) Panobinostat docked with native and (b) Panobinostat docked with mutant

Table 4: Binding energy calculated by prime MM-GBSA

	Binding energy (kcal mol^{-1})	
Ligands	Native (1 Pkg)	Mutant (D816 V)
Sunitinib	-15.78	-17.52
Trichostatin A	-20.27	-25.99
Panobinostat	-18.47	-22.30

this protocol has been used for productively re-scoring hits that may be ranked unfavorably by docking programs during virtual screening (Graves *et al.*, 2008; Guimaraes and Cardozo, 2008).

CONCLUSION

Defects in genes/enzymes could lead to chronic diseases in humans. Discovering the causes of this defect will provide a clue for its prevention or control of some disorders. In this study, we have analyzed the *KIT* gene through computational methods to predict its importance towards gastrointestinal stromal tumours and the influence of functional SNPs were also evaluated. The results obtained from SIFT, PolyPhen and I-Mutant2.0 servers, showed that 11 nsSNPs were found to cause damage to the *KIT* gene. Among these nsSNPs, we found the major mutation occurs in D/V at the position 816 and this deviation in the protein structure was caused by deleterious nsSNP rs121913507. Moreover, the docking study and binding affinity with sunitinib and HDAC inhibitors were carried out and result shows that the HDAC inhibitors exhibit better binding affinity with mutant type when compared to that of the native type. Hence from our analysis, we concluded that the above mutation is the major target for GISTs caused by *KIT* gene

and in addition, HDAC inhibitors can be served as the best drug for mutant as compared to native protein.

REFERENCES

Abbasi, S., C. Azimi, F. Othman, M.R.N. Daloii and Z.O. Ashtiani *et al.*, 2009a. Estrogen receptor-α gene codon 10 (T392C) polymorphism in Iranian women with breast cancer: A case study. Trends Mol. Sci., 1: 1-10.

Abbasi, S., P. Ismail., F. Othman., R. Rosli and C. Azimi, 2009b. Estrogen receptor-α gene, codon 594 (G3242A) polymorphism among Iranian women with breast cancer: A case control study. Asian J. Sci. Res., 2: 51-60.

Al-Khedhairy, A.A.A., M. Arfin , B.A. Alahmadi and M.A. Al-Jumah, 2005. Single nucleotide polymorphism associated with late-onset Alzheimer's disease. J. Medical Sci., 5: 275-279.

Ali, S. and S. Ali, 2007. Role of *c-KIT*/SCF in cause and treatment of gastrointestinal stromal tumors (GIST). Gene, 401: 38-45.

Bava, K.A., M.M. Gromiha, H. Uedaira, K. Kitajima and A. Sarai, 2004. ProTherm, version 4.0: Thermodynamic database for proteins and mutants. Nucleic Acids Res., 32: D120-D121.

Capriotti, E., P. Fariselli and R. Casadio, 2005. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res., 33: W306-W310.

Cmarik, J.L., 2008. From bioinformatics to bioassays: Gleaning insights into protein structure-function from disease-associated nsSNPs. Mol. Interv., 8: 162-164.

Collins, F.S., L.D. Brooks and A. Chakravarti, 1998. A DNA polymorphism discovery resource for research on human genetic variations. Genome Res., 8: 1229-1231.

De Alencar, S.A. and J.C.D. Lopes, 2010. A comprehensive in silico analysis of the functional and structural impact of SNPs in the IGF1R gene. J. Biomed. Biotechnol., Vol. 2010, 10.1155/2010/715139.

Demetri, G.D., A.T. van Oosterom, C.R. Garrett, M.E. Blackstein and M.H. Shah *et al.*, 2008. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: A randomised controlled trial. Lancet, 368: 1329-1338.

Duensing, A., M.C. Heinrich, C.D. Fletcher and J.A. Fletcher, 2004. Biology of gastrointestinal stromal tumors: *KIT* mutations and beyond. Cancer Invest., 22: 106-116.

Fredman, D., M. Siegfried, Y.P. Yuan, P. Bork, H. Lehvaslaiho and A.J. Brookes, 2002. HGVbase: A human sequence variation database emphasizing data quality and a broad spectrum of data sources. Nucleic Acids Res., 30: 387-391.

Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren and J.J. Klicic *et al.*, 2004. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem., 47: 1739-1749.

Gonzalez, I., E.J. Andreu, A. Panizo, S. Inoges and A. Fontalba *et al.*, 2004. Imatinib inhibits proliferation of Ewing tumor cells mediated by the stem cell factor/*KIT* receptor pathway and sensitizes cells to vincristine and doxorubicin-induced apoptosis. Clin. Cancer Res., 10: 751-761.

Graves, A.P., D.M. Shivakumar, S.E. Boyce, M.P. Jacobson, D.A. Case and B.K. Shoichet, 2008. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing. J. Mol. Biol., 377: 914-934.

Guimaraes, C.R.W. and M. Cardozo, 2008. MM-GB/SA rescoring of docking poses in structure-based lead optimization. J. Chem. Inf. Model., 48: 958-970.

Hirota, S., K. Isozaki, Y. Moriyama, K. Hashimoto and T. Nishida *et al.*, 1998. Gain-of-function mutations of c-*KIT* in human gastrointestinal stromal tumors. Science, 279: 577-580.

Hornick, J.L. and C.D. Fletcher, 2007. The role of KIT in the management of patients with gastrointestinal stromal tumors. Hum. Pathol., 38: 679-687.

Javed, R. and Mukesh, 2010. Current research status, databases and application of single nucleotide polymorphism. Pak. J. Biol. Sci., 13: 657-663.

Khan, M. and K. Jamil, 2008. Study on the conserved and the polymorphic sites of MTHFR using bioinformatics approaches. Trends Bioinform., 1: 7-17.

Khan, S., K. Jamil, G.P. Das , C.M. Vamsy and S. Murthy, 2007. Polymorphic sites (1236 and 3435) in multi drug resistance gene 1 influencing drug response in breast cancer patients. Int. J. Pharmacol., 3: 453-460.

Lasota, J., M. Jasinski, M. Sarlomo-Rikala and M. Miettinen, 1999. C-KIT mutations occur preferentially in malignant vs. benign gastrointestinal stromal tumors and do not occur in leiomyomas and leiomyosarcomas. Am. J. Pathol., 154: 53-60.

Lee, J.M., 1998. Inhibition of p53-dependent apoptosis by the KIT tyrosine kinase: Regulation of mitochondrial permeability transition and reactive oxygen species generation. Oncogene, 17: 1653-1662.

Lux, M.L., B.P. Rubin, T.L. Biase, C.J. Chen and T. Maclure *et al.*, 2000. KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors. Am. J. Pathol., 156: 791-795.

Ma, Y., S. Zeng, D.D. Metcalfe, C. Akin and S. Dimitrijevic *et al.*, 2002. The c-KIT mutation causing human mastocytosis is resistant to STI571 and other KIT kinase inhibitors: Kinases with enzymatic site mutations show different inhibitor sensitivity profiles than wild-type kinases and those with regulatory-type mutations. Blood, 99: 1741-1744.

Maiti, R., G.H. van Domselaar, H. Zhang and D.S. Wishart, 2004. SuperPose: A simple server for sophisticated structural superposition. Nucleic Acids Res., 32: W590-W594.

Majumder, S., K. Brown, F.H. Qiu and P. Besmer, 1988. c-kit protein, a transmembrane kinase: Identification in tissues and characterization. Mol. Cell Biol., 8: 4896-4903.

Mol, C.D., K.B. Lim, V. Sridhar, H. Zou and E.Y.T. Chien *et al.*, 2003. Structure of a c-kit product complex reveals the basis for kinase transactivation. J. Biol. Chem., 278: 31461-31464.

Mol, C.D., D.R. Dougan, T.R. Schneider, R.J. Skene and M.L. Kraus *et al.*, 2004. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. J. Biol. Chem., 279: 31655-31663.

Muhlenberg, T., Y. Zhang, A.J. Wagner, F. Grabellus and J. Bradner *et al.*, 2009. Inhibitors of deacetylases suppress oncogenic KIT signaling, acetylate HSP90 and induce apoptosis in gastrointestinal stromal tumors. Cancer Res., 69: 6941-6950.

Ng, P.C. and S. Henikoff, 2003. SIFT: Predicting amino acid changes that affect protein function. Nucl. Acids Res., 31: 3812-3814.

Nocka, K., J. Buck, E. Levi and P. Besmer, 1990. Candidate ligand for the c-kit transmembrane kinase receptor: KL, a fibroblast derived growth factor stimulates mast cells and erythroid progenitors. EMBO J., 9: 3287-3294.

Nuchnoi, P., D. Nantakomol, V. Chumchua, C. Plabplueng, C. Isarankura-Na-Ayudhya and V. Prachayasittikul, 2011. The identification of functional non-synonymous SNP in human ATP binding cassette (ABC), subfamily member 7 gene: Application of bioinformatics tools in biomedicine. J. Bioanal. Biomed., 3: 26-31.

Paulson, R.F., S. Vesely, K.A. Siminovitch and A. Bernstein, 1996. Signalling by the *W/Kit* receptor tyrosine kinase is negatively regulated *in vivo* by the protein tyrosine phosphatase Shp1. Nat. Genet., 13: 309-315.

Prajapat, R., R.K. Gaur, R. Raizada and V.K. Gupta, 2010. *In silico* analysis of genetic diversity of begomovirus using homology modelling. J. Biol. Sci., 10: 217-223.

Prokunina, L., C. Castillejo-Lopez, F. Oberg, I. Gunnarsson and L. Berg *et al.*, 2002. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat. Genet., 32: 666-669.

Prokunina, L. and M.E. Alarcon-Riquelme, 2004. Regulatory SNPs in complex diseases: Their identification and functional validation. Expert Rev. Mol. Med., 6: 1-15.

Rajasekaran, R., C.G.P. Doss, C. Sudandiradoss, K. Ramanathan and R. Sethumadhavan, 2008a. *In silico* analysis of structural and functional consequences in p16INK4A by deleterious nsSNPs associated *CDKN2A* gene in malignant melanoma. Biochimie, 90: 1523-1529.

Rajasekaran, R., C.G.P. Doss, C. Sudandiradoss, K. Ramanathan, R. Purohit and R. Sethumadhavan, 2008b. Effect of deleterious nsSNP on the HER2 receptor based on stability and binding affinity with herceptin: A computational approach. Comptes Rendus Biol., 331: 409-417.

Rajasekaran, R., G.P. Doss, C. Sudandiradoss, K. Ramanathan, P. Rituraj and R. Sethumadhavan, 2008c. Computational and structural investigation of deleterious functional SNPs in breast cancer BRCA2 gene. Sheng Wu Gong Cheng Xue Bao, 24: 851-856.

Ramensky, V., P. Bork and S. Sunyaev, 2002. Human non-synonymous SNPs: Server and survey. Nucl. Acids Res., 30: 3894-3900.

Rubin, B.P., S. Singer, C. Tsao, A. Duensing and M.L. Lux *et al.*, 2001. KIT activation is a ubiquitous feature of gastrointestinal stromal tumors. Cancer Res., 61: 8118-8121.

Schrodinger, 2009. Ligprep, User Manual. Version 2.3. Schrodinger Press, LLC, New York.

Sherry, S.T., M. Ward and K. Sirotkin, 1999. DbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res., 9: 677-679.

Srivastava, S.K., M. Agrawal and M. Grover, 2009. Genome wide single nucleotide polymorphism analysis of mycobacterium species and subspecies. Res. J. Microbiol., 4: 112-121.

Stitziel, N.O., Y.Y., Tseng, D. Pervouchine D. Goddeau, S. Kasif and J. Liang, 2003. Structural location of disease-associated single-nucleotide polymorphisms. J. Mol. Biol., 327: 1021-1030.

Sun, J., M. Pedersen and L. Ronnstrand, 2009. The D816V mutation of c-kit circumvents a requirement for Src family kinases in c-Kit signal transduction. J. Biol. Chem., 284: 11039-11047.

Tchin, B.L., W.S. Ho, S.L. Pang and J. Ismail, 2011. Gene-associated single nucleotide polymorphism (SNP) in cinnamate 4-hydroxylase (C4H) and cinnamyl alcohol dehydrogenase (CAD) genes from *Acacia mangium* superbulk trees. Biotechnology, 10: 303-315.

Wang, Z. and J. Moult, 2001. SNPs, protein structure and disease. Hum. Mutat., 17: 263-270.

Yasuoka, R., C. Sakakura, K. Shimomura, Y. Fujita and M. Nakanishi *et al.*, 2003. Mutations in exon 11 of the c-kit gene in a myogenic tumor and a neurogenic tumor as well as in gastrointestinal stromal tumors. Utility of c-kit mutation as a prognostic biomarker for gastrointestinal mesenchymal tumor. Dig. Surg., 20: 183-191.

Yuan, H.Y., J.J. Chiou, W.H. Tseng, C.H. Liu and C.K. Liu *et al.*, 2006. FASTSNP: An always up-to-date and extendable service for SNP function analysis and prioritization. Nucl. Acids Res., 34: W635-W641.